

The evidence for word order correlations:  
a response to Dunn, Greenhill, Levinson and Gray's paper in *Nature*

Matthew S. Dryer  
University at Buffalo

## I. Introduction

1. Dunn et al argue for the following:
  - a. There is no evidence that there are universal correlations between the order of various pairs of elements (such as the order of verb and object and the order of adposition and noun phrase) for which a universal correlation has been claimed.
  - b. Among the four families they examined (Indo-European, Bantu, Austronesian, and Uto-Aztecan),
    - i. a correlation is never found in more than two families;
    - ii. a correlation is generally not found in more than one family;
    - iii. in some instances, a correlation is found in no family;and thus correlations are specific to particular lineages.
  - c. Existing theories of word order correlations predict that most families should exhibit correlations but this prediction is false.
  - d. Correlations that Dryer (1992) predicts should not exist are found in some families, such as a correlation between the order of adjective and noun and the order of relative clause and noun.
  - e. "at least with respect to word order, cultural evolution is the primary factor that determines linguistic structure, with the current state of a linguistic system shaping and constraining future states"
  - f. Cognitive factors do not play a major role in determining linguistic structure, at least as far as word order is concerned.
2. I will argue that
  - a. There is good evidence that there *are* universal word order correlations.
  - b. The correlations which they find which they say are a problem for Dryer (1992) are not a problem; these correlations are fully consistent with that paper. [I will not argue for this in that talk.]
  - c. There is no evidence that correlations are specific to particular lineages.
  - d. The method they use from biological evolution does not apply to language evolution because it ignores the powerful effect that contact plays on typological change. The method they use requires that changes in different phylogenetic branches be

independent, a requirement that is met with biological evolution but not by language evolution, because of the role of contact.

- e. The differences they find between different language families are well within the range of random variation.
- f. Their method can lead to the false conclusion that a family lacks a correlation because of the problem in (d).
- g. Their method can lead to the false conclusion that a family does show a dependency between two typological parameters only because the two parameters are ones for which particular values for those parameters are common features of a linguistic area in which part of that family falls.
- h. Existing theories of correlations do not predict that most families will exhibit a correlation in their sense; the numbers of families exhibiting a correlation in their sense is what we would expect under existing theories of correlations.
- i. Their claims (e) and (f) are entirely correct. More specifically, I claim
  - 1. the primary factor determining word order changes is not cognitive factors but rather the specific cultural history of speakers of the language
  - 2. while I claim that cognitive factors do play some role in word order change, the ways in which they do varies with other properties of grammatical system of the language
- j. Lineage plays no role in determining word order changes. Rather, it is the factors just mentioned in (i).

## II. Evidence for correlations

### 3. numbers of languages

	Po	Pr
OV	472	14
VO	42	456

### numbers of genera

	Po	Pr
OV	203	10
VO	27	121

### 4. Evidence for a universal correlation between the order of verb and object and the order of adposition and noun phrase (numbers represent numbers of genera)

	Africa	Eurasia	SEAsia&Oc	Aus-NGui	NAmer	SAmer	Total	#Lgs
OV&Po	<span style="border: 1px solid black; padding: 2px;">22</span>	<span style="border: 1px solid black; padding: 2px;">29</span>	<span style="border: 1px solid black; padding: 2px;">14</span>	<span style="border: 1px solid black; padding: 2px;">70</span>	<span style="border: 1px solid black; padding: 2px;">28</span>	<span style="border: 1px solid black; padding: 2px;">40</span>	203	472
OV&Pr	3	2	1	3	0	1	10	14
Prop Po	.88	.94	.93	.96	1.00	.98	Avg=.95	
VO&Po	8	6	0	3	4	6	27	42
VO&Pr	<span style="border: 1px solid black; padding: 2px;">28</span>	<span style="border: 1px solid black; padding: 2px;">8</span>	<span style="border: 1px solid black; padding: 2px;">35</span>	<span style="border: 1px solid black; padding: 2px;">15</span>	<span style="border: 1px solid black; padding: 2px;">23</span>	12	121	456
Prop Po	.22	.43	.00	.17	.15	.33	Avg=.22	

5. Comparing proportions of genera that are Po in each area

	Africa	Eurasia	SEAsia&Oc	Aus-NewGui	NAmer	SAmer	Avg.
OV	.88	.94	.93	.96	1.00	.98	.95
VO	.22	.43	.00	.17	.15	.33	.22

6. a. If OV then Po, and if VO then Pr.  
b. If OV then Po, and if Po then OV.

7. numbers of languages

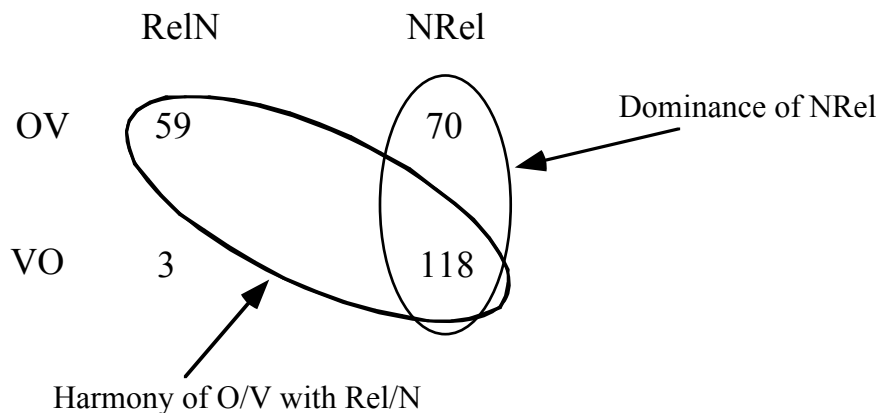
	RelN	NRel
OV	132	113
VO	5	415

- numbers of genera

	RelN	NRel
OV	59	70
VO	3	118

8. a. If VO then NRel  
b. *not* If OV then RelN  
c. *not* If NRel then VO

- 9.



### III. Dunn et al's notion of a correlation

19. Three types of language families:
- one-type families*: families in which all languages are one of the four types (e.g. Bantu, where all languages are VO and Pr)
  - correlating families*: families where two types occur that are the opposite of each other in the sense that the two types differ in the values of both parameters, but where the other two types are "relatively infrequent" if they exist at all (e.g. Austronesian, where all languages are VO&Pr or OV&Po)
  - noncorrelating families*: families where there is a pair of nonopposite types the less frequent of which is "relatively frequent" compared to the opposite of the most frequent type (e.g. Uto-Aztecan, where VO&Po languages are "relatively frequent" relative to VO&Pr)
20. Dunn et al: A family exhibits a correlation if it is a correlating family. It does not exhibit a correlation if it is a one-type family or a noncorrelating family.

21. a. The claim that there is a universal or crosslinguistic correlation does not predict that most families should be correlating families.  
 b. Rather it predicts that most families should be one of the following:  
 i. one-type families where that one type is a consistent type  
 ii. correlating families where the two types are consistent types (where the putative crosslinguistic correlation defines which types are consistent)  
 iii. noncorrelating families in which the number of languages which are consistent "outnumber" the number of languages which are not consistent  
 c. It predicts that we should not find many families of the following sort:  
 i. correlating families where the two types are inconsistent ones  
 ii. one-type families where that type is an inconsistent type  
 iii. noncorrelating families where the inconsistent languages "outnumber" the consistent languages
22. If every language family was a one-type family, then Dunn et al's method would say there was no crosslinguistic correlation.
23. Number of one-type families of each of the four types

	OV	VO
Postp	88	3
Prep	0	26

24. Families which are not one-type families

	OV&Po	OV&Pr	VO&Po	VO&Pr	consistent (OV&Po or VO&Pr)	total	% consistent
<i>Mixe-Zoque</i>	0	0	1	1	1	2	50%
<i>Uralic</i>	8	0	6	0	8	14	57%
Arawakan	2	0	5	5	7	12	58%
<i>Solomons East Papuan</i>	2	0	1	0	2	3	67%
<i>Tupian</i>	11	1	3	0	11	15	73%
<i>Khoisan</i>	4	0	1	0	4	5	80%
Uto-Aztecan	10	0	3	4	14	17	82%
Torricelli	1	0	1	4	5	6	83%
<i>Tucanoan</i>	6	0	1	0	6	7	86%
<i>West Papuan</i>	0	1	0	7	7	8	88%
Australian	16	2	1	5	21	24	88%
Niger-Congo	19	0	11	97	116	127	91%
NILO-SAHARAN	12	1	3	27	39	43	91%
INDO- EUROPEAN	26	4	1	30	56	61	92%
AFRO-ASIATIC	11	4	0	46	57	61	93%
AUSTRONESIAN	12	1	0	132	144	145	99%

*Italics: Apparently noncorrelating families*

ALL CAPS: PROBABLY CORRELATING FAMILIES

Other: Possibly noncorrelating

26. a. In every family in (24), at least 50% of the languages are one of the consistent types.  
 b. There is a positive correlation between family size and percentage of languages which are consistent. This is expected if there is a universal correlation.
28. Three types of genera that OV&Pr and VO&Po languages occur in:  
 a. genera with only one lg in sample, e.g. Peba-Yaguan  
 b. genera with more than one lg in sample, and all lgs inconsistent, e.g. Balto-Finnic  
 c. genera with more than one lg in sample, and some lgs are consistent, e.g. Indic
29. Number of genera of these three types:

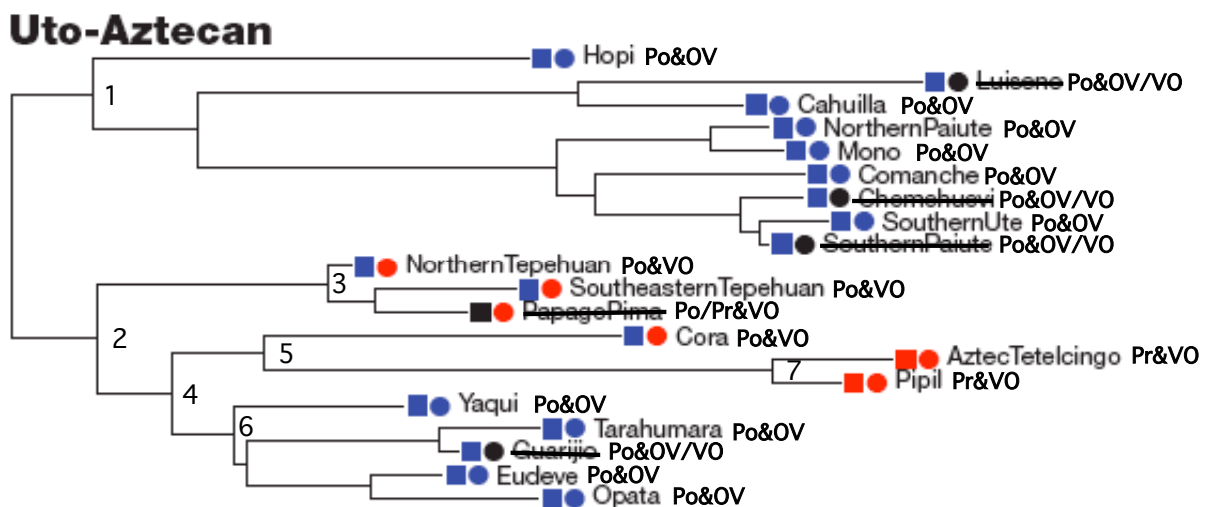
	sole lg in genus	all lgs in genus inconsistent	some other lgs in genus are consistent
OV&Pr	1	0	9
VO&Po	9	2	14
<i>Total</i>	<i>10</i>	<i>2</i>	<i>23</i>

30. The only two genera where all lgs are inconsistent VO&Po: Balto-Finnic and Kwa
31. However, Kwa also has "mixed" types (and some of the VO&Po languages allow OV word order in some syntactic contexts)  
 VO&Po: Baule, Nkonya, Lelemi, Ewe, Adiukrou.  
 VO&Po/Pr: Akan, Fongbe, Gungbe.  
 OV/VO&Po: Ajagbe.

32. For Finnish, Karlsson (1999) lists 50 postpositions and 15 prepositions.
33. What is the source of the differences among different families?

#### IV. Are the differences among families due to random variation?

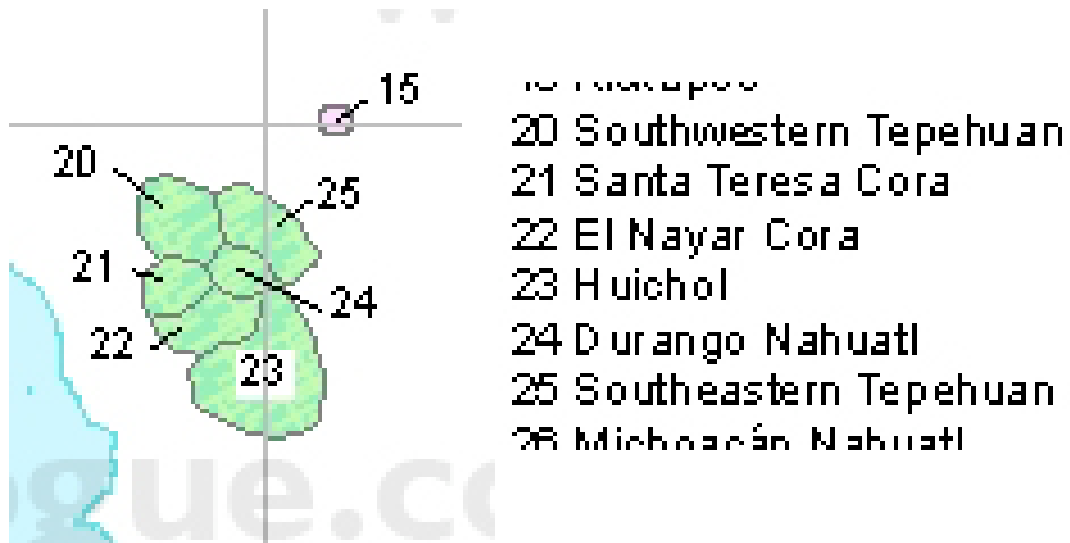
34.



35. Plausible set of changes from a phylogenetic point of view  
 a. OV&Po -> VO&Po in group 3 (Tepiman)  
 b. OV&Po -> VO&Po in Cora  
 c. OV&Po -> VO&Pr in group 7 (Aztec)

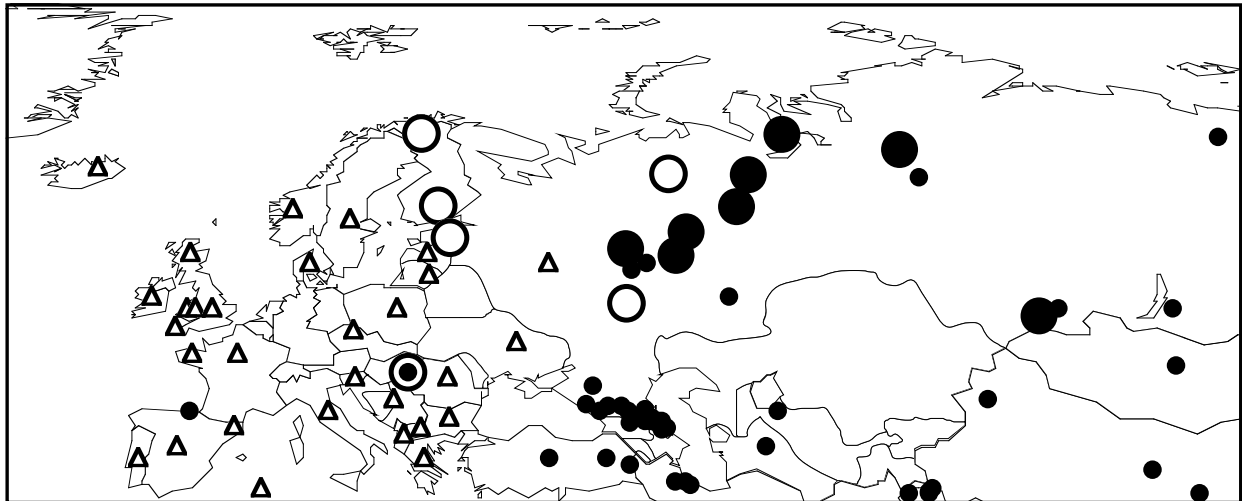
## V. The role of contact

36. Map of small area in Mexico about half way up the west side of the country (from Ethnologue):



37. Cora, like the Tepehuan languages, is also verb-initial
38. There is another language on the same branch as Cora, Huichol, which is OV
40. Uralic
- Samoyedic
    - Nenets: OV&Po
    - Enets: OV&Po
    - Kamass: OV&Po
  - Finno-Ugric
    - Ugric
      - Khanty: OV&Po
      - Mansi: OV&Po
      - Hungarian: OV/VO&Po*
    - Finnic
      - Balto-Finnic
        - Finnish: VO&Po*
        - Estonian: VO&Po*
      - Saami
        - Northern Saami: VO&Po*
      - Permic
        - Udmurt: OV&Po
        - Komi-Permyak: OV&Po
        - Komi-Zyrian: VO&Po*
      - Mari
        - Mari (Meadow): OV&Po
      - Mordvin
        - Mordvin (Erzya): VO&Po*

42. Map showing order of verb and object and order of adposition and noun phrase in Uralic and other languages



Δ - VO&Pr      ○ - VO&Po: Uralic      ● - OV&Po  
 ⊙ - OV/VO&Po: Uralic      ● - OV&Po: Uralic

Large symbols are Uralic; smaller symbols are non-Uralic; triangles are Pr; circles are Po; filled symbols are OV; symbols with white in the middle are VO; symbols which have black inside white in the middle are OV/VP.

43. The split between VO and OV within Uralic is largely geographical.
44. Dunn et al's method would conclude that there have been independent changes from OV&Po to VO&Pr in many subgroups of Uralic. But the pattern is best understood areally: both in Uralic and outside Uralic, there is a general pattern of VO to the west and OV to the east in Eurasia. I.e. these were not really independent changes.
45. Suppose the Uralic languages that have changed to VO have acquired some other typological trait associated with Indo-European languages to the west. Then Dunn et al's method would imply that this trait and VO word order are dependent on each other.
46. If multiple branches of a family all acquire traits associated with a linguistic area, then Dunn et al's method will say that these traits are dependent on each other in that family.
48. The primary impact of contact on word order is not contact-induced change, but contact-induced non-change. Contact between languages with the same word order causes these languages to remain the same.

51.

Value 1	Value 2	same genus		same family, different genus		different family	
		<500	>500	<500	>500	<500	>500
subject-verb word order	verb-subject word order	95	83	95	80	94	77
numeral-noun	noun-numeral	81	64	87	66	79	49
subject pronouns obligatory	subject pronouns not obligatory	83	83	84	74	82	79
tense/aspect affixes	no tense/aspect affixes	83	78	89	72	80	76
subject affixes on verbs	no subject affixes on verbs	75	69	74	55	73	58
little or no affixation	more than a little affixation	80	70	81	68	84	74
Mean		83	75	85	69	82	69

57. The method Dunn et al use from biological evolution does not apply to language evolution because it ignores the powerful effect that contact plays on typological change. The method they use requires that changes in different phylogenetic branches be independent, a requirement that is met with biological evolution but not by language evolution, because of the role of contact.

VI. External factors influencing the frequency of different language types are necessarily relatively weak

59. "... evolutionary processes of language diversification explore alternative ways to construct coherent language systems unfettered by tight universal constraints." (Dunn et al)

60. When external factors are strong, they will cause all languages to conform to them.

61. If not all languages conform to an external factor, then that external factor will be relatively weak.

62. If the word order correlations are partly due to processing factors (i.e. if inconsistent types are more likely to have structures which are harder to parse, as argued for by Dryer 1992 and Hawkins 1994), then, because these correlations are purely statistical, the processing factors are necessarily relatively weak.

63. Q: If it is so difficult to process sentences in languages with inconsistent word order types, then why do inconsistent languages exist at all?

- A:
- i. the processing difficulties will in many contexts be rather weak
  - ii. the difficulties arise only in infrequently-used structures
  - iii. speakers will find other ways to express relevant meanings
  - iv. exactly when processing difficulties might arise depends considerably on specific properties of the grammatical system

64. Furthermore, whenever we have competing motivations, these motivations are necessarily relatively weak, since it is common for them not to be satisfied.



65. Examples from phonology

a. competing constraints (cf. Optimality Theory)

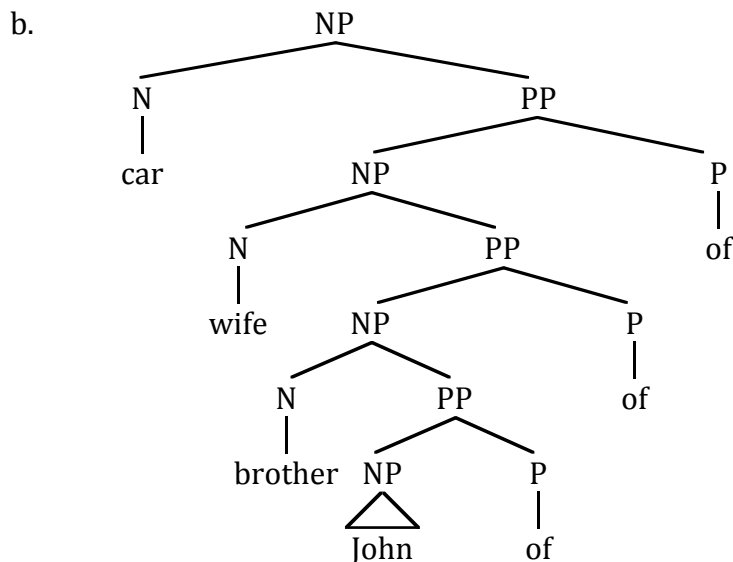
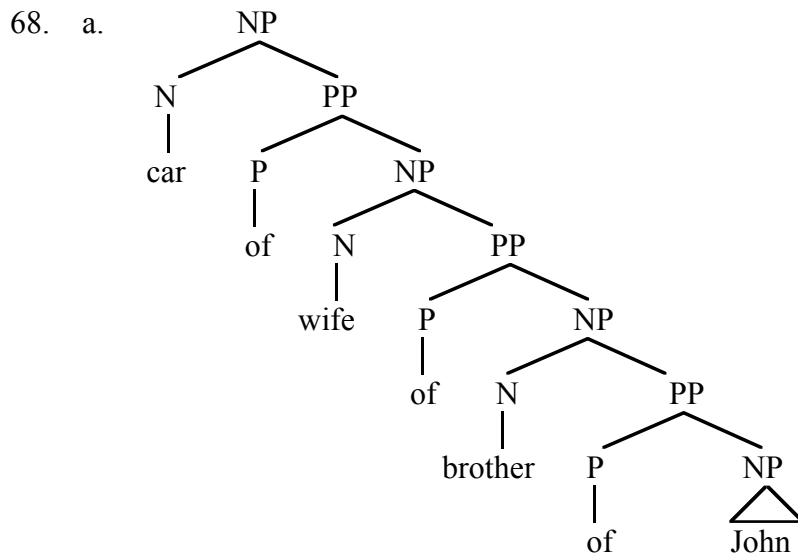
b.        p        t        k  
          m        n        (ŋ)

c. English [ɹ]

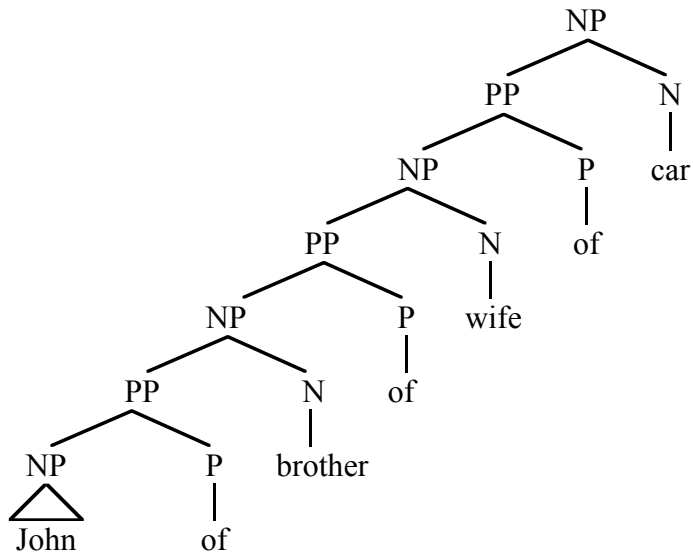
d. non-coarticulated apico-velar

66. If the correlation between the order of adposition and noun phrase and the order of noun and genitive is motivated by processing, then why are there 67 languages which are Po&NGen or Pr&GenN?

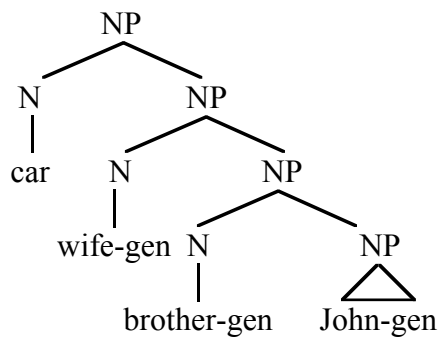
67.                    GenN   NGen  
Postp            442    13  
Prep             54    351



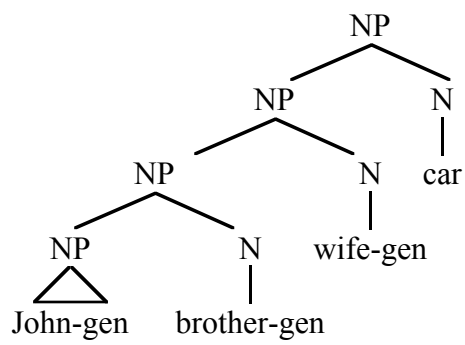
69.



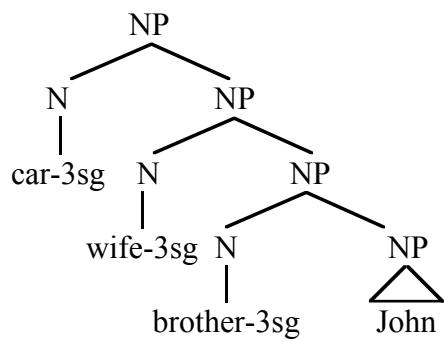
70.



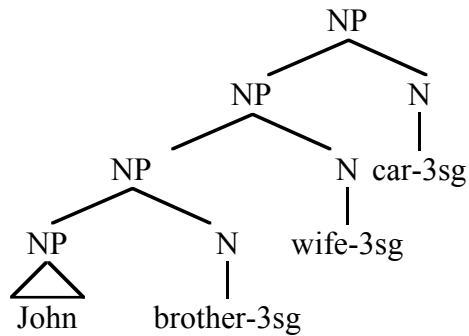
71.



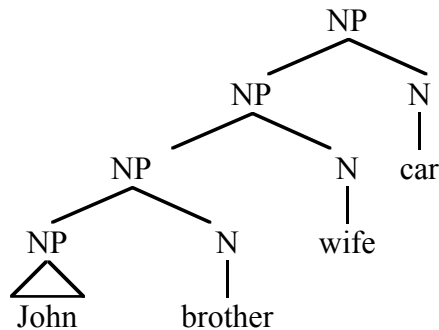
72.



73.



74.



75. Hatam (North Central Bird's Head; Indonesian New Guinea): GenN&amp;Pr

Tuhan Allah de mun  
 Lord God POSS thing  
 'God's things' (Reesink 1999: 80)

76. John's mother's car

77. a. Dunn et al seem to assume that if there are cognitive biases, they will apply uniformly across languages.  
 b. But the fact that processing difficulties associated with GenN&Pr depend on the particular genitive construction shows that this is not the case.  
 c. This shows how other features of the grammatical system play a role in determining just how potential processing difficulties will show up in a language.
78. a. My database contains 67 GenN&Pr or NGen&Po lgs, 54 of which I have data on the particular genitive construction  
 b. No GenN&Pr lgs which use Pr in genitive construction  
 c. Only 4 NGen&Po lgs which use Po in genitive construction: Kanuri, Tubu, Majang, Kugu Nganhcara

Kanuri (Saharan; Nigeria)

79. ya-nze                      Ali=be  
 mother-3SG.POSS      Ali=GEN  
 'Ali's mother' (Hutchinson 1976: 7)

80. fátò                      [kâm kùrà]=ve  
 compound              [man big]=GEN  
 'the big man's compound' (class handout, Russell Schuh 1976)

82. \*kulo [ya-nze Ali=be]=be  
farm mother-3SG.POSS Ali=GEN=GEN  
'Ali's mother's farm' (Hutchinson 1976: 10)
83. Ali=be ya-nze=be kulo-nze  
Ali=GEN mother-3SG.POSS=GEN farm-3SG.POSS  
'Ali's mother's farm'
84. kulo ya-nze Ali=be  
farm mother-3SG.POSS Ali=GEN  
'Ali's mother's farm'
85. a. %[a friend of John's]'s book  
b. %[a friend of John's] book  
c. a book that belongs to a friend of John's
86. Languages deal with the problem of structures that are difficult to process by providing alternative ways to express the relevant meanings
87. Also, structures that would be difficult to process (such as nominal possessors of possessors) would arise with fairly low frequency.

## VII. What factors determine how languages resolve competing motivations?

88. a. cultural history, especially those aspects governing contact situations  
b. specific properties of the grammatical system
89. Both orders of genitive and noun are common in SVO languages. Both SVO&GenN and SVO&NGen are common.
90. a. harmony of O/V with Gen/N  
b. harmony of S/V with Gen/N
91. SVO&GenN languages are adjacent to OV&GenN languages more often than SVO&NGen languages are.
- 93.
- |           | Africa | Eurasia | SEAsia&Oc | Aus-NGui | NAmer | SAmer | Total   | #Lgs |
|-----------|--------|---------|-----------|----------|-------|-------|---------|------|
| OV&RelN   | 6      | 21      | 11        | 11       | 3     | 7     | 59      | 132  |
| OV&NRel   | 21     | 5       | 3         | 17       | 15    | 9     | 70      | 113  |
| Prop RelN | .22    | .81     | .79       | .39      | .17   | .44   | Avg=.47 |      |
94. If NAdj, then NRel.

95.
  - a. Tibeto-Burman languages are most often NAdj
  - b. Tibeto-Burman languages are in a large area where OV&ReIN predominates over OV&NRel
  - c. 25 of the 33 NAdj&ReIN languages in my database are Tibeto-Burman
96. Dunn et al's claims (1e) and (1f) are entirely correct. More specifically, the primary factors determining word order changes are not cognitive factors but rather
  - i. the specific cultural history of the language
  - ii. the entire grammatical system of the language

## IX. The role of lineage

97. "Instead [of cognitive factors], it [linguistic diversity] is the product of cultural evolution, canalized by the systems that have evolved during diversification, so that future states lie in an evolutionary landscape with channels and basins of attraction that are specific to linguistic lineages."
98. Lineage plays no role in determining word order changes.
99.
  - a. The transition probabilities of particular word order changes for languages in the same lineage but in different contact situations are very different.
  - b. The transition probabilities of particular word order changes for languages in the same lineage but with different grammatical properties are different.
  - c. The transition probabilities of particular word order changes for languages in different lineages but in the same contact situation are similar.
  - d. The transition probabilities of particular word order changes for languages in different lineages but with similar grammatical properties are similar.
100. Instead [of cognitive factors] *and linguistic lineages*, it [linguistic diversity] is the product of cultural evolution, canalized by the systems that have evolved during diversification, ~~so that~~ *and* future states lie in an evolutionary landscape with channels and basins of attraction that are specific to ~~linguistic lineages~~ *particular languages, reflecting their specific cultural history and overall grammatical system.*