Coordination and Processing

Scott Drellishak

sfd@u.washington.edu

1 Introduction

One of the ongoing research aims of modern linguistics is accounting for the range of possible phenomena in human language. In particular, generative grammarians working in Chomsky's Principles and Parameters framework have often sought to explain typological generalizations by positing principles of Universal Grammar (UG) that require them, or parameters that can take two or more values that correspond to observed language variation. Such a research program has an obvious appeal: if typological generalizations can be explained by a small set of broad, simple principles and parameters, then linguists will have gone a long way towards characterizing the precise contents of the human language faculty.

This research program, however, has tended to overlook an alternative source of explanations that can often account for typological generalizations, namely limitations of the human language processor. John Hawkins (2004) has proposed a set of processing principles that are intended to account both for preferences within languages for certain kinds of constructions, and for the distribution of typological features across languages. These processing principles offer a way of accounting for statistical universals — that is, "universals" that hold less than one hundred percent of the time — because they assert a preference for more-easily-processed structures without ruling out the alternatives. If the processing principles can account for language variation, a theory based on them is a clear improvement over a theory that requires the positing of new principles or parameters to account for newly discovered phenomena.

In this paper, I will apply Hawkins' processing principles in an attempt to account for an apparent universal in coordination strategies in the world's languages. The second

 University of Washington Working Papers in Linguistics vol. 24 (2005), eds. Daniel J. Jinguji and Steven Moran, pp 1-14, Seattle, WA.

section is a description of the universal in question. The third section briefly lists and explains the processing principles of Hawkins (2004). The fourth section examines the observed universal in light of the processing principles and attempts to determine whether it can be accounted for in terms of them. The fifth section discusses the findings of this attempt and proposes possible explanations for what is found.

2 The Universal

Stassen (2000) contains a survey of noun phrase coordination in a genetically diverse sample containing 270 of the world's languages. As indicated by its title, Stassen's article broadly divides the world's languages into two groups: AND-languages, in which NP coordination is accomplished with a syntactically balanced structure (similar to those marked in English and the other Indo-European languages by *and* and its cognates), which he calls the *Coordinate Strategy*; and WITH-languages, in which NP-coordination structures are imbalanced, with one of the two coordinands marked in a way that carries comitative meaning, which he calls the *Comitative Strategy*. This paper does not focus on WITH-languages, but rather on a universal Stassen observed in the AND-languages in his sample.

Because the domain of Stassen's survey included only the coordination of two items, there are a finite number of possible morpheme orders in AND-language marking strategies. These strategies can be categorized in two ways: by the number of marked coordinands, and by the position of the marking morpheme. In some strategies, coordinated items are simply juxtaposed without marking (that is, the number of marked coordinands is zero); this is referred to as *asyndeton*. In other strategies, there is a single marking morpheme for the entire coordinated phrase; this is referred to as *monosyndeton*. In still other strategies, one marking morpheme appears for each coordinand; this is referred to as *polysyndeton*. As for position of the coordinating morphemes, they can either precede or follow each of the coordinands. Among the strategies Stassen found in his survey, various possibilities were attested. These included the very common medial monosyndeton, as found in Finnish (Uralic, Balto-Finnic[1]):

[1] The language classifications included here are Stassen's.

(1) Pentti ja Pirkko
Pentti and Pirkko
'Pentti and Pirkko' (Stassen 2000:11)

Stassen also found examples of final monosyndeton, as in Pitjantjatjara (Australian, Pama-Nyungan):

(2) Henry-ku mama ngunytju puru
Henry-GEN mother father and
'Henry's father and mother' (Stassen 2000:15)

Among polysyndeton strategies, Stassen found examples of languages in which the mark followed the coordinands and, more rarely, examples in which it preceded them. The former pattern can be seen in Abkhaz (North-West Caucasian):

(3) s-ànə-y s-àbə-y
my-mother-and my-father-and
'my mother and my father' (Stassen 2000:12)

The latter can be seen in a strategy in Sedang (Mon-Khmer) that marks coordinands with dual pronouns:

(4) préi klá préi koa
3DU tiger 3DU turtle
'the tiger and the turtle' (Stassen 2000:17)

In spite of this variety, in all the languages in his survey, Stassen failed to find any occurrences among the AND-languages of the remaining coordination marking pattern: initial monosyndeton. As he puts it:

> To round off the discussion of the various manifestations of the Coordinate Strategy, I can note that monosyndetic preposing on the first NP is not attested at all in the

> sample. That is, there do not seem to be languages which conform to the AND-NP NP scheme. (Stassen 2000:15)

Given that there exist languages that exhibit the NP NP-AND strategy, it is curious that no languages exhibit a strategy with the opposite order, especially since that is just the sort of variation commonly observed between head-initial and head-final languages. Why should this typological asymmetry exist? Because the generalization is apparently exceptionless, we might be tempted to assert the existence of a universal principle to account for it; however, such a universal would have to be phrased in such a way that its existence seems improbable. First, rather than broadly applying to any type of construction, the principle would have to address coordination and nothing else. On top of that, the supposedly universal principle would be most economically phrased as a *negative* universal, so that it would specifically rule out initial monosyndeton, but allow medial and final monosyndeton, preposed and postposed polysyndeton, and asyndeton. This would leave us with a principle proposed to be universal, but phrased so narrowly that it would only apply to a particular kind of AND-coordination and to no other construction. Proposing such a principle has little or no explanatory power, because it makes no claims outside the narrow domain of coordination strategies; instead, we should look for other explanations, such as one provided by Hawkins' processing principles.

3 Hawkins's Processing Principles

John Hawkins (2004) proposes a set of performance principles in order to explain the distribution of various structures in human language. The central idea underlying these principles is his Performance-Grammar Correspondence Hypothesis (PGCH):

> Grammars have conventionalized syntactic structures in proportion to their degree of preference in performance, as evidenced by patterns of selection in corpora and by ease of processing in psycholinguistic experiments. (Hawkins 2004:3)

According to this hypothesis, a structure that is preferred according to performance criteria should be more common not only within a single language but also cross-linguistically. However, for the purposes of this paper it is important to note that the

converse of this implication does not necessarily hold: alternatives that occur more often in a corpus or across the world's languages need not have a preference in performance, but might have some other cause (common origin, accident of history, etc.). In order to determine whether Stassen's coordination universal is explainable by performance factors, then, we need to evaluate it according to the principles Hawkins spells out.

The first of these principles is Minimize Domains (MiD), which Hawkins defines as follows:

> The human processor prefers to minimize the connected sequences of linguistic forms and their conventionally associated syntactic and semantic properties in which relations of combination and/or dependency are processed. The degree of this preference is proportional to the number of relations whose domains can be minimized in competing sequences of structures, and to the extent of the minimization difference in each domain. (Hawkins 2004:31)

A simple example of this principle in operation is in the preference for short prepositional phrase adjuncts before long ones in English. Consider examples (5) and (6) from Hawkins (2004:104):

(5) The man $_{VP}$[waited $_{PP1}$[for his son] $_{PP2}$[in the cold but not unpleasant wind]]

(6) The man $_{VP}$[waited $_{PP2}$[in the cold but not unpleasant wind] $_{PP1}$[for his son]]

Sentence (5) is preferred to (6) because the domain of the VP (i.e. the range of lexical items that must be processed in order to recognize it; see §4.1 below for a more formal definition) contains five words, from *waited* to *in*, in sentence (5), but nine words, from *waited* to *for*, in sentence (6). According to MiD, the sentence with the smaller domain is preferred.

The second of Hawkins' principles is Minimize Forms (MiF), which is based on the straightforward idea that it is easier to process less material than to process more. Hawkins' formal definition is as follows:

> The human processor prefers to minimize the formal complexity of each linguistic form F (its phoneme, morpheme, word, or phrasal units) and the number of forms with unique conventionalized property assignments, thereby assigning more properties to fewer forms. These minimizations apply in proportion to the ease with which a given property P can be assigned in processing to a given F. (Hawkins 2004:38)

MiF prefers structures with less material to those with more. For example, in a sentence in which the grammatical role of a given NP can be recognized by its position, MiF would prefer no marking to the presence of a case-marking morpheme.

The third, and most complex, of the performance principles proposed by Hawkins is Maximize On-line Processing (MaOP), which he defines as follows:

> The human processor prefers to maximize the set of properties that are assignable to each item X as X is processed, thereby increasing O(n-line) P(roperty) to U(ltimate) P(roperty) ratios. The maximization difference between competing orders and structures will be a function of the number of properties that are unassigned or misassigned to X in a structure/sequence S, compared to the number in an alternative. (Hawkins 2004:51)

Because the name of this principle is perhaps less transparent than those of the other two, it requires a bit more explanation. Hawkins' idea is that, as a sentence is being processed, various properties are being assigned to the items in the sentence. When the sentence is finished, some total number of properties has been assigned. Depending on the facts of the language, at some points during processing properties can be assigned immediately, but at other points this assignment is delayed. Consider two hypothetical SOV languages, one in which subject NPs are marked in some way, and another in which they are not. In the first language, an initial NP can be identified as the subject immediately after it occurs, while in the second, it cannot be identified until later—it might turn out to be the object of the verb in an optional-subject language, for example. According to MaOP, structures that maximize the ratio of the number of properties assignable during processing to the final number of properties are preferred. This

principle formalizes the intuition that it is easier to process sentences that do not contain ambiguous forms or garden paths.

4 Coordination and Processing

Now that we have Stassen's observed language universal and Hawkins' processing principles for evaluating universals, it remains to analyze the universal in light of each of the three principles. However, first a description of the structures that will be analyzed and a statement of some assumptions are necessary. The following sections contain a comparison of two coordination structures: initial monosyndeton, which takes the form AND NP* (i.e. a single coordinator followed by any number of noun phrases); and final monosyndeton, which takes the form NP* AND (i.e. any number of noun phrases followed by a coordinator). Other possible locations for the coordinator, including the common NP* AND NP structure, generally fall between the two peripherally marked strategies according to the processing principles. To deal with any variation arising from basic word order, the processing principles are tested on verb phrases from two hypothetical languages: one language whose basic word order is OV and another that is VO, where in each case the O is a coordinated NP. (Subjects are omitted to avoid the possibility of the subject NP being confused for a part of the object in some cases.) For each principle, therefore, four utterances will be considered:

(7) V NP NP AND (VO, final monosyndeton)

(8) V AND NP NP (VO, initial monosyndeton)

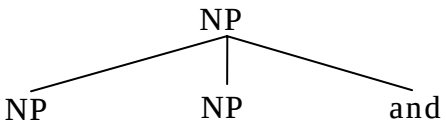
(9) NP NP AND V (OV, final monosyndeton)

(10) AND NP NP V (OV, initial monosyndeton)

For simplicity, I will initially assume that the NPs are all single words, although that obviously need not be the case. In addition, where necessary I will also assume a phrase structure for coordinated phrases that is flat and in which the dominating node has the

same category as its coordinands. (11) shows an example of such a structure for two NPs with a final monosyndeton coordinator:

(11)



It is important to note that this structure violates two commonly held assumptions of many syntactic theories, namely X-bar structure and binary branching, and also that it does not assume the existence of a CoordP or &P maximal projection. It is assumed here, however, because it is in keeping with the sorts of phrase structures that appear in Hawkins (2004) (such as (12) below, in which the 0 represents a gap site), and a complete reassessment of his ideas using a different theory of syntax is beyond the scope of this paper.

(12) [NP VP[V O]] and [NP VP[V NP]] (Hawkins 2004:94)

4.1 Minimize Domains

In order to apply MiD, we first need an understanding of what a domain is. Hawkins (2004) describes them in some detail, but for the purposes of this paper, this definition will suffice: the domain of a node M is "the smallest set of terminal and non-terminal nodes that must be parsed in order to recognize M" (Hawkins 2004:32).

Given this definition, we can apply the principle MiD to the example sentences. In each case, there are two domains to be considered: the domain of the coordinated phrase and the domain of the verb phrase (i.e. the verb and its object). Let us first consider the domain of the coordinated NP, whose extent is controlled by the recognition of the coordinated phrase and all its coordinands. In the examples in which the coordinator is initial, (8) and (10), the domain extends from the coordinator to the last item it coordinates—that is, across the entire coordinated phrase. In the examples in which the coordinator is final, (7) and (9), this remains true: the whole coordinated phrase cannot be recognized until the final coordinator is seen. Next, let us consider the domain of the verb phrase, whose extent is controlled by the recognition of the verb and all its arguments. In examples (7) and (10), in which the coordinator is on the opposite

side of the verb phrase from the verb, the domain of the VP covers all four words, because the listener cannot be said to have recognized it until hearing either the coordinator in (7) or the verb in (10). Examples (8) and (9) are slightly more problematic because it is not immediately clear whether to consider the domain to extend across the whole object to the farthest coordinated NP, or just to the coordinator. Hawkins' analysis of constituent-order preferences in head-initial and head-final languages (2004:104-111) gives us some guidance. For both word orders, he assumes that the domain for the recognition of a VP that contains a verb and an argument PP extends from the verb to the adposition, and not across the whole PP. Extending this assumption to coordination implies that the domain of the VP in (8) and (9) includes just the verb and the adjacent coordinator.

All domains, therefore, are the same size in each pair of examples that contrasts coordinator order. Accordingly, the processing principle Minimize Domains is indifferent to the difference between initial and final monosyndeton.

4.2 Minimize Forms

Applying MiF is even more straightforward. In all four examples, the amount of material is the same, and so MiF, like MiD, is indifferent between initial and final monosyndeton. In addition, it is interesting to note that according to Minimize Forms, asyndeton coordination is the most preferred strategy, since it has one less morpheme than monosyndeton coordination.

4.3 Maximize On-line Processing

Unlike the other two principles, MaOP prefers one of the coordination orders over the other. To show this, we need to consider the processing of each example sentence word by word, keeping track of which final properties have been assigned in order to calculate the On-line Property to Ultimate Property ratio (OP/UP). The ratio for each of the four examples is calculated in the following tables, using the notation of Hawkins (2004:56). Each column shows the properties (Categories, Phrases, Attachments, and Relations) that have been assigned as the sentence is recognized sequentially, as well as the current number of assigned relations and the current OP/UP ratio. NP_1 and NP_2 are used for the coordinand noun phrases, and NP_c for the coordinated NP.

Table 1: OP/UP Ratios

Example (7)	V	NP	NP	AND
Categories	V	NP	NP	Coord
Phrases	VP	NP_1	NP_2	NP_c
Attachments	VP[S]			NP_c[VP], NP_1[NP_c], NP_2[NP_c]
Relations				NP_c=OBJ-V
# Assigned	3	5	7	13
OP/UP	3/13 = 23%	5/13 = 38%	7/13 = 54%	13/13 = 100%

Example (8)	V	AND	NP	NP
Categories	V	Coord	NP	NP
Phrases	VP	NP_c	NP_1	NP_2
Attachments	VP[S]	NP_c[VP]	NP_1[NP_c]	NP_2[NP_c]
Relations		NP_c=OBJ-V		
# Assigned	3	7	10	13
OP/UP	3/13 = 23%	7/13 = 54%	10/13 = 77%	13/13 = 100%

Example (9)	NP	NP	AND	V
Categories	NP	NP	Coord	V
Phrases	NP_1	NP_2	NP_c	VP
Attachments			NP_1[NP_c], NP_2[NP_c]	NP_c[VP], VP[S]
Relations				NP_c=OBJ-V
# Assigned	2	4	8	13
OP/UP	2/13 = 15%	4/13 = 31%	8/13 = 62%	13/13 = 100%

Example (10)	AND	NP	NP	V
Categories	Coord	NP	NP	V
Phrases	NP_c	NP_1	NP_2	NP_c
Attachments		NP_1[NP_c]	NP_2[NP_c]	NP_c[VP], VP[S]
Relations				NP_c=OBJ-V
# Assigned	2	5	8	13
OP/UP	2/13 = 15%	5/13 = 38%	8/13 = 62%	13/13 = 100%

There are several caveats about these tables that should be mentioned. First, the attachment of the VP in the sentence is shown in each case, although the sentence

otherwise does not appear; however, this makes no difference in the analysis of the difference between final and initial polysyndeton, because the VP[S] attachment appears in the same place in pairs of examples contrasting that property. Second, I have assumed that the coordinated NP (NP_c) is constructed immediately upon the occurrence an initial coordinator, following Hawkins, who assumes that a PP can be constructed immediately upon the occurrence of a preposition. It could be argued in the initial monosyndeton examples that NP_c cannot actually be constructed until the appearance of the first NP lets the listener know the category of element being coordinated. This is especially true in (10), in which there is no context to give the processor a clue as to the category, although in (8) it might be argued that, since a verb has already been heard, the processor is expecting a noun phrase. However, this would affect the calculations only slightly, since it simply delays the identification of the NP_c for one word. It could also be argued that, because many languages allow coordination by juxtaposition, the occurrence of two adjacent NPs in (7) and (9) is enough for listeners to construct the NP_c. If this were true, however, then we would expect that overt coordination marking does not make the listener's task any easier, and so according to MiF, asyndeton should be the most common strategy across the world's languages, but this is not the case (but see below for further discussion of the historical origin of coordination).

Several patterns are apparent in the data. First, there is a strong preference for (7) over (8): in (8), the initial coordinator allows the assignment of two more properties after the second word and three more after the third word. Second, there is a very slight preference for (10) over (9): in (10), the initial coordinator allows the assignment of one additional property after the second word. MaOP, therefore, reveals a preference in performance for one of the coordination strategies. Interestingly, this preference is the opposite of Stassen's observed universal: MaOP prefers initial monosyndeton marking to final monosyndeton marking, either strongly or weakly depending on the basic word order of the language. Note also that this preference only increases if there are more than two coordinated elements: if we extend the tables by adding more NPs, this creates more columns in which the number of assigned properties is higher in the initial monosyndeton examples.

5 Analysis

Having applied Hawkins' three processing principles, therefore, we have come to a surprising conclusion: the principles either do not prefer one strategy over the other (MiD, MiF), or else prefer the unattested strategy to the attested one (MaOP). We cannot therefore account for Stassen's universal using Hawkins' processing principles—in fact, it appears to be a counterexample to them, because an increase in the ease of processing of a structure is associated with a *decrease* in the frequency of occurrence of that structure cross-linguistically. Clearly, we have to look elsewhere for an explanation of Stassen's universal, but where? In this section, I suggest two possible alternative explanations.

The first is, like Hawkins' principles, based on the PGCH. Recall that the PGCH attempts to correlate the frequency of language features with "their degree of preference in performance" (Hawkins 2004:3). I suggest that this principle includes a broader range of processing than what is described by MaOP—the method of calculating the OP/UP ratio is based on operations are be performed by the hearer rather than by the speaker. Perhaps ease of production should also be taken into account when evaluating the degree of preference in performance.

Considering production has two effects on a MaOP analysis above. Initial monosyndeton is no longer preferred over final monosyndeton, because speakers can construct a coordinated NP_c immediately upon beginning to pronounce it, unlike hearers, who must wait until a marker of coordination occurs. Second, production considerations may actually favor final monosyndeton. Coordination allows the inclusion of arbitrarily many phrases of a given type in a position usually occupied by a single phrase of that type. When speakers are constructing coordinated structures, especially those that coordinate longer utterances (e.g. sentences), it requires less working memory if they can decide to add another coordinand as an "afterthought". If coordination were marked initially, speakers would need to know before the first coordinand whether another will follow, but since it is, in fact, marked medially or finally, speakers can delay making this decision until the following coordinand. In other words, final monosyndeton allows the speakers to leave their options open and coordinate as an afterthought, rather than having to plan it out beforehand.

The second alternative explanation has to do with the origin of coordinate structures. Mithun (1988) discusses the various origins of coordinate constructions in the world's languages. She suggests that speakers of languages that have coordination only by juxtaposition often develop overtly marked coordination when they come into contact with speakers of languages that have overtly marked coordination, or with written language. Coordination by juxtaposition, she argues, is actually coordination marked by a special intonation contour, the "comma intonation" (Mithun 1988:332), and while this strategy is sufficient in spoken language, to be unambiguous in written language, coordination must be marked somehow. Often, the sources of the newly grammaticized coordinators are comitative markers on noun phrases (the Comitative Strategy of Stassen's WITH-languages).

Let us suppose that coordination is only recently grammaticized in many languages, that comitative marking is often its source, and that, in addition, obliquely marked noun phrases tend to follow the subjects of sentences. If so, we would expect the structures that precede coordinate structures to consist of one NP followed by another NP marked as comitative. Depending on whether the language is head-initial or head-final, this produces one of two patterns:

(13) NP NP-WITH

(14) NP WITH-NP

If structures like (13) and (14) undergo reanalysis into balanced syntactic coordination and the WITH-language becomes an AND-language, the result will be coordinate structures with the word orders attested in Stassen's survey, but not the unattested AND-NP NP order.

6 Conclusion

In this paper, I have attempted to account for a language universal observed by Stassen (2000) using principles of processing proposed by Hawkins (2004), and found that they do not adequately explain it. I have offered two alternative explanations. The first is based on the idea of taking another aspect of performance, namely production,

into account when analyzing language structures. Taking production more formally into account seems like a reasonable extension of Hawkins' ideas, but much more work, including the analysis of many more typological generalizations, would need to be done before a new production-based principle could be proposed. The second explanation is based on work on the grammaticization of coordination in Mithun (1988), and it suggests that the unattested coordination structure may be lacking because of the historical origin of coordination structures. This explanation may work for languages in which coordination is a recent development, but if Hawkins' performance-based explanation is correct, why do we still see *no* examples of AND-NP NP structures, even after, in some cases, thousands of years of language change after the grammaticization of coordination? As always in linguistic typology, there remains more work to be done, both in the collection and analysis of language data and in the formulation of theories to account for variation among the world's languages.

References

Hawkins, John A. 2004. *Efficiency and complexity in grammars*. Oxford: Oxford University Press.

Mithun, Marianne. 1988. The grammaticization of coordination. *Clause combining in grammar and discourse*, Ed. John Haiman and Sandra A. Thompson. Amsterdam: Benjamins. 331–359.

Stassen, Leon. 2000. AND-languages and WITH-languages. *Linguistic Typology*, 4: 1–54.