Typology: Lecture VI
Word Order

Christian Bentz
University of Tübingen

May 24, 2017
OVERVIEW

INTRODUCTION
The Basic Order of Subject, Object, and Verb

DATA AVAILABLE
Inventory-Based Accounts of Word Order
Distribution-Based Accounts of Word Order

RECENT RESEARCH
Information-Theoretic Considerations
Word Order in “Proto-World?”
Basic Order of Agent, Patient, Action
Word Order Case-Marking Trade-Off
Phylogenetics
THE BASIC ORDER OF SUBJECT, OBJECT AND VERB

In a prototypical transitive sentence the

- **Subject**
  shows *agreement* with the predicate,
  directly *preceedes* the predicate,
  is the *topic* of the sentence,
  and represents the *semantic role of agent*

- **Object**
  does not agree with the predicate,
  represents the *patient*

- **Verb** represents the *action*

e.g. [s The dog] [v chased] [o the cat]

Note that there are counterxamples for any of these criteria: e.g. in passive sentences the subject might be the patient of an action.

[s The dog] [v was chased] [o by the cat]
**Word Order at Different Levels of Abstraction**

- **Syntactic role level**:
  
  \[S][V][O]\n
  \[[s \text{ colorless green ideas}][v \text{ escape}][o \text{ meaning}]\]

- **Phrase level**:

  \[NP][VP][NP]\n
  \[[\text{NP colorless green ideas}][\text{VP escape}][\text{NP meaning}]\]

- **Word level**:

  \[\text{colorless}][\text{green}][\text{ideas}][\text{escape}][\text{meaning}]\]

  “colorless green ideas escape meaning”

Actually: typologists talk about “syntactic role order” or “phrase order” rather than “word order”
Greenberg’s Word Order Universals

Greenberg (1963) gave a set of statistical universals (i.e., trends rather than absolute statements) of word order. He used a relatively small sample of geographically and genetically diverse languages.

Problems with this approach:

- It is somewhat hard to define “basic word order” (most notorious case: German, with different word order in different clause types).
- There arguably are languages without fixed word order (Slavic languages, Latin, …), for which we can only talk about preferred word order.

The Six Possible Orders

SOV

(1) Ainu (Isolate: Japan)

kamuy aynu rayke
bear person kill

“The bear killed the person”

SVO

(2) Matuumbi (Niger-Congo)

abũnwaаsĩ aachéngite ŋũũmba
PN he.built house

“Abumwas built a house”

**The Six Possible Orders**

**VSO**

(3) Irish (Indo-European)

tógann Máire an cat

“Mary lifts the cat”

**VOS**

(4) Cèmuhî (Austronesian)

ɛ ālī-hī ā-li mwà ɔ pā-li āpūlīp

3SG see-TR ART:NEUT-DEF house SUBJ ART:NF-DEF man

“The man saw the house”

THE SIX POSSIBLE ORDERS

OVS

(5) Päri (Nilo-Saharan)

dháagɔ̀ á-yàaŋ ùbúrr-ì
cwoman COMPL-insult PN-ERG

“Ubur insulted the woman”

OSV

(6) Warao (Isolate: Venezuela)

erike hube abun-ae
PN snake bite-PAST

“A snake bit Enrique”

**Flexible Word Order**

Nhanda (Pama-Nyungan)

```
abarla-lu  wumba-yi  wur’a-tha
child-ERG  steal-PPERF money-1SGOBL
S          V          O
abarla-lu  wur’a-tha wumba-yi
S          O          V
wumba-yi  wur’a-tha  abarla-lu
V          O          S
wumba-yi  abarla-lu  wur’a-tha
V          S          O
wur’a-tha  wumba-yi  abarla-lu
O          V          S
wur’a-tha  abarla-lu  wumba-yi
O          S          V
```

“The child stole my money”

TWO BASIC ACCOUNTS OF WORD ORDER

Inventory-based

- Assign one of 7 categories (SOV, SVO, VSO, VOS, OSV, OVS, non dominant) based on a grammatical description of the language or a small “corpus” of collected sentences

Distribution-based

- Give the frequencies of sentences with the 7 word order options in a corpus of the language
INTRODUCTION

DATA AVAILABLE

RECENT RESEARCH

INVENTORY-BASED

WALS Chapter 81
Basic order of Subject, Object and Verb for 1377 languages
WALS Chapter 81
Papua New Guinea
WALS Chapter 81
South America
**Introduction**

Data Available

Recent Research

Hammarstöm (2013, 2016)

Basic order of Subject, Object and Verb for more than 5000 languages (not yet available online)

SOV
SVO
VSO
VOS
OVS
OSV

No Dominant Order
DISTRIBUTION-BASED

Account by Östling (2015) based on 987 parallel texts of the Bible Corpus. Collects the frequencies of word orders occurring in the texts.

Examples

(7) German

Im Anfang schuf Gott Himmel und Erde
in.the beginning [create.PERF]V [god]S [heaven and earth]O

“In the beginning god created the heavens and the earth”

Finsternis lag über der Tiefe
[darkness]S [lay.PERF]V over [the deep]O

“Darkness was on the face of the deep”

## Distribution-Based

<table>
<thead>
<tr>
<th>ISO</th>
<th>SOV</th>
<th>SVO</th>
<th>OSV</th>
<th>OVS</th>
<th>VSO</th>
<th>VOS</th>
<th>Total</th>
<th>SOV_perc</th>
<th>MAX</th>
<th>WALS</th>
</tr>
</thead>
<tbody>
<tr>
<td>aai</td>
<td>64</td>
<td>8</td>
<td>12</td>
<td>6</td>
<td>1</td>
<td>2</td>
<td>93</td>
<td>0.69</td>
<td>SOV</td>
<td>_</td>
</tr>
<tr>
<td>aak</td>
<td>81</td>
<td>17</td>
<td>20</td>
<td>33</td>
<td>5</td>
<td>9</td>
<td>165</td>
<td>0.49</td>
<td>SOV</td>
<td>_</td>
</tr>
<tr>
<td>aau</td>
<td>37</td>
<td>1</td>
<td>7</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>46</td>
<td>0.80</td>
<td>SOV</td>
<td>SOV</td>
</tr>
<tr>
<td>abt</td>
<td>29</td>
<td>7</td>
<td>5</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>51</td>
<td>0.57</td>
<td>SOV</td>
<td>SOV</td>
</tr>
<tr>
<td>aby</td>
<td>53</td>
<td>11</td>
<td>10</td>
<td>11</td>
<td>6</td>
<td>5</td>
<td>96</td>
<td>0.55</td>
<td>SOV</td>
<td>_</td>
</tr>
<tr>
<td>acc</td>
<td>4</td>
<td>27</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>41</td>
<td>0.10</td>
<td>SVO</td>
<td>_</td>
</tr>
<tr>
<td>acd</td>
<td>5</td>
<td>30</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>42</td>
<td>0.12</td>
<td>SVO</td>
<td>_</td>
</tr>
<tr>
<td>ace</td>
<td>3</td>
<td>70</td>
<td>10</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>101</td>
<td>0.03</td>
<td>SVO</td>
<td>No dominant order</td>
</tr>
<tr>
<td>acf</td>
<td>9</td>
<td>89</td>
<td>13</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>113</td>
<td>0.08</td>
<td>SVO</td>
<td>_</td>
</tr>
<tr>
<td>acn</td>
<td>103</td>
<td>2</td>
<td>26</td>
<td>12</td>
<td>2</td>
<td>1</td>
<td>146</td>
<td>0.71</td>
<td>SOV</td>
<td>_</td>
</tr>
<tr>
<td>acr</td>
<td>4</td>
<td>22</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>35</td>
<td>0.11</td>
<td>SVO</td>
<td>_</td>
</tr>
<tr>
<td>acu</td>
<td>21</td>
<td>5</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>31</td>
<td>0.68</td>
<td>SOV</td>
<td>_</td>
</tr>
<tr>
<td>ade</td>
<td>10</td>
<td>121</td>
<td>9</td>
<td>2</td>
<td>9</td>
<td>10</td>
<td>161</td>
<td>0.06</td>
<td>SVO</td>
<td>_</td>
</tr>
<tr>
<td>adh</td>
<td>16</td>
<td>94</td>
<td>4</td>
<td>1</td>
<td>5</td>
<td>10</td>
<td>130</td>
<td>0.12</td>
<td>SVO</td>
<td>_</td>
</tr>
<tr>
<td>adj</td>
<td>8</td>
<td>39</td>
<td>10</td>
<td>2</td>
<td>6</td>
<td>5</td>
<td>70</td>
<td>0.11</td>
<td>SVO</td>
<td>SVO</td>
</tr>
<tr>
<td>adz</td>
<td>48</td>
<td>15</td>
<td>11</td>
<td>9</td>
<td>7</td>
<td>5</td>
<td>95</td>
<td>0.51</td>
<td>SOV</td>
<td>SVO</td>
</tr>
<tr>
<td>aey</td>
<td>72</td>
<td>9</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>9</td>
<td>97</td>
<td>0.74</td>
<td>SOV</td>
<td>SOV</td>
</tr>
<tr>
<td>afr</td>
<td>523</td>
<td>142</td>
<td>52</td>
<td>13</td>
<td>44</td>
<td>2</td>
<td>776</td>
<td>0.67</td>
<td>SOV</td>
<td>_</td>
</tr>
<tr>
<td>agd</td>
<td>69</td>
<td>18</td>
<td>18</td>
<td>27</td>
<td>13</td>
<td>6</td>
<td>151</td>
<td>0.46</td>
<td>SOV</td>
<td>SOV</td>
</tr>
<tr>
<td>agg</td>
<td>39</td>
<td>9</td>
<td>8</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>62</td>
<td>0.63</td>
<td>SOV</td>
<td>SOV</td>
</tr>
</tbody>
</table>

**Distribution-Based**

<table>
<thead>
<tr>
<th>ISO</th>
<th>Name</th>
<th>SOV</th>
<th>SVO</th>
<th>OSV</th>
<th>OVS</th>
<th>VSO</th>
<th>VOS</th>
<th>Total</th>
<th>SOV_perc</th>
<th>SVO_perc</th>
<th>MAX</th>
<th>WALS</th>
</tr>
</thead>
<tbody>
<tr>
<td>hin</td>
<td>Hindi</td>
<td>194</td>
<td>14</td>
<td>18</td>
<td>14</td>
<td>2</td>
<td>1</td>
<td>243</td>
<td>0.8</td>
<td>0.06</td>
<td>SOV</td>
<td>0.06 SOV</td>
</tr>
<tr>
<td>eng</td>
<td>English</td>
<td>457</td>
<td>291</td>
<td>23</td>
<td>74</td>
<td>14</td>
<td>5025</td>
<td>0.01</td>
<td>0.91</td>
<td>0.28</td>
<td>VSO</td>
<td>0.28 VSO</td>
</tr>
<tr>
<td>haw</td>
<td>Hawaiian</td>
<td>3</td>
<td>31</td>
<td>2</td>
<td>2</td>
<td>70</td>
<td>3</td>
<td>111</td>
<td>0.03</td>
<td>0.07</td>
<td>VSO</td>
<td>0.07 VSO</td>
</tr>
<tr>
<td>mav</td>
<td>Sateré-Mawé (Tupian)</td>
<td>3</td>
<td>9</td>
<td>4</td>
<td>7</td>
<td>8</td>
<td>11</td>
<td>42</td>
<td>0.07</td>
<td>0.21</td>
<td>VOS</td>
<td>0.21 VOS</td>
</tr>
<tr>
<td>czt</td>
<td>Zotung Chin (Sino-Tibetan)</td>
<td>8</td>
<td>35</td>
<td>25</td>
<td>4</td>
<td>9</td>
<td>115</td>
<td>0.3</td>
<td>0.07</td>
<td>0.07</td>
<td>OSV</td>
<td>0.07 OSV</td>
</tr>
<tr>
<td>cny</td>
<td>Ngawn Chin (Sino-Tibetan)</td>
<td>6</td>
<td>14</td>
<td>33</td>
<td>4</td>
<td>4</td>
<td>90</td>
<td>0.32</td>
<td>0.07</td>
<td>0.07</td>
<td>OSV</td>
<td>0.07 OSV</td>
</tr>
<tr>
<td>deu</td>
<td>German</td>
<td>1571</td>
<td>1391</td>
<td>251</td>
<td>529</td>
<td>105</td>
<td>3900</td>
<td>0.4</td>
<td>0.36</td>
<td>0.36</td>
<td>SOV</td>
<td>No dominant order</td>
</tr>
</tbody>
</table>

DISTRIBUTION-BASED

DISTRIBUTION-BASED

1. Information-Theoretic Considerations
2. Word Order in “Proto-World?”
3. Basic Order of Agent, Patient, Action
4. Word Order and Case-Marking Trade-Off
5. Phylogenetics
**Word Order Change and Evolution: The Permutation Ring**

[Ferrer-i-Cancho (2017). The placement of the head that maximizes predictability. An information-theoretic approach.]
TWO COMPETING PRESSURES THAT SHAPE WORD ORDER

Dependency-length minimization
The head of a sentence/phrase (e.g. the verb) should be placed in a way that **minimizes** dependency lengths

Predictability maximization
The head of a sentence/phrase should be placed in a way that **maximizes** its predictability

[Ferrer-i-Cancho (2017). The placement of the head that maximizes predictability. An information-theoretic approach.]
Placing the verb (head) in the **middle position** minimizes dependency lengths.

```
[head]

S  V  O

[NP colorless green ideas]    [VP escape]    [NP meaning]
```
Placing the verb (head) in the **last position** maximizes its predictability.
**Word order in “Proto-World”?**

The origin and evolution of word order

Murray Gell-Mann\(^a,1\) and Merritt Ruhlen\(^b,1\)

\(^a\)Santa Fe Institute, Santa Fe, NM 87501; and \(^b\)Department of Anthropology, Stanford University, Stanford, CA 94305

Contributed by Murray Gell-Mann, August 26, 2011 (sent for review August 19, 2011)

Recent work in comparative linguistics suggests that all, or almost all, attested human languages may derive from a single earlier language. If that is so, then this language—like nearly all extant languages—most likely had a basic ordering of the subject (S), verb (V), and object (O) in a declarative sentence of the type “the man (S) killed (V) the bear (O).” When one compares the man”) and uses prepositions. (Nowadays, these correlations are described in terms of head-first and head-last constructions.) In light of such correlations it is often possible to discern relic traits, such as GN order in a language that has already changed its basic word order from SOV to SVO. Later work (7) has shown that diachronic pathways of grammaticalization often reveal relic

![Diagram of word order evolution](image)

**Fig. 1.** Evolution of word order.

## Word order in “Proto-World”


<table>
<thead>
<tr>
<th>Language Family</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>World</td>
<td>1008-770-164-[40-16-13]</td>
</tr>
<tr>
<td>Khoisan</td>
<td>22-11-1-[0-0-0]</td>
</tr>
<tr>
<td>Congo-Saharan</td>
<td>61-318-16-[0-1-0]</td>
</tr>
<tr>
<td>Niger-Kordofanian</td>
<td>39-279-1-[0-0-0]</td>
</tr>
<tr>
<td>Nilo-Saharan</td>
<td>22-39-15-[0-1-0]</td>
</tr>
<tr>
<td>Indo-Pacific</td>
<td>223-25-1-[0-0-2]</td>
</tr>
<tr>
<td>Australian</td>
<td>59-20-1-[3-1-1]</td>
</tr>
<tr>
<td>Austric</td>
<td>30-220-67-[16-0-2]</td>
</tr>
<tr>
<td>Austroasiatic</td>
<td>8-34-0-[1-0-0]</td>
</tr>
<tr>
<td>Miao-Yao</td>
<td>0-4-0-[0-0-0]</td>
</tr>
<tr>
<td>Daic</td>
<td>1-19-0-[0-0-0]</td>
</tr>
<tr>
<td>Austronesian</td>
<td>21-163-67-[15-0-2]</td>
</tr>
<tr>
<td>Dene-Caucasian</td>
<td>157-13-0-[0-0-0]</td>
</tr>
<tr>
<td>Basque</td>
<td>1-0-0-[0-0-0]</td>
</tr>
<tr>
<td>Caucasian</td>
<td>29-0-0-[0-0-0]</td>
</tr>
<tr>
<td>Burushaski</td>
<td>1-0-0-[0-0-0]</td>
</tr>
<tr>
<td>Sino-Tibetan</td>
<td>84-13-0-[0-0-0]</td>
</tr>
<tr>
<td>Ket</td>
<td>1-0-0-[0-0-0]</td>
</tr>
<tr>
<td>Na-Dene</td>
<td>41-0-0-[0-0-0]</td>
</tr>
<tr>
<td>Nostratic-Amerind</td>
<td>456-163-78-[21-14-8]</td>
</tr>
<tr>
<td>Afro-Asiatic</td>
<td>58-37-14-[0-0-0]</td>
</tr>
<tr>
<td>Nostratic</td>
<td>182-59-6-[0-0-0]</td>
</tr>
<tr>
<td>Kartvelian</td>
<td>4-0-0-[0-0-0]</td>
</tr>
<tr>
<td>Dravidian</td>
<td>28-0-0-[0-0-0]</td>
</tr>
<tr>
<td>Eurasian</td>
<td>149-59-6-[0-0-0]</td>
</tr>
<tr>
<td>Indo-Hittite</td>
<td>79-47-6-[0-0-0]</td>
</tr>
<tr>
<td>Uralic</td>
<td>10-10-0-[0-0-0]</td>
</tr>
<tr>
<td>Altaic</td>
<td>50-1-0-[0-0-0]</td>
</tr>
<tr>
<td>Ainu</td>
<td>1-0-0-[0-0-0]</td>
</tr>
<tr>
<td>Gilyak</td>
<td>1-0-0-[0-0-0]</td>
</tr>
<tr>
<td>Chukchi-Kamchatkan</td>
<td>2-1-0-[0-0-0]</td>
</tr>
<tr>
<td>Eskimo-Aleut</td>
<td>6-0-0-[0-0-0]</td>
</tr>
<tr>
<td>Amerind</td>
<td>216-67-58-[21-14-8]</td>
</tr>
</tbody>
</table>

The numbers after each family represent the number of languages with SOV, SVO, VSO, VOS, OVS, and OSV orders, given in that order, with the final three word orders in brackets. Note that we have chosen one of the several definitions of Nostratic.
The natural order of events: How speakers of different languages represent events nonverbally

Susan Goldin-Meadow*,†, Wing Chee So‡, Aslı Özyürek§, and Carolyn Mylander*

*Department of Psychology, University of Chicago, 5730 South Woodlawn Avenue, Chicago, IL 60637; †Department of Psychology, National University of Singapore, Republic of Singapore 119077; ‡Department of Linguistics, Radboud University Nijmegen, 6525 HT Nijmegen, The Netherlands; §Max Planck Institute for Psycholinguistics, 6500 AH Nijmegen, The Netherlands; and †Department of Psychology, Koc University, 34450 Istanbul, Turkey

Edited by Rochel Gelman, Rutgers, The State University of New Jersey, Piscataway, NJ, and approved May 8, 2008 (received for review November 12, 2007)

To test whether the language we speak influences our behavior Gesture Task (6). Forty adults [10 English speakers, 10 Turkish
**Semantics and Word Order: Extensional and Intensional Events**

extensional (above), intensional (below)

**Fig. 1.** Example item: intensional event: 'Plate thrown guitar'.

**Fig. 2.** Example item: extensional event: 'Cook thinks of sock'.

**Fig. 3.** Results: extensional events. Mean proportions of SOV and SVO gesturing orders for extensional events. Error bars indicate standard error of the mean.

**Fig. 4.** Results: intensional events. Mean proportions of SOV and SVO gesturing orders for intensional events. Error bars indicate standard error of the mean.

[Schouwstra, Marieke, and De Swart, Henriette (2014) The semantic origins of word order. *Cognition* 131.]
**Semantics and Word Order: Reversible and Nonreversible Events**

[Fig. 2. Proportions of responses in basic three-word orders for nonreversible events (inanimate objects).]

[Fig. 3. Proportions of responses in basic three-word orders for reversible events (human objects).]

**Word Order and Case-Marking: Experimental Evidence**

**Word Order and Case-Marking: The Cross-Linguistic Evidence**

**Word Order and Case Marking: Statistical Analysis**

**Typological Analysis**

- balanced sample of 50 languages
- data from WALS

<table>
<thead>
<tr>
<th></th>
<th>M-</th>
<th>M+</th>
</tr>
</thead>
<tbody>
<tr>
<td>WO-</td>
<td>3</td>
<td>22</td>
</tr>
<tr>
<td>WO+</td>
<td>8</td>
<td>17</td>
</tr>
</tbody>
</table>

χ² = 2.91, p = 0.088

WO-: word order does not disambiguate S and O
WO+: word order does disambiguate S and O
M-: no morphological marking
M+: morphological marking

**Word Order Change and Evolution: Phylogenetics**

- **VO order**
- **OV order**
- **no dominant order**
- **AdN order (prepositions)**
- **NAd order (postpositions)**
- **no dominant order**

THANKS. SEE YOU NEXT WEEK!