Complexity trade-offs in core argument marking

Kaius Sinnemäki
University of Helsinki

Languages have often been claimed to trade off complexity in one area with simplicity in another. The present paper tests this claim with a complexity metric based on the functional load of different coding strategies (head/dependent marking and word order) that interact in core argument marking. Data from a sample of 50 languages showed that the functional use of word order had a statistically significant inverse dependency with the presence of morphological marking, especially with dependent marking. Most other dependencies were far from statistical significance and in fact provide evidence against the trade-off claim, leading to its rejection as a general all-encompassing principle. Overall, languages seem to adhere more strongly to distinctiveness than to economy.

1. Introduction

The recent wake of interest in language complexity has inspired many linguists to develop methods for assessing the complexity of languages (McWhorter 2001; Kusters 2003; Dahl 2004; Hawkins 2004).[1] One of the central interests has been to evaluate the validity of the old claim that all languages are equally complex, expressed in a succinct way by Crystal (1997: 6): "All languages have a complex grammar: there may be relative simplicity in one respect (e.g., no word-endings), but there seems always to be relative complexity in another (e.g., word-position)." In other words, the locus of complexity varies but overall, there is a balance across languages: if one area of grammar (e.g., morphology) is complex, another area (e.g., syntax) is in turn simple, varying from language to language.

1. I would like to thank Fred Karlsson, Matti Miestamo, Seppo Kittilä, and one anonymous reviewer for helpful comments on earlier versions of this paper. Research for this article has been funded by the Academy of Finland under grant 201601 "Complexity in natural languages" and the Finnish Graduate School in Language Studies (Langnet). Their support is gratefully acknowledged. Earlier versions of this paper have been presented at three conferences: ALT VI in Padang, July 21–25, 2005, Approaches to Complexity in Language in Helsinki, August 24–26, 2005, and Leipzig Students' Conference in Linguistics, March 25–26, 2006. Thanks are due to the audiences for helpful comments. I alone am responsible for any remaining errors.

This kind of complexity *trade-off* (also called compensation or balance) is often called for as a proof for equal complexity of languages. Countless scholars have subscribed to similar views without evaluating them more seriously or testing them systematically.

My purpose is to test whether complexity trade-offs are a general all-encompassing principle in language or whether they occur merely in some subdomains or not at all, henceforth called the *trade-off hypothesis*. I suspect trade-offs are more likely between functionally connected variables. For instance, it is intuitively more conceivable that syllable complexity would depend on the size of consonant inventory rather than e.g., the complexity of gender marking. In fact, Maddieson (2006) established a statistically significant positive correlation between the members of the former pair. Moreover, it is unnecessary to study a plethora of features to establish the presence or absence of trade-offs as a general all-encompassing principle. If such generality was to be found, every pair of variables in each area of language should manifest trade-offs, and therefore it should suffice to study even a single area to capture this generality.

To test the hypothesis, I confine myself to studying the interaction of coding strategies involved in the marking of a single functional domain. A functional domain (in the sense of Givón 1981) is a set of closely related semantic or pragmatic functions that are linguistically encoded by at least some languages (e.g., aspect or passive) (cf. Miestamo 2007: 293). Due to this restriction the chosen domain should be as universal as possible. This requirement is met by core argument marking, one of the most "universal" and best documented functional domains in languages. Core arguments are the arguments of a prototypical two-place transitive predicate, one being more agent-like (A) and the other more patient-like (P) (Comrie 2005: 398).

Four linguistic strategies are employed in the marking of core arguments: dependent marking (e.g., case affixes), head marking (e.g., person agreement on the verb), word order, and lexicon.[2] The first three of these are structural strategies and they are the focus of this paper. Although these structural strategies operate on different *additional* functional domains – for example head marking is relevant to referentiality, dependent marking to affectedness (Næss 2004), and word order to focus marking – they interact in core argument marking. According to the trade-off hypothesis, their interaction should not be arbitrary but such that complexity in one strategy is balanced out by simplicity in another. I test this claim statistically with data from 50 languages. The sample is genealogically and areally stratified and randomly chosen; the data is drawn from available reference grammars.

The rest of the paper has the following structure. The language sample is presented in Section 2, followed by a discussion of trade-offs and of the definition of complexity in Section 3. Section 4 formulates the hypothesis for the statistical tests, and in Section 5 I define the parameters to be tested. Section 6 discusses the complexity metric and the data. The results are presented and discussed in Section 7 and conclusions in Section 8.

---

2. Strictly speaking, word order and lexicon do not encode but rather distinguish the core arguments from one another. The role of non-linguistic features, such as context, is not discussed.

Table 1. The distribution and number of genera in the WALS and the language sample.

	Number of genera	
Macro-areas	**WALS**	**Sample**
Africa	64	7
Eurasia	38	4
South East Asia and Oceania	46	5
Australia and New Guinea	125	13
North America	93	10
South America	93	11
Total	459	50

2. Sampling

In order to study trade-offs statistically, a representative and well-balanced random sample is needed. A random and genealogically as well as areally balanced sample is difficult to obtain by sampling mere languages, so it is better to sample groups of languages. For this purpose, Dryer's (1992) method of grouping languages into genera (genealogic subgroupings of languages roughly equivalent with e.g., Germanic or Romance languages) provides a workable solution. He further divides the world into six macro-areas that have relatively equal proportions of genealogical and typological variation. This classification is also used in *The World Atlas of Language Structures* (WALS, Dryer 2005), which I used as a basis for stratification. The total number of genera of spoken languages in the WALS is 459; Table 1 shows their distribution according to macro-areas.[3]

The sample of 50 languages was created by taking from each macro-area a number of genera that is proportional to the total number of genera in that macro-area (Miestamo 2005: 31-39). Each macro-area is therefore represented to the same proportion. The sample is genealogically stratified so that no two languages come from the same genus, subfamily, or family (Table 2).[4] Since creoles have received a lot of attention in the recent discussion on language complexity, one creole was also sampled (one genus was consequently reduced from Australia and New Guinea).

3. The proportions of the sampled genera in North America and South America differ from one another because the one creole sampled, Berbice Dutch Creole, was included in the numbers for South America. Otherwise, the numbers for South America would have been 92 (WALS column) and 10 (sample column).

4. In two macro-areas, in Africa as well as in Australia and New Guinea, small concessions were made at the highest levels of classification due to the unavailability of suitable grammars at the time of sampling.

Table 2. The language sample.

Area	Family	Genus	Language	Source
AFR	Afro-Asiatic	Berber	Middle Atlas Berber	(Penchoen 1973)
		Eastern Cushitic	Somali	(Saeed 1999)
	Khoisan	Central Khoisan	Khoekhoe	(Hagman 1977)
	Niger-Congo	Bantoid	Babungo	(Schaub 1985)
		Southern Atlantic	Kisi	(Childs 1995)
	Nilo-Saharan	Lendu	Ngiti	(Kutsch Lojenga 1994)
		Nubian	Dongolese Nubian	(Armbruster 1960)
EUR	Dravidian	Southern Dravidian	Kannada	(Sridhar 1990)
	Finno-Ugric	Ugric	Hungarian	(Rounds 2001)
	Indo-European	Celtic	Welsh	(King 1993)
	Kartvelian	Kartvelian	Georgian	(Harris 1981)
ANG	Australian	Bunuban	Gooniyandi	(McGregor 1990)
		Maran	Warndarang	(Heath 1980)
		Pama-Nyungan	Diyari	(Austin 1981)
	Dagan	Dagan	Daga	(Murane 1974)
	Kuot	Kuot	Kuot	(Lindström 2002)
	Sepik-Ramu	Lower Sepik	Yimas	(Foley 1991)
	Sko	Western Sko	Sko	(Donohue 2004)
	Solomons	Solomons	Lavukaleve	(Terrill 2003)
	East Papuan	East Papuan		
	Torricelli	Kombio-Arapesh	Arapesh	(Conrad and Wogiga 1991)
	Trans-New Guinea	Awju-Dumut	Kombai	(de Vries 1993)
		Timor-Alor-Pantar	Adang	(Haan 2001)
	West Papuan	North-Central Bird's Head	Maybrat	(Dol 1999)
	Yele	Yele	Yelî Dnye	(Henderson 1995)
SAO	Austro-Asiatic	Munda	Korku	(Nagaraja 1999)
	Austronesian	Sundic	Indonesian	(Sneddon 1996)
	Hmong-Mien	Hmong-Mien	Mien	(Court 1985)
	Sino-Tibetan	Qiangic	Qiang	(LaPolla 2003)
	Tai-Kadai	Kam-Tai	Thai	(Iwasaki and Ingkaphirom 2005)
NAM	Algonquian	Algonquian	Plains Cree	(Dahlstrom 1991)
	Eskimo-Aleut	Eskimo-Aleut	West Greenlandic	(Fortescue 1984)
	Hokan	Yuman	Maricopa	(Gordon 1986)

(*Continued*)

Table 2. Continued.

Area	Family	Genus	Language	Source
	Mayan	Mayan	Tzutujil	(Dayley 1985)
	Muskogean	Muskogean	Koasati	(Kimball 1991)
	Na-Dene	Athapaskan	Slave	(Rice 1989)
	Penutian	Miwok	Southern Sierra Miwok	(Broadbent 1964)
	Siouan	Siouan	Osage	(Quintero 2005)
	Uto-Aztecan	Corachol	Cora	(Casad 1984)
	Wakashan	Southern Wakashan	Nuuchahnulth	(Nakayama 2001)
SAM	Aymaran	Aymaran	Jaqaru	(Hardman 2000)
	Cariban	Cariban	Hixkaryana	(Derbyshire 1979)
	Chibchan	Aruak	Ika	(Frank 1990)
	Mura	Mura	Pirahã	(Everett 1986)
	Panoan	Panoan	Shipibo-Konibo	(Valenzuela 1997)
	Peba-Yaguan	Peba-Yaguan	Yagua	(Payne and Payne 1990)
	Quechuan	Quechuan	Imbabura Quechua	(Cole 1985)
	Trumai	Trumai	Trumai	(Guirardello 1999)
	Tupian	Tupi-Guaraní	Urubú-Kaapor	(Kakumasu 1986)
	Warao	Warao	Warao	(Romero-Figueroa 1997)
	Creole	Creole	Berbice Dutch Creole	(Kouwenberg 1994)

3. The definition of complexity

In order to define and measure complexity in a meaningful way, a good understanding of the notion of trade-off is first needed. Whereas trade-offs may be viewed simply as negative correlations between the complexities of different grammatical structures (however measured), it is more instructive to consider the underlying principles that could trigger them. Two underlying principles for trade-offs may be recognized, namely economy and distinctiveness. Both principles relate to the well-established principle of one-meaning–one-form but from different perspectives.

The principle of economy may be simply defined as one meaning encoded by only one form. Encoding via more than one form is avoided as well as unnecessary marking of distinctions. Violations of economy increase synonymy and redundancy (and hence complexity, Dahl 2004: 9–11, 51–55, 120–121) (see Miestamo, this volume, for detailed sub-criteria of the violations of the one-meaning–one-form principle). If in language change a new form emerges to compete for the same niche with an older one, either should be dispensed with: free variation between the strategies and the partitioning of the functional load violate the principle of economy (cf. Dahl 2004: 128–134). The principle of distinctiveness, on the other hand, is defined as one meaning encoded by at least some form. Distinctiveness is violated by one meaning being encoded by a non-unique form, i.e., if meaning distinctions made in the grammar are encoded by identical forms. Such insufficient encoding of meaning distinctions leads to homophony and ambiguity and increase processing difficulties (inefficiency in terms of Hawkins 2004). If in language change a coding strategy is lost, a new strategy should emerge to take over so that meaning distinctions are not lost.

The notion of trade-off assumes that the interaction between coding strategies interacting in the same functional domain is equally controlled by economy and distinctiveness: their combined outcome gears grammar towards adhering to the principle of one-meaning–one-form. Although their operation may be interpreted from the point of view of speaker and hearer preferences, I follow Dahl (2004: 39) in keeping difficulty and complexity apart from one another: difficulty is a matter of language use while complexity is an absolute property of the language system.

Complexity is typically measured as the number of parts or rules in a system (cf. Miestamo, this volume, for a detailed discussion). Less often the metrics have paid any attention to the functional load of transparent marking. Functional load is the measure of how often a contrast – a particular meaning distinction made with particular forms – is employed in a language. For instance, Hungarian always marks P with case and thus has a greater functional load than e.g., English, which has case marking of P only for personal pronouns. From an information theoretic point of view, great functional load means high frequency and thus great expectedness – a sign of low complexity. However, a metric based on such a definition of complexity would yield English case marking of P more complex than that of Hungarian. If, on the other hand, we take into account the asymmetries in the frequency and structural marking of the arguments, we come closer to a complexity metric that pays attention to functional loads.

In core argument marking, A is most frequently animate and definite whereas P is inanimate and indefinite and both typically lack overt structural marking. A is less frequently inanimate and indefinite and P animate and definite but these are more often structurally marked (Comrie 1989). In most contexts, A and P could be distinguished solely by semantic properties: where semantic properties satisfy distinctiveness, structural marking is unnecessary according to economy. According to Haspelmath (2006) however, structural asymmetries may be explained via frequency asymmetries instead of relying on (semantic) markedness: the most frequent members of a category are typically not structurally marked (probably due to their high expectedness), whereas less frequent members are. Arguments are thus expected to be distinguished only when A is inanimate and/or indefinite or P is animate and/or definite. Consequently, marking that occurs with the most frequent A and P is redundant and thus more complex (but in some ways also rather unexpected). Thus, the greater the functional load of a coding strategy, the more redundant and therefore the more complex it is.

For the present purposes, it is most relevant to measure how often a particular coding strategy distinguishes A from P. This translates roughly into a measure of differential argument marking. The details of measuring the functional load of the coding strategies will be discussed in Section 6.

4. Hypothesis for the statistical tests

Having discussed the nature of trade-offs, we may now formulate the hypothesis for statistical tests:

> the complexity values of different coding strategies depend on one another in an inverse way.

According to this formulation, the complexity values of different coding strategies should correlate negatively and in a statistically significant way. If a correlation is not statistically significant, the sign of correlation (+/–) might still be relevant. If the correlation is positive, we must reject the hypothesis. The closer the correlation is to zero, the smaller the relationship is between the variables leading to the rejection of the hypothesis.

Things are not this simple, however. Research by Fenk-Oczlon and Fenk (this volume) has indicated that trade-offs are more likely between coding strategies from different coding domains (e.g., one morphological and the other syntactic) than between strategies form the same coding domain. The hypothesis should be slightly modified to accommodate these tendencies:

> a negative correlation is expected for a pair of variables from different coding domains but a positive correlation is possible only domain-internally.

5. Coding strategies: Definitions and constraints

Three constraints, typical in word order studies, are observed in the analysis of the coding strategies. Firstly, the focus is on simple active affirmative indicative main clauses; marking outside this focus is excluded. Secondly, the marking involved with pronouns is excluded as well, ruling out coding strategies which occur only with pronouns or in the absence of full NPs. Although clauses with two full core NPs are rare in the spoken form of most languages (Du Bois 1987), two reasons justify excluding pronouns from the analysis. For one, this variation in argument marking is a consequence of how languages mark the functional domain of topic continuity; focusing on core argument marking of full NPs rules out the effects of a neighbouring functional domain that interacts with it. A more serious reason is that especially in the case of clitics it may be impossible to determine whether a pronoun instantiates head or dependent marking. A third constraint is that an argument is understood as being part of the clause proper only if it is not obligatorily separated from the rest of the clause by a pause and if there is no pronoun in situ replacing a transposed argument. Although earlier studies have noted that word order rules increase complexity, they have not been subjected to systematic cross-linguistic measures before.

The three structural coding strategies under investigation have traditionally been labelled case marking, agreement, and word order (WO). Since the first two terms have typically been associated with affixal morphology, I adopt and slightly modify Nichols' (1992) terms head and dependent marking, which cover both affixal and non-affixal marking.

For the present purpose, I define dependent marking (DM) and head marking (HM) in the following way. An argument NP is dependent-marked when an NP functions as a core argument (A or P) of a prototypical two-place transitive predicate and occurs in a particular form distinct from that of the other argument. The definition measures the degree of differential argument marking and therefore the arguments are treated together also in the statistical tests, i.e., only one column stands to represent dependent marking in Table 3. The definition pays no attention to the formal way in which the argument is marked as dependent. For instance, Kannada (1) uses an accusative suffix *-annu*, Urubú-Kaapor (2) uses a postposition *ke*, but Somali (3) uses different tones (*Cáli* in citation form, with high-low tone pattern, and *Cali* in the nominative, with low-low tone pattern).

Kannada (Southern Dravidian, Sridhar 1990: 86)

(1) *cirate* *mariy-annu* *nekkutti-de*
Leopard cub-ACC lick.PRS-3SG.N
'The leopard is licking the cub.'

Urubú-Kaapor (Tupi-Guaraní, Kakumasu 1986: 351)

(2) *pe* *tapĩ* *xa'e* ***ke*** *kutuk* *tĩ*
and Tapĩ Xa'e OBJ 3.pierce also
'And Tapĩ also pierced Xa'e (with an arrow).'

Somali (Eastern Cushitic, Saeed 1999: 59, 65, 229)

(3) ***Cali*** *warqáddíi* ***wuu*** *íi* *dhiibay*
Ali.NOM letter.ABS DECL.3 1SG.to pass.PST
'Ali passed the letter to me.'

Sometimes the formal marker occurs on some other constituent than the one whose role it indicates. In Yagua (4), the clitic *-níí* marks P as third person singular. However, it is attached neither to the predicate nor the NP itself but to the constituent preceding the NP. Since P functions as a reference point for the clitic placement, I treat it as dependent-marked.

Yagua (Peba-Yaguan, Payne and Payne 1990: 255)

(4) a. ***sa****-jimyiy* *Alchíco-****níí*** *quiiva*
3SG.A-eat Alchico-3SG.P fish
'Alchico is eating the fish.'

b. ***sa****-jimyiy-****níí*** *quiiva*
3SG.A-eat-3SG.P fish
'He is eating the fish.'

Head marking is defined as follows: an argument NP of a prototypical two-place transitive predicate is head-marked when the form of the predicate transparently determines that the argument has (a) certain feature(s). According to this definition the form of the predicate does not determine the role of the argument, but rather its form determines that the argument has particular properties: e.g., the A argument has properties third person singular neuter in (1). The definition pays no attention to the formal ways in which the features of the argument are marked on the head. For instance, Kannada (1) uses the suffix *-de*, Yagua (4) uses the proclitic *sa-*, but in Somali (3) the marking does not occur on the main verb, but on the declarative marker *waa*, which has fused with the third person singular marker to yield *wuu*. In the minimal case, the absence of overt formal marking (i.e., zero marking) on the predicate often indicates that the argument is in the third person. Languages also vary in how many features of the argument the predicate determines. Some determine only person (Urubú-Kaapor (2)), others person and number (Somali (3)), and still others person, number, and gender (Kannada (1)).

The third morphosyntactic strategy is word order. It distinguishes the core arguments when the role of the noun argument is determined by its position relative to the predicate, whereby the arguments may not occupy each other's position. In other words, if any of the *reversible word order pairs* AVP/PVA, APV/PAV, and VAP/VPA occur in a language, word order cannot distinguish the roles of the arguments. For instance, Thai (5) employs AVP order when the A is topical: reversing the order of the NPs changes their roles and thus PVA is impossible.

Thai (Kam-Tai, Iwasaki and Ingkaphirom 2005: 110)

(5) *lék tè nɔ̀ɔy*
Lek kick Noy
'Lek kicks Noy.' *'Noy kicks Lek.'

Thai may also topicalize the P by placing it clause-initially, but since this does not lead to a reversible order – AVP and PAV do not form a reversible word-order pair – word order distinguishes core arguments from one another. One should, of course, check all possible reversible word order pairs, not only the reversible word order pair involving the canonical word order. In Koasati, the canonical word order is APV but its reverse (PAV) does not occur (Kimball 1991: 513–517). Because another reversible word order pair is allowed, namely AVP and PVA, it is unlikely that word order distinguishes the roles of the arguments in Koasati at all.

Some languages allow for a reversible word order pair but either one or both orders are marked by changing the properties of head/dependent marking that occur in the canonical word order. By changes I mean removal of overt marking or a new set of markers, i.e., some marking which does not occur in the canonical order. Since the change in morphological marking parallels a change in word order, these orders would probably not be allowed for without the change in morphological marking. This leads us to think that word order helps to distinguish the roles in the canonical order. As an example, consider word order in Slave (6).

Slave (Athapaskan, Rice 1989: 1197)

(6) a. *lį* *ˀehkee* *kayįhshu*
dog (A) boy (P) 3.bit
'The dog bit the boy.'

b. *ˀehkee* *lį* *kayeyįhshu*
boy (P) dog (A) 3.bit.4
'The boy, a dog bit him.'

The canonical order is APV (6a), but when the object NP occurs clause-initially (6b), i.e., in its non-canonical position, the predicate is head-marked with the co-indexing pronoun *ye*-.

The three strategies defined in this section are morphosyntactic strategies. However, they are not the only *linguistic* strategies: the inherent lexical features of the arguments are often used as well. Some languages employ morphosyntactic marking only when the lexical features fail (Kittilä 2005). For instance, the agent is optionally dependent-marked in Qiang only when P is moved to a clause initial position or higher on the person hierarchy than A (7).

Qiang (Qiangic, LaPolla 2003: 78, 79, 122)

(7) a. *qa* *zawa* *ɦo-ɣlu-a*
1sg stone dir-roll-1sg
'I rolled the stone down.'

b. *mi-wu* *qa* *zə-dʐi*
person-a 1sg dir-hit
'Somebody hit me.'

c. *the:-tɕ* *pi:-xsə-la* *sum-wu* *de-l-ji* *ŋuə*
3sg-gen pen-three-clf teacher-a dir-give-csm cop
'*The teacher* gave him three pens.'

This paper focuses on morphosyntactic encoding and excludes the role of the lexicon in core argument marking.

6. Measuring functional load – and complexity

In Section 2, complexity of a coding strategy was defined as its functional load (FL). FL could of course be counted from actual texts but this was not possible within the limits of this paper. I rather estimated the FL of a coding strategy based on the grammar descriptions of each language.

Functional load forms a continuum whose end-poles are defined as *none* – a coding strategy never distinguishes the arguments in a particular language – and *full* – a coding strategy always distinguishes the arguments from one another. The FL of coding strategies is limited in various ways whereby it would be uninformative to assign the same values to all coding strategies with limited FL. The FL of argument marking is also often split between the coding strategies, even in a complementary fashion. Consequently, the complexity metric should capture trade-offs regardless of the FL being carried by a single strategy or by more than one complementary strategy. This is achieved by distinguishing *marginal* FL from *extensive* FL; impressionistically, the FL of a coding strategy is marginal when it distinguishes the arguments in less than half of the contexts, while it is extensive when it encodes them in roughly half or more of the contexts.

The values marginal and extensive are more precisely defined in terms of typological markedness (Croft 1990). Typological markedness is defined using three criteria one of which is the frequency criterion: the unmarked member of a category occurs at least as frequently as the marked member in a given language. Using the frequency criterion helps us to avoid problems with semantically based definitions of both markedness (Haspelmath 2006) and differential argument marking (cf. Næss 2004). The marked member never occurs more frequently than the unmarked member. Translating this into our metric, e.g., the most frequent P (inanimate and indefinite) is unmarked and it never occurs less frequently than the marked P (animate and definite). If a coding strategy encodes an argument when it is expected to, that is, when it is marked (and the least frequent), it is analysed as having marginal FL. If a coding strategy encodes an unmarked argument, it has extensive FL. For instance, in Kannada human patients are dependent-marked but because non-humans are optionally dependent-marked as well (Sridhar 1990: 86), the functional load of dependent marking is analysed as extensive. In (6) we saw that Slave uses word order to encode the core arguments only in the canonical (i.e., the most frequent and thus unmarked) order. Consequently, the FL of word order in Slave is extensive.

When the frequency criterion of typological markedness is irrelevant to the type of restriction, FL will be estimated by counting the number of relevant language specific contexts and estimating whether e.g., optionality refers to the evoking or the dropping of marking. For instance, Arapesh marks the person, number, and class of the patient on the head in transitive verb classes 1 and 2 but optionally in classes 3 and 4.[5] Class 1 is relatively small but classes 2–4 are rather large: this would indicate a marginal FL. However, Conrad and Wogiga (1991: 78) note that "... [w]ith relatively few exceptions, verb morphology is the same for all verbs, regardless of their distribution in a sentence and regardless of different patterns of affixation [of the verb classes, KS] ... " Dropping the verb morphology seems rather an exception than the norm in Arapesh, which leads us to analyse the FL of patient head-marking as extensive.

Note that the metric is able to capture complementary functional loads. In West Greenlandic the arguments are analysed as dependent-marked when the number of

5. Also in verb classes 7 and 8, but these classes contain only one verb each.

either or both is singular but not when both arguments are plural because the case marker for plural agent and patient is homophonous *-(i)t* (Fortescue 1984: 206–210). This results in extensive FL for dependent marking.[6] The noun arguments cannot occur in reversed order in this situation but only in AP order. The restrictions on the FL of word order are the mirror image of the restrictions of dependent marking, resulting in marginal FL.

The instances of marginal and extensive FL are described in detail in the lists below, containing only the ones not already discussed.

Marginal FL:

1. Adang: the person of P is head-marked by a small closed class of rather frequently used verbs. The arguments are also dependent-marked when focused, A optionally but P obligatorily.
2. Ika (A), Shipibo-Konibo (A), and Slave (P): plural number of the (parenthesized) argument is optionally head-marked.
3. Kannada: WO distinguishes the arguments when both are inanimate.
4. Korku: the person and number of animate patients are head-marked.
5. Sko: the gender and number of the patient are optionally head-marked when the patient is possessed.
6. Slave: A is head-marked when third person singular and in perfective mode but only with one conjugation marker, and optionally as third person plural when animate.
7. Urubú-Kaapor: P is optionally dependent-marked.
8. Yagua: (roughly) definite patients are dependent-marked.

Extensive FL:

9. Cora, Kuot, Slave, Trumai, Yagua: word order distinguishes the arguments in canonical order.
10. Gooniyandi: inanimate agents are dependent-marked, animate agents optionally. The person and number of the patient are head-marked when A is plural and when one classifier is in future and present tenses.
11. Ika (P), Qiang, and Sko: when arguments could be confused (e.g., non-canonical WO), either is dependent-marked, optionally elsewhere.
12. Kisi: A is dependent-marked with a subject noun class pronoun, optionally for animate singular. P is optionally dependent-marked with object noun class pronouns. There is also tonal downstep between the verb and the following direct object.

6. Note that our definition of functional load is hearer-oriented: homophonous forms violate the principle of distinctiveness.

13. Korku: animate patients are dependent-marked, inanimate optionally.
14. Lavukaleve: the person and number of the agent are head-marked in present tense and in the habitual aspect, optionally elsewhere.
15. Somali: a non-focused P is dependent-marked with a tone pattern distinct from that of A.
16. Warao: the person and number of the agent are head-marked when aspect is not marked or when tense is either implied or marked.
17. Warndarang: the person (and number) of the patient is head-marked when the number of either or both arguments is plural.
18. Yelî Dnye: the person (and partially number) of the agent is head-marked for continuous aspect and partially for punctiliar aspect.
19. Yimas: the person and number of the arguments are obligatorily head-marked when they are new referents but optionally when established.

Two problematic cases not treated in the lists are discussed here. In Daga, an animate P is head-marked when third person singular but inanimate patients are only optionally marked as totally affected (Murane 1974: 44). The problem is how to assess the FL of such a heterogeneous marking. Contrasting this FL with one that has already been analysed might help. Obligatory head marking in Daga is more limited than e.g., obligatory dependent marking in Kannada. Optional head marking in Daga, on the other hand, is impressionistically rather similar in scope with optional dependent marking in Kannada. Since FL of dependent marking in Kannada was analysed as extensive, FL of patient head-marking cannot be analysed as extensive but rather as marginal. In Maricopa the number of the patient is head-marked by suppletive stems for some verbs and by an optional plural prefix (Gordon 1986: 22, 100). Since the source did not specify what type of verbs had suppletive verb forms for number, I interpreted "some" literally and classified FL as marginal.

For the purposes of the statistical tests, the four FL values – none, marginal, extensive, full – were converted into ordinal values 1, 2, 3, and 4, respectively (Table 3).

Table 3. The data.

Language	**WO_A/P**	**DM_A/P**	**HM_A**	**HM_P**
Adang	none	marg	none	marg
Arapesh	full	none	full	ext
Babungo	none	none	none	none
Berbice Dutch Creole	full	none	none	none
Cora	ext	none	full	full
Daga	none	none	full	marg
Diyari	none	full	none	none
Georgian	none	full	full	none
Gooniyandi	none	ext	full	ext

(*Continued*)

Table 3. Continued.

Language	WO_A/P	DM_A/P	HM_A	HM_P
Hixkaryana	full	none	full	full
Hungarian	none	full	full	full
Ika	none	ext	marg	full
Indonesian	full	none	full	none
Jaqaru	none	full	full	full
Kannada	marg	ext	full	none
Khoekhoe	none	full	none	none
Kisi	full	ext	none	none
Koasati	none	full	full	none
Kombai	full	none	full	none
Korku	none	ext	none	marg
Kuot	ext	none	full	full
Lavukaleve	full	none	ext	full
Maricopa	none	full	full	marg
Maybrat	full	none	full	none
Middle Atlas Berber	full	full	full	none
Mien	full	none	none	none
Ngiti	full	none	full	none
Dongolese Nubian	none	full	full	none
Nuuchahnulth	none	none	full	none
Osage	full	full	full	full
Pirahã	full	none	none	none
Plains Cree	none	none	full	full
Qiang	none	ext	full	full
Quechua	none	full	full	none
Shipibo-Konibo	none	full	marg	none
Sko	full	ext	full	marg
Slave	ext	none	marg	marg
Somali	none	ext	full	none
Southern Sierra Miwok	none	full	full	none
Thai	full	none	none	none
Trumai	ext	full	none	none
Tzutujil	none	none	full	full
Urubú-Kaapor	none	marg	full	none
Warao	none	none	ext	none
Warndarang	full	none	full	ext
Welsh	full	full	full	none
West Greenlandic	marg	ext	full	full
Yagua	ext	marg	full	none
Yelî Dnye	full	full	ext	full
Yimas	none	none	ext	ext

7. Results and discussion

According to the hypothesis, statistically significant trade-offs should exist between the different variables involved in core argument marking. Word order (WO), dependent marking (DM) and the head-marking of agent (HM_A) and patient (HM_P) were taken as variables as such. Other variables were the "normalized" head marking (HM_N) and morphological marking (the sum of dependent marking (DM) and head marking (HM_N)). These five variables combine into nine pairs (Table 4). HM_N means that the functional load of head marking is normalized to the lowest value of head marking of A/P in each language, roughly measuring the degree of differential argument marking by HM (or "rich agreement"). For instance, the FL of HM_A e.g., in Maricopa is 3 and that of HM_P is 1, but the FL of normalized HM is 1.

To test the hypothesis, I used Kendall's (tau) nonparametric correlation test in the open-source statistical computing environment R (R Development Core Team 2006). Kendall's tau measures the difference between the probabilities that two variables are in the same or different order. If Kendall's tau was statistically significant for a particular pair of variables, the Chi-Square test and its tabulated data was used to double-check the result and help in the interpretation. For the Chi-Square test, the values of the coding strategies were classified into two groups: a) those with no FL and b) those with non-zero FL, corresponding straightforwardly to "simple" and "complex", respectively.

Table 4. The Kendall's tau values for each pair of variables.

Pair of variables	tau	p
WO–DM	–0.302	0.017
WO–HM_A	–0.035	0.787
WO–HM_P	–0.038	0.768
WO–HM_N	–0.005	0.967
WO–M	–0.249	0.041
DM–HM_A	0.066	0.606
DM–HM_P	–0.094	0.452
DM–HM_N	–0.114	0.365
HM_A–HM_P	0.201	0.118

Table 4 summarizes the results for Kendall's tau. The values were statistically significant only for two correlation pairs: those between WO and morphological marking and WO and DM. The correlation between head marking of A and P was strong enough to validate closer inspection. For all the other pairs Kendall's tau was less than ± 0.12 and far from being statistically significant. In fact some values (especially that between WO and HM_N) were so close to zero as to render the trade-off hypothesis vacuous: almost no interaction occurred between these variables. We may still observe that the sign of correlation conformed to the hypothesis for all non-significant pairs: WO had a very small negative correlation with HM_A, HM_P, and HM_N, whereas DM correlated positively with head marking of A but negatively with HM_ P and HM_N.

The variables representing the different coding domains (WO and morphological marking) correlated negatively in a statistically significant way (tau = –0.249, $p < 0.05$). The Chi-Square value was not statistically significant but still close to it ($p = 0.088$). According to Table 5, eight sample languages relied exclusively on word order and three lacked full FL for any coding strategy. Two of these, Warao and Nuuchahnulth, had extensive and full head marking of A, respectively, but none for P. Babungo, on the contrary, had no head or dependent marking (a floating tone might serve as a dependent marker, but there was not enough data to confirm this). It also allows for a reversible word order pair AVP/PVA: the canonical word order is AVP, but the reverse PVA may be used to topicalize the P. Thus, only context tells when the sentence is AVP or PVA (8):

Babungo (Bantoid, Schaub 1985: 141, 250 n1)

(8) *Wèe wî jìa Làmbí*
child that hold-PFV Lambi
'As for that child, Lambi held him.'
or 'That child held Lambi.'

It seems almost a norm for languages to have at least some morphological marking of the core arguments. The absence of morphological marking seems a good indicator that WO has some FL but the presence of morphological marking indicates only weakly that WO has a zero FL. The overall conclusion is that the smaller the FL of morphological marking is the more likely WO carries a greater functional load in core argument marking. This result conforms well to the trade-off hypothesis.

The correlation between WO and DM was –0.302 ($p < 0.05$). The Chi-Square value was also statistically significant ($p < 0.05$). Interpreting from Table 5, the absence of DM seems a good indicator of the presence of WO, and the absence of WO seems a good indicator of the presence of DM. It is probably safe to conclude that the smaller the functional load of DM, the more likely WO has a greater functional load, and the greater the functional load of DM, the less likely WO has a smaller (rather than greater) functional load. A strong inverse dependency occurs between these variables. It is probably no coincidence they have often been seen as complementary devices in argument marking (e.g., Vennemann 1973). Since the correlation of WO and HM was so small, the correlation between WO and DM explains why WO correlated also with overall morphological marking.

Table 5. Chi-square values for a few pairs of variables.

	M–	M+	DM–	DM+
WO–	3	22	7	18
WO+	8	17	15	10
	χ^2 =2.91, p = 0.088		χ^2 = 5.19, p = 0.023	
	HM_A–	**HM_A+**	**WO/DM–**	**WO/DM+**
HM_P–	9	18	5	22
HM_P+	2	21	12	11
	χ^2 = 4.39, p = 0.036		χ^2 = 6.27, p = 0.012	

Since the data contained many tied values, the reliability of Kendall's tau was further studied with a resampling procedure called bootstrap.[7] This procedure was performed for the statistically significant pairs WO and DM and WO and morphological marking. Languages were randomly sampled with replacement 10,000 times and the value of Kendall's tau was computed each time. Confidence limit for 0.95 was computed by extracting the 250th and 9750th resampled tau values. The confidence limits at level 0.95 were [–0.249 and 0.250] for WO and DM and [–0.244 and 0.246] for WO and morphological marking thus confirming the confidences of the original values (–0.302 and –0.249).

Although the correlation between the head-marking of A and P (tau = 0.201) was not statistically significant, the Chi-Square value was ($p < 0.05$). The presence of HM_P seems a good indicator of the presence of HM_A, and the absence of HM_A seems a good indicator of the absence of HM_P. This positive dependency between the variables provides evidence for dependencies between the head-marking strategies. The result reflects the agreement hierarchy (subject < object), which tells that the presence of object agreement implies the presence of subject agreement but subject agreement can exist without object agreement (Croft 1990: 101–107)

The trade-off hypothesis predicts that HM should correlate strongly with both WO and DM. Notably neither variable had a significant relationship with the head-marking of A or P or their neutralized sum. The strong dependency between WO and DM might suggest languages prefer distinguishing the core arguments via either of them and rely on HM only when they are insufficient. Accordingly, the trade-off hypothesis was adjusted to expect HM when core arguments were not fully distinguished by WO and DM. Since HM_A is so frequent cross-linguistically (occurred in 39 sample languages), HM_P was expected instead (note that it often implies the presence of HM_A as well). This adjusted prediction was borne out by the results (Table 5): when WO and

7. I am grateful to Stefan Werner for advice with the bootstrap method.

DM fully distinguish the core arguments (WO/DM+, i.e., their combined FL is 3 or more), HM_P is less likely to occur than not, and when WO and DM cannot fully distinguish the core arguments (WO/DM-) HM_P is more likely to occur than not. The Chi-Square value ($\chi^2 = 6.27$) for this inverse dependency was statistically significant ($p < 0.05$). HM_A seems to occur in languages regardless of its "dispensability" in core argument marking, but HM_P does so less often.

Finally, a few words should be said about how languages overall conformed to the principles of economy and distinctiveness. The functional loads of the coding strategies (WO, DM, and HM_N) were added up in each language. According to the principle of distinctiveness, languages should distinguish core arguments in all contexts. In other words, the sum of the FLs should not be less than 3. In nine sample languages the overall FL was less than this, violating the principle of distinctiveness. The most notable exceptions were Warao, Nuuchahnulth, and Babungo (already discussed above). In 22 languages, the overall functional load was exactly 3, adhering perfectly to the two principles. The remaining 19 languages (38 %) violated the principle of economy to varying degrees, most notably Osage and Yelî Dnye, which exhibit at least extensive FL for all coding strategies.[8] Overall, adherence to distinctiveness seems stronger than adherence to economy, which is another way of saying that languages allow for redundancy more than ambiguity.

8. Conclusions

The claim that languages trade off complexity in one area with simplicity in another has been criticised in the recent discussion on language complexity (McWhorter 2001; Kusters 2003; Shosted 2006). This study is one of the first attempts at investigating the claim with morphosyntactic variables and with a large balanced sample. The results justify rejecting trade-offs as an all-encompassing principle in languages: most of the correlations were small or even approaching zero, indicating no relationship between the variables.

However, the results support the trade-off hypothesis in a weaker form in that trade-offs occur in some subdomains: word order correlated in a statistically significant way with morphological marking and dependent marking. The statistically significant positive dependency between the head-marking of agent and patient further conforms to the predictions of the hypothesis. So does the sign of correlation (+/–), which was as predicted for all pairs of variables. Lastly, the presence of the head-marking of P depended in an inverse and statistically significant way on whether word order and dependent marking fully distinguished the core arguments. Overall, it would be fair to

8. I am a bit sceptical about the analysis of Yelî Dnye and Osage, since the sources did not cover word order phenomena in great detail.

say that languages manifest at least some trade-offs in core argument marking, most notably that between WO and DM.

The results agree with the work of Fenk-Oczlon and Fenk (this volume, and the references therein). They have established several negative correlations between e.g., the "complexities" of syllable structure, word structure, and clause structure counted in terms of numbers of phonemes, syllables and words. The results disagree with Shosted (2006), who found no evidence for trade-offs in a cross-linguistic study on the complexities between syllable structure and inflectional synthesis on the verb. However, this discrepancy in our results might stem from the choice of parameters: I intentionally chose variables that were functionally connected whereas the parameters studied by Shosted (2006) were functionally rather dissimilar. Trade-offs seem limited to functionally connected variables but even then they are not an all-encompassing principle.

It is not at all clear that equal complexity of languages could be shown by a synchronic snapshot on trade-offs alone. In addition to trade-offs, one should show for instance that simplifying and complexifying tendencies are equally frequent in language change. In other words, languages should follow the principles of economy and distinctiveness to the same degree both synchronically and diachronically. What the present study on trade-offs has hopefully shown is to what degree languages adhere to the principles of economy and distinctiveness synchronically in marking the core arguments. According to the results, most languages conform to the principle of one-meaning–one-form, but those that do not conform to it adhere to distinctiveness at the expense of economy. This may reflect the tendency in languages to accumulate complexity during their long histories, i.e., that they exhibit fewer simplifying than complexifying tendencies (McWhorter 2001; Dahl 2004), but it may as well reflect the tendency in small, isolated, and tight-knit communities to preserve complexity, i.e. that they exhibit fewer simplifying tendencies than languages spoken by large communities with loose social networks and adult language learning by outsiders (Kusters 2003: 5–9).

Abbreviations

1	first person
3	third person
4	fourth person (only in Slave)
A	agent
ABS	absolutive
ACC	accusative
CLF	classifier
COP	copula
CSM	change of state marker
DECL	declarative
DIR	directional prefix

GEN	genitive
N	neuter
NOM	nominative
OBJ	object marker
P	patient
PST	past tense
PFV	perfective
PRS	present tense
SG	singular

References

Armbruster, C.H. 1960. *Dongolese Nubian: A Grammar*. Cambridge: CUP.
Austin, P. 1981. *A Grammar of Diyari, South Australia*. Cambridge: CUP.
Broadbent, S.M. 1964. *The Southern Sierra Miwok Language*. Berkeley CA: University of California Press.
Casad, E.H. 1984. Cora. In *Studies in Uto-Aztecan Grammar*, Vol. 4: *Southern Uto-Aztecan Grammatical Sketches*, R.W. Langacker (ed.), 153–459. Dallas TX: Summer Institute of Linguistics.
Childs, G.T. 1995. *A Grammar of Kisi: A Southern Atlantic Language*. Berlin: Mouton de Gruyter.
Cole, P. 1985. *Imbabura Quechua*. London: Croom Helm.
Comrie, B. 1989. *Language Universals and Linguistic Typology*. 2nd edn. Oxford: Blackwell.
Comrie, B. 2005. Alignment of case marking. In *The World Atlas of Language Structures*, M. Haspelmath, M. Dryer, D. Gil & B. Comrie (eds), 398–405. Oxford: OUP.
Conrad, R.J. & Wogiga, K. 1991. *An Outline of Bukiyip grammar*. Canberra: Australian National University.
Court, C. 1985. *Fundamentals of Iu Mien (Yao) Grammar*. PhD Dissertation, University of California at Berkeley.
Croft, W. 1990. *Typology and Universals*. Cambridge: CUP.
Crystal, D. 1997. *The Cambridge Encyclopedia of Language*. 2nd edn. Cambridge: CUP.
Dahl, Ö. 2004. *The Growth and Maintenance of Linguistic Complexity*. Amsterdam: John Benjamins.
Dahlstrom, A. 1991. *Plains Cree Morphosyntax*. New York NY: Garland Publishing.
Dayley, J. 1985. *Tzutujil Grammar*. Berkeley CA: University of California Press.
Derbyshire, D.C. 1979. *Hixkaryana*. Amsterdam: North-Holland.
Dol, P. 1999. *A Grammar of Maybrat: A Language of the Bird's Head, Irian Jaya*. PhD Dissertation, University of Leiden.
Donohue, M. 2004. Ms. A Grammar of the Skou Language of New Guinea. National University of Singapore. [http://www.papuaweb. org/dlib/tema/bahasa/skou/ (November 13, 2006)].
Dryer, M. 1992. The Greenbergian word order correlations. *Language* 68: 81–138.
Dryer, M. 2005. Genealogical language list. In *The World Atlas of Language Structures*, M. Haspelmath, M. Dryer, D. Gil & B. Comrie (eds), 584–644. Oxford: OUP.
Du Bois, J. 1987. The discourse basis of ergativity. *Language* 63: 805–855.

Everett, D.L. 1986. Pirahã. In *Handbook of Amazonian Languages,* Vol. 1, D.C. Derbyshire & G.K. Pullum (eds), 200–325. Berlin: Mouton de Gruyter.
Foley, W.A. 1991. *The Yimas Language of New Guinea*. Stanford CA: Stanford University Press.
Fortescue, M. 1984. *West Greenlandic*. London: Croom Helm.
Frank, P. 1990. *Ika Syntax*. Arlington TX: Summer Institute of Linguistics and the University of Texas at Arlington.
Givón, T. 1981. Typology and functional domains. *Studies in Language* 5: 163–193.
Gordon, L. 1986. *Maricopa Morphology and Syntax*. Berkeley CA: University of California Press.
Guirardello, R. 1999. *A Reference Grammar of Trumai*. PhD Dissertation, Rice University, Houston, Texas.
Haan, J.W. 2001. *The Grammar of Adang: A Papuan Language Spoken on the Island of Alor, East Nusa Tenggara – Indonesia*. PhD Dissertation, University of Sidney. [http://www-personal.arts. usyd.edu.au/jansimps/haan/adang-index.htm (November 13, 2006)].
Hagman, R.S. 1977. *Nama Hottentot Grammar*. Bloomington IN: Indiana University Press.
Hardman, M.J. 2000. *Jaqaru*. München: Lincom Europa.
Harris, A. 1981. *Georgian Syntax: A Study in Relational Grammar*. Cambridge: CUP.
Haspelmath, M. 2006. Against markedness (and what to replace it with). *Journal of Linguistics* 42: 25–70.
Hawkins, J.A. 2004. *Efficiency and Complexity in Grammars*. Oxford: OUP.
Heath, J. 1980. *Basic Materials in Warndarang: Grammar, Texts and Dictionary*. Canberra: Australian National University.
Henderson, J. 1995. *Phonology and Grammar of Yele, Papua New Guinea*. Canberra: Australian National University.
Iwasaki, S. & Ingkaphirom, P. 2005. *A Reference Grammar of Thai*. Cambridge: CUP.
Kakumasu, J. 1986. Urubu-Kaapor. In *Handbook of Amazonian Languages*, Vol. 1, D.C. Derbyshire and G.K. Pullum (eds), 326–403. Berlin: Mouton de Gruyter.
Kimball, G.D. 1991. *Koasati Grammar*. Lincoln NB: University of Nebraska Press.
King, G. 1993. *Modern Welsh: A Comprehensive Grammar*. London: Routledge.
Kittilä, S. 2005. Optional marking of arguments. *Language Sciences* 27: 483–514.
Kouwenberg, S. 1994. *A Grammar of Berbice Dutch Creole*. Berlin: Mouton de Gruyter.
Kusters, W. 2003. *Linguistic Complexity: The Influence of Social Change on Verbal Inflection*. PhD Dissertation, University of Leiden.
Kutsch Lojenga, C. 1994. *Ngiti: A Central-Sudanic Language of Zaire*. Köln: Rüdiger Köppe Verlag.
LaPolla, R.J. with Huang, C. 2003. *A Grammar of Qiang*. Berlin: Mouton de Gruyter.
Lindström, E. 2002. *Topics in the Grammar of Kuot, a Non-Austronesian Language of New Ireland, Papua New Guinea*. PhD Dissertation, Stockholm University.
Maddieson, I. 2006. Correlating phonological complexity: Data and validation. *Linguistic Typology* 10: 106–123.
McGregor, W. 1990. *A Functional Grammar of Gooniyandi*. Amsterdam: John Benjamins.
McWhorter, J. 2001. The world's simplest grammars are creole grammars. *Linguistic Typology* 5: 125–156.
Miestamo, M. 2005. *Standard Negation: The Negation of Declarative Verbal Main Clauses in a Typological Perspective*. Berlin: Mouton de Gruyter.
Miestamo, M. 2007. Symmetric and asymmetric encoding of functional domains, with remarks on typological markedness. In *New Challenges in Typology: Broadening the Horizons and Redefining the Foundations*, M. Miestamo & B. Wälchli (eds), 293–314. Berlin: Mouton de Gruyter.

Murane, E. 1974. *Daga Grammar, from Morpheme to Discourse*. Norman OK: SIL.
Næss, Å. 2004. What markedness marks: The markedness problem with direct objects. *Lingua* 114: 1186–1212.
Nagaraja, K.S. 1999. *Korku Language: Grammar, Texts, Vocabulary*. Tokyo: Tokyo University of Foreign Studies.
Nakayama, T. 2001. *Nuuchahnulth (Nootka) Morphosyntax*. Berkeley CA: University of California Press.
Nichols, J. 1992. *Linguistic Diversity in Space and Time*. Chicago IL: The University of Chicago Press.
Payne, D.L. & Payne, T.E. 1990. Yagua. In *Handbook of Amazonian Languages*, Vol. 2, D.C. Derbyshire & G.K. Pullum (eds), 249–474. Berlin: Mouton de Gruyter.
Penchoen, T. 1973. *Tamazight of the Ayt Ndhir*. Los Angeles CA: Undena Publications.
Quintero, C. 2005. *Osage Grammar*. Lincoln NB: University of Nebraska Press.
R Development Core Team 2006. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. [http://www.R-project.org].
Rice, K. 1989. *A Grammar of Slave*. Berlin: Mouton de Gruyter.
Romero-Figueroa, A. 1997. *A Reference Grammar of Warao*. München: Lincom.
Rounds, C. 2001. *Hungarian: An Essential Grammar*. London: Routledge.
Saeed, J.I. 1999. *Somali*. Amsterdam: John Benjamins.
Schaub, W. 1985. *Babungo*. London: Croom Helm.
Shosted, R.K. 2006. Correlating complexity: A typological approach. *Linguistic Typology* 10: 1–40.
Sneddon, J.N. 1996. *Indonesian: A Comprehensive Grammar*. London: Routledge.
Sridhar, S.N. 1990. *Kannada*. London: Routledge.
Terrill, A. 2003. *A Grammar of Lavukaleve*. Berlin: Mouton de Gruyter.
Valenzuela, P.M. 1997. *Basic Verb Types and Argument Structures in Shipibo-Konibo*. MA Thesis, University of Oregon.
Vennemann, T. 1973. Explanation in syntax. In *Syntax and Semantics*, Vol. 2, J. Kimball (ed.), 1–50. New York NY: Academic Press.
Vries, L. de 1993. *Forms and Functions in Kombai, an Awyu Language of Irian Jaya*. Canberra: Australian National University.