The evolution of word-order universals: Some word-order correlation are lineage specific - others might be universal

Gerhard Jäger, Gwendolyn Berger, Isabella Boga, Thora Daneyko & Luana Vaduva

Tübingen University

Association for Linguistic Typology, Canberra

December 15, 2017
Introduction
Word order correlations

- Greenberg, Keenan, Lehmann etc.: general tendency for languages to be either consistently head-initial or consistently head-final

- alternative account (Dryer, Hawkins): phrases are consistently left- or consistently right-branching

- can be formalized as collection of implicative universals, such as With overwhelmingly greater than chance frequency, languages with normal SOV order are postpositional. (Greenberg’s Universal 4)

- both generativist and functional/historical explanations in the literature
Phylogenetic non-independence

- languages are phylogenetically structured
- if two closely related languages display the same pattern, these are not two independent data points
  ⇒ we need to control for phylogenetic dependencies

(from Dunn et al., 2011)
Maslova (2000):

“If the A-distribution for a given typology cannot be assumed to be stationary, a distributional universal cannot be discovered on the basis of purely synchronic statistical data.”

“In this case, the only way to discover a distributional universal is to estimate transition probabilities and as it were to ‘predict’ the stationary distribution on the basis of the equations in (1).”
The phylogenetic comparative method
Modeling language change

Markov process
Modeling language change

Markov process

Phylogeny
Modeling language change

Markov process

Phylogeny

Branching process
Estimating rates of change

- if phylogeny and states of extant languages are known...
Estimating rates of change

- if phylogeny and states of extant languages are known...
- ... transition rates, stationary probabilities and ancestral states can be estimated based on Markov model
Correlation between features

Pagel and Meade (2006)

- construct two types of Markov processes:
  - independent: the two features evolve according to independent Markov processes
  - dependent: rates of change in one feature depends on state of the other feature

- fit both models to the data
- apply statistical model comparison
Dunn et al. (2011)
all 28 pairs of 8 word-order features considered

4 language families: Austronesian, Bantu, Indo-European, and Uto-Aztecan

main finding: wildly different results between families

conclusion: word-order correlations are lineage-specific
Universal and lineage-specific models
This study

Experiments

1. replication of Dunn et al. (2011) with different data
2. model comparison: universal vs. lineage-specific correlations
3. word-order correlations across a world-tree of languages
4. automatically identifying lineage-specificity
Data

- **word-order data**: WALS
- **phylogeny**:
  - ASJP word lists (Wichmann et al., 2016)
  - feature extraction (automatic cognate detection, *inter alia*) $\leadsto$ character matrix
  - Maximum-Likelihood phylogenetic inference with Glottolog (Hammarström et al., 2016) tree as backbone
  - advantages over hand-coded Swadesh lists
    - applicable across language families
    - covers more languages than those for which expert cognate judgments are available
- 1004 languages in total
- Austronesian: 123; Bantu: 41; Indo-European: 53; Uto-Aztecan: 13
Replication of Dunn et al.

![Graph showing Bayes Factor comparison between this study and Dunn et al.]

- **Family**:
  - Austronesian
  - Bantu
  - Indo-European
  - Uto-Aztecan

**Universal and lineage-specific models**
Comparing universal and lineage-specific models

- so far: fitting a separate model for each language family
  - **advantage**: good fit of the lineage-specific data
  - **disadvantage**: many parameters (8 per family for a dependent model)

- statistical model comparison: quantifying to what degree the data support the excess parameters of lineage-specific models

- models to be compared:
  - **universal**: one set of rates (8 parameters), applying to all 4 families
  - **lineage specific**: a separate set of rates for each family

- comparison via **Bayes Factor**
  (implementation with RevBayes; Höhna et al. 2016)
Results

- very strong evidence for universality:
  - noun-adjective ↔ noun-numeral
  - adposition-noun ↔ verb-object
- strong evidence for universality:
  - adposition-noun ↔ verb-object ↔ noun-genitive ↔ noun-relative clause
- strong or very strong evidence for lineage specificity:
  - behavior of noun-adjective and noun-numeral
Results

universal (PN/VO)

lineage-specific (NG/NNum)

Jäger et al. (Tübingen)

Word-order Universals

ALT2017 18 / 26
Using the world tree

Glottolog family
- Atlantic-Congo
- Mande
- Afro-Asiatic
- Nuclear_Trans_New_Guinea
- Pama-Nyungan
- Timor-Alor-Pantar
- Otomanguean
- Indo-European
- Uto-Aztecan
- Tai-Kadai
- Mayan
- Austronesian
- Austroasiatic
- Sino-Tibetan
- Quechuan

Macro-Area
- Africa
- Papunesia
- Eurasia
- South America
- North America
- Australia
Results

- strong evidence for dependent model for 21 out of 28 feature pairs
- no evidence for independent model
- strongest evidence (BF > 100) supports Dryer (1992)
Automatically identifying lineage-specificity

- Lineages with different dynamics can be inferred automatically on the world tree.
- Latest version of *BayesTraits* (v. 3) implements a model (“discrete covarion model”) where languages can be either in a dependent or an independent state.
- Statistical model comparison between universal and lineage-dependent model (in this sense).

### Feature Pair Bayes Factor

<table>
<thead>
<tr>
<th>Feature Pair</th>
<th>Bayes Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>PN-NRe</td>
<td>0.42</td>
</tr>
<tr>
<td>NG-NRe</td>
<td>-0.90</td>
</tr>
<tr>
<td>PN-NG</td>
<td>-1.37</td>
</tr>
<tr>
<td>PN-VO</td>
<td>-2.29</td>
</tr>
<tr>
<td>VO-NRe</td>
<td>-4.86</td>
</tr>
<tr>
<td>NA-ND</td>
<td>-11.66</td>
</tr>
<tr>
<td>NA-NRe</td>
<td>-21.42</td>
</tr>
<tr>
<td>ND-NNum</td>
<td>-22.86</td>
</tr>
<tr>
<td>ND-NRe</td>
<td>-23.16</td>
</tr>
<tr>
<td>NG-VO</td>
<td>-25.20</td>
</tr>
<tr>
<td>PN-VO</td>
<td>-25.70</td>
</tr>
<tr>
<td>ND-VO</td>
<td>-28.63</td>
</tr>
<tr>
<td>NG-NS</td>
<td>-29.05</td>
</tr>
<tr>
<td>VO-NS</td>
<td>-29.74</td>
</tr>
<tr>
<td>PN-ND</td>
<td>-30.35</td>
</tr>
<tr>
<td>ND-VO</td>
<td>-30.90</td>
</tr>
<tr>
<td>NA-ND</td>
<td>-31.42</td>
</tr>
<tr>
<td>ND-NG</td>
<td>-37.75</td>
</tr>
<tr>
<td>NA-NS</td>
<td>-40.18</td>
</tr>
<tr>
<td>NRc-NS</td>
<td>-44.06</td>
</tr>
<tr>
<td>NA-PN</td>
<td>-44.25</td>
</tr>
<tr>
<td>NNum-NRc</td>
<td>-45.30</td>
</tr>
<tr>
<td>NA-VO</td>
<td>-49.34</td>
</tr>
<tr>
<td>NNum-NRc</td>
<td>-53.38</td>
</tr>
<tr>
<td>PN-NNum</td>
<td>-55.88</td>
</tr>
<tr>
<td>NA-NNum</td>
<td>-58.86</td>
</tr>
<tr>
<td>NNum-VO</td>
<td>-64.76</td>
</tr>
<tr>
<td>NG-NNum</td>
<td>-66.61</td>
</tr>
</tbody>
</table>
Automatically identifying lineage-specificity

- no evidence for truly universal dependent model
- equivocal evidence for 5 feature pairs
- define a cluster for which there was strong evidence for universality in experiment 2

<table>
<thead>
<tr>
<th>feature pair</th>
<th>Bayes Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>PN-NRc</td>
<td>0.42</td>
</tr>
<tr>
<td>NG-NRc</td>
<td>-0.90</td>
</tr>
<tr>
<td>PN-NG</td>
<td>-1.37</td>
</tr>
<tr>
<td>PN-VO</td>
<td>-2.29</td>
</tr>
<tr>
<td>VO-NRc</td>
<td>-4.86</td>
</tr>
<tr>
<td>NA-ND</td>
<td>-11.66</td>
</tr>
<tr>
<td>NA-NRc</td>
<td>-21.42</td>
</tr>
<tr>
<td>ND-NNum</td>
<td>-22.86</td>
</tr>
<tr>
<td>ND-NRc</td>
<td>-23.16</td>
</tr>
<tr>
<td>NG-VO</td>
<td>-25.20</td>
</tr>
<tr>
<td>PN-VS</td>
<td>-25.70</td>
</tr>
<tr>
<td>ND-VS</td>
<td>-28.63</td>
</tr>
<tr>
<td>NG-VS</td>
<td>-29.05</td>
</tr>
<tr>
<td>VO-VS</td>
<td>-29.74</td>
</tr>
<tr>
<td>PN-ND</td>
<td>-30.35</td>
</tr>
<tr>
<td>ND-VO</td>
<td>-30.90</td>
</tr>
<tr>
<td>NA-NNum</td>
<td>-31.42</td>
</tr>
<tr>
<td>ND-NG</td>
<td>-37.75</td>
</tr>
<tr>
<td>NA-VS</td>
<td>-40.18</td>
</tr>
<tr>
<td>NRc-VO</td>
<td>-44.06</td>
</tr>
<tr>
<td>NA-PN</td>
<td>-44.25</td>
</tr>
<tr>
<td>NNum-VO</td>
<td>-45.30</td>
</tr>
<tr>
<td>NA-VO</td>
<td>-49.34</td>
</tr>
<tr>
<td>NNum-NRc</td>
<td>-53.38</td>
</tr>
<tr>
<td>PN-NNum</td>
<td>-55.88</td>
</tr>
<tr>
<td>NA-NG</td>
<td>-58.86</td>
</tr>
<tr>
<td>NNum-VO</td>
<td>-64.76</td>
</tr>
<tr>
<td>NG-NNum</td>
<td>-66.61</td>
</tr>
</tbody>
</table>
Universal and lineage-specific models

What the dependencies look like
Conclusion
Conclusion

- empirical
  - *universal* vs. *lineage-specific* is not an absolute distinction, but a matter of degree
  - some “classical” word-order correlation fall very close to the universal end

- methodological
  - important to fit statistical model across language-families
Our co-authors

Thora Daneyko

Gwendolyn Berger

Luana Vaduva

Gwendolyn Berger
Conclusion


