Between OV and VO: exploring word-order variation and change across the Western Asian Transition Zone (WATZ)

Geoffrey Haig, University of Bamberg

Paper at OV-IS 2019
Corrélâts de l’ordre OV et structure de l’information
20-21.11.2019, Inalco, Paris
Overview of presentation

1. The VO / OV asymmetry in word order typology
2. The Western Asian Transition Zone (WATZ)
3. Data overview, coding and analysis
4. Results
   Role:
   a. DO (Direct Object)
   b. PLACE
   c. GOAL

Information structure and DO’s:

a. Definiteness
b. Weight
The VO / OV asymmetry in word order typology

<table>
<thead>
<tr>
<th>Harmonic Order</th>
<th>OV Basic Order</th>
<th>VO Basic Order</th>
</tr>
</thead>
<tbody>
<tr>
<td>XOV</td>
<td>VOX</td>
<td></td>
</tr>
<tr>
<td>OXV</td>
<td>VXO</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Disharmonic Order</th>
<th>OVX</th>
<th>XVO</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Harmonic Order</th>
<th>Disharmonic Order</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>VO</td>
<td>189 (98%)</td>
<td>2%</td>
</tr>
<tr>
<td>OV</td>
<td>68 (65%)</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>3 (XVO)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>37 (OVX)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>192</td>
<td></td>
</tr>
<tr>
<td></td>
<td>105</td>
<td></td>
</tr>
</tbody>
</table>

(Figures based on Hawkins 2008: 169, excludes languages with no dominant order)
• OV languages are vastly more tolerant of disharmonic orders (verbal complements on either side of the verb) than VO languages (almost exclusively VOX attested)

• No truly compelling explanation for this asymmetry is available.

• Has implications for understanding shifts between OV and VO (certain pathways are less likely than others)

• Focus here on **disharmonic orders in OV languages**
The Western Asian Transition Zone (WATZ)

- Cf. Stilo’s ‘Buffer’, or ‘overlap zone’, Haig 2017, Haig & Khan 2018
- Area of intersection and overlap at the border of two regions that are maximally distinct across a given set of typological parameters
Combined map of adpositional orders and VO/OV (WALS, Dryer 2013a, b)
Language genera in WATZ

- Iranian
- Indo-Aryan (Domari)
- Armenian
- Nakh-Daghestanian (Udi)
- Kartvelian
- Turkic
- Semitic
The data sample, coding, analysis

• Convenience sample of ongoing data compilation (target sample: 25-30 languages), from the project:

• *Post-predicate elements in Iranian and neighbouring languages: Inheritance, contact, and information structure* (Alexander von Humboldt Foundation, 2019-2021, Co-PI Mohammad Rasekhmahand)

• Most of the data here compiled and coded by Don Stilo

• Addition of two varieties of Turkic, coded for GOALs, RECIPIENTs, ADDRESSEEs
The data sample, coding, analysis

<table>
<thead>
<tr>
<th>FAMILY, GROUP</th>
<th>VARIETY, LOCATION</th>
<th>SOURCE, NOTES</th>
<th>N</th>
<th>%VX</th>
<th>%VG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semitic, NENA</td>
<td>Christian Barwar, N. Iraq</td>
<td>Khan 2008</td>
<td>961</td>
<td>74</td>
<td>95</td>
</tr>
<tr>
<td>Kartvelian</td>
<td>Georgian, colloquial standard</td>
<td>Recording Stilo</td>
<td>1324</td>
<td>42</td>
<td>68</td>
</tr>
<tr>
<td>Armenian, East</td>
<td>Lorri group, Shnogh, Armenia</td>
<td>Armenian folk tales, Vol. 8</td>
<td>715</td>
<td>23</td>
<td>47</td>
</tr>
<tr>
<td>Armenian, West</td>
<td>orig. Erzurum, relocated to Georgia</td>
<td>Armenian folk tales, Vol. 9</td>
<td>395</td>
<td>17</td>
<td>53</td>
</tr>
<tr>
<td>Iranian, NW</td>
<td>Mazanderani</td>
<td>Borjian &amp; Borjian 2008</td>
<td>676</td>
<td>16</td>
<td>63</td>
</tr>
<tr>
<td>Iranian, NW, Kurdish</td>
<td>Northern Kurdish, Lachin, Azerbaijan</td>
<td>Bakaev 1965</td>
<td>772</td>
<td>18</td>
<td>80</td>
</tr>
<tr>
<td>Turkic, Oghuz</td>
<td>Bayādestān, Iran</td>
<td>Bulut, in prep.</td>
<td>835</td>
<td>17</td>
<td>64</td>
</tr>
</tbody>
</table>

**ONLY CODED FOR GOALS, RECIP., ADDR.**

| Turkic, Oghuz       | Erzincan                                               | Reetz 2015                                         | 316  | na. | 29  |
| Turkic, Oghuz       | Tabriz                                                 | Reetz 2015                                         | 261  | na. | 65  |

**TOTAL NON-SUBJECT CONSTITUENTS, NOUNS & PRONOUNS:** 6255
The data sample, coding, analysis

All non-subject constituents coded in the sample text

Coded for the following variables:

• Genetic affiliation and location of the source variety
• Pronoun vs. noun
• Person and animacy
• Weight
• Role (direct objects also distinguish ‘definiteness’)
• Flagging
• Position (0=before; 1=after predicate)
Structural factors determining post-predicate placement

(1) Role

- DO
- PLACE
- GOAL
### PLACE

Any phrase indicating static location:

<table>
<thead>
<tr>
<th>iran_mazand</th>
<th>43a</th>
<th>(\varepsilon t=\varepsilon a \ \varepsilon q\varepsilon r\varepsilon n \ p\varepsilon l \ m\varepsilon \ j\varepsilon f \ d\varepsilon \varepsilon \varepsilon)</th>
<th>‘there was a single qeran in my pocket’</th>
</tr>
</thead>
<tbody>
<tr>
<td>nena_barwar</td>
<td>2</td>
<td>(h\varepsilon r=p\varepsilon \i\varepsilon \varepsilon \varepsilon \ v\varepsilon u \ ?\varepsilon \varepsilon r\varepsilon x\varepsilon)</td>
<td>‘he stayed in the mill.’</td>
</tr>
</tbody>
</table>
Note here and other graphics: ‘Turkish (Erzurum)’ should be ‘Turkish (Erzincan)’.
Post-predicate DO and PLACE, nouns only
Post-predicate DO, PLACE and GOAL, nouns only
Motion vs. caused motion GOALs

% post-predicate

Turkish (Erzurum), Turkish (Tabriz), Turkish (Bayat), Iranian (Mazand.), Kurmanji (Lachin), Armenian (Erzurum), Armenian (Lorri), Georgian, NENA (Barwar)

goals of motion
goals of caused motion
Summary DO

- **Iranian and Turkic**: little divergence from canonical OV order (<5%)
- **Armenian**: predominantly OV, but greater divergence
- **Georgian**: !!!???
- **Neo-Aramaic**: predominantly though not consistently VO, in accordance with expected Semitic values; see later slides
Summary PLACE

- **Iranian, Turkic, Armenian**: some data gaps, little difference to DO, slightly higher postpredicate tendency
- **Georgian and Neo-Aramaic**: more postpredicate than DO
Summary GOAL

- GOAL differs significantly from PLACE; no general pattern for spatial arguments
- Iranian, Turkic, except Erzurum: The OVG pattern typical for much of WATZ: consistent OV, dominant VG
- Armenian: less extreme form of WATZ pattern
- Georgian and Neo-Aramaic: mixed
- All languages: GOALs last (possibly universal?)
- Transitivity (caused vs. non-caused motion): irrelevant except in Turkic
Factors determining post-predicate placement

(1) Information structure and DO’s

Effects of *givenness* on position.

Three grades of givenness distinguished:

(a) Indefinite NP
(b) Definite NP
(c) Anaphoric pronoun
Givenness and DO

• For consistently OV languages (Turkic, Iranian):
  
  more definite > more post-predicate (weak effect)

• For consistently VO language (Neo-Aramaic):
  
  more definite > less post-predicate (strong effect)

• Mixed (Armenian, Georgian):
  
  mixed
Factors determining post-predicate placement

(2) Effects of weight on DO’s

• Probably the most discussed factor in variationist accounts of word order variation
• However, in naturalistic spoken data, there is little variation in weight on NP’s; most DO’s are NP’s consisting of 1-2 words.
• Two grades distinguished here:
  (a) light (one prosodic word)
  (b) heavy (three words or more)
<table>
<thead>
<tr>
<th>Kurmanji Kurdish of Lachin, Azerbaijan</th>
<th>English Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>dizidæ fikʰir dækin kö</em></td>
<td>“the thieves get the idea (lit: think) to steal <em>the apples</em> from him”</td>
</tr>
<tr>
<td><em>sevan žæ wi bídizin</em></td>
<td>“they pick up <em>that man’s apples</em> (and) leave”</td>
</tr>
<tr>
<td><em>sevedae wi mærika</em></td>
<td></td>
</tr>
<tr>
<td><em>hiltæynin dærævin</em></td>
<td></td>
</tr>
</tbody>
</table>
Post-predicate DO, nouns only

### Armenian Lorri

<table>
<thead>
<tr>
<th></th>
<th>OV</th>
<th>VO (postpredicate)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIGHT</td>
<td>105</td>
<td>12 (10.3%)</td>
</tr>
<tr>
<td>HEAVY</td>
<td>44</td>
<td>9 (17%)</td>
</tr>
<tr>
<td></td>
<td>149</td>
<td>21</td>
</tr>
</tbody>
</table>

The Fisher exact test statistic value is 0.2196. The result is not significant at p < .05.

### Georgian

<table>
<thead>
<tr>
<th></th>
<th>OV</th>
<th>VO (POSTPREDICATE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIGHT</td>
<td>125</td>
<td>87 (41%)</td>
</tr>
<tr>
<td>HEAVY</td>
<td>10</td>
<td>17 (63%)</td>
</tr>
<tr>
<td></td>
<td>135</td>
<td>104</td>
</tr>
</tbody>
</table>

The Fisher exact test statistic value is 0.0389. The result is significant at p < .05.
Combining role and information structure and DO

Frommer (1981) on Persian:

• Corpus of 2595 clause units of spoken colloquial Persian (in addition to other genres not considered here)

• Investigates (a) syntactic functions, and (b) the effects of focal vs. non-focal information status

• Wrts to (a): largely confirmed in our investigations of other Iranian languages
**Which roles are post-posed?** (% post-predicate, Frommer 1981)
% non-focal items in post-predicate position, according to role
Summary

OV

• Turkish and Iranian: little divergence from OV (<5%)
• minor, though consistent, effects of information structure (given DO’s leak rightwards)
• Converse tendency strongly observable in historically VO Neo-Aramaic
Summary

GOALs Last

• best predictor of post-predicate position is role:
  
  GOALS > other obliques > DO

• impervious to information structure

• rates of post-predicate GOAL sensitive to language origin, and geographic location: most frequent in Semitic and Iranian in contact with Semitic, southwestern periphery of WATZ (N. Iraq, Syria)
Summary

Why GOALs Last?

• Motivated by iconicity in syntax: event endpoints map onto clause endpoints (Haig 2014, to appear)

• Hence little effect of transitivity, or information structure

• Goals Last regularly manifested in VO languages, which only permit the order VOG as a basic order (Hawkins 2008 notes *VXO)

• In OV languages, in competition with Unified Head Directionality (all verbal arguments should be on the same side of the predicate) hence varying and areally sensitive degrees of manifestation
From OV to VO and back in WATZ: Sequence of changes

1 Historically VO languages in contact with OV:
   copula complements > definite DO > indefinite DO > GOAL
Wholesale shift documented in Neo Aramaic (e.g. Jewish dialect of Urmiye)

2 Historically OV in contact with VO:
   GOAL > other obliques > definite DO (>indefinite DO)
Wholesale shift not attested in the region (?), instead stable OVG

3 Armenian and Georgian ...
References
References (cont.)
Haig, G., Nils Schiborr, Stefan Schnell and Maria Vollmer. (in prep.) Introducing new referents: A corpus-based cross-linguistic perspective.


Schnell, S., Schiborr, N. & Haig, G. (under review). Efficiency in discourse processing: Does morphosyntax adapt to accommodate new referents?


