Stability and adaptivity of word order in the Western Asian Transition Zone: Evidence from West Iranian

Geoffrey Haig, University of Bamberg

Presentation at the workshop
Tracing contact in closely related languages

Universität Zürich, 19-20th November 2020
Transition zone

A geographic region of intersection of two (or more) established linguistic areas which are characterized by diametrically opposing values on a particular typological parameter.

A research focus within dialectology (Jeszensky et al 2018), but relatively little in areal typology, where the focus has been on establishing areas (and the theoretical and methodological challenges involved, see Bickel & Nichols 2006, Dedio et al 2019, among many others), rather than exploring the (comparatively small) number of languages at the overlaps.
The Western Asian Transition Zone (WATZ):

**VO & Prep vs. OV & Postp**

(Combination Features 83A Object/Verb, and 85A Adposition/NP, WALS)
West Iranian (Iranian, Indo-European)

- Internal grouping of Iranian remains controversial (Korn 2016)
- Difficulties frequently assigned to multiple contact settings, and to pervasive influence of Persian as a hegemonic language across the region for more than 2000 years
- Basic lexicon relatively uniform across West Iranian, remarkable differences in morphosyntax
- Most widely-dispersed language group in WATZ, hence of pivotal interest in the present project

Transition zones could, in principle, yield four outcomes:

<table>
<thead>
<tr>
<th>1. <strong>ABRUPT</strong></th>
<th>Prepositions</th>
<th>Postpositions</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. <strong>HYBRID FORMS</strong></td>
<td>Prepositions</td>
<td>Circumpositions</td>
</tr>
<tr>
<td>3. <strong>FUNCTIONAL SPLIT</strong></td>
<td>Prepositions</td>
<td>Prepositions in context X</td>
</tr>
<tr>
<td>4. <strong>FREQUENCY GRADED</strong></td>
<td>Prepositions 100%</td>
<td>Prepositions 50%</td>
</tr>
</tbody>
</table>
Unified Head Directionality (UHD)

OV/VO is commonly assumed to correlate closely with at least three other parameters:

- Gen/Noun
- Adp/NP
- Auxiliary/Lexical verb

However, very little research has been conducted on correlations with relative ordering of other verbal constituents, for example:

- Recipients / Verb
- Locations / Verb
- Goals / Verb
- Copular complements / copula
**General assumption:**

All verbal arguments (and to a lesser extent, adjuncts) linearize on the same side of the verb (Unified Head Directionality),

For example, OV would also imply:

- Goal-Verb
- Recipient-Verb
- Location-Verb etc.

But does this really happen in actual usage? And can these ‘minor word order’ patterns be utilized as diagnostics for contact linguistics?
Exploratory research questions

• When OV languages and VO languages co-exist in a transition zone, which verbal constituents exhibit greatest adaptivity, and which the highest stability?

• To what extent do minor word order patterns align with OV/VO ordering (as predicted by UHD) in contact situations?
Methodology

Corpus-based (or token-based) typological (CBT) to word order (Wälchli 2009, Futrell et al 2015, Levshina 2019), applied to contact linguistics (Stilo 2018, Haig, to appear).

Most CBT based on written corpora (in particular UD corpora, see Haitao 2010, Futrell et al 2015, among many others), and biased towards well-researched and generally standardized languages (Levshina 2019).

We draw on a corpus of spoken, non-elicited, non-scripted language, generally monologic (traditional narratives, oral history); special thanks to Don Stilo for making much of this material available).
Data overview: 25 doculects from WATZ

<table>
<thead>
<tr>
<th>family</th>
<th>doculect</th>
<th>tokens</th>
<th>source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iranian</td>
<td>Balochi, Koroshi</td>
<td>485</td>
<td>-- Silko (2018)</td>
</tr>
<tr>
<td>Iranian</td>
<td>Balochi, Turkmen</td>
<td>200</td>
<td>--</td>
</tr>
<tr>
<td>Iranian</td>
<td>Gorani, Gawaraju</td>
<td>--</td>
<td>-- Silko (2018)</td>
</tr>
<tr>
<td>Iranian</td>
<td>Kurmangi</td>
<td>224</td>
<td>--</td>
</tr>
<tr>
<td>Iranian</td>
<td>Mazanderani</td>
<td>584</td>
<td>--</td>
</tr>
<tr>
<td>Iranian</td>
<td>N. Kurdish, Lachin</td>
<td>642</td>
<td>--</td>
</tr>
<tr>
<td>Iranian</td>
<td>N. Kurdish, Muş</td>
<td>532</td>
<td>--</td>
</tr>
<tr>
<td>Iranian</td>
<td>S. Kurdish, Bijar</td>
<td>558</td>
<td>--</td>
</tr>
<tr>
<td>Iranian</td>
<td>Talyshi</td>
<td>475</td>
<td>--</td>
</tr>
<tr>
<td>Iranian</td>
<td>Dag Kushcu Tat</td>
<td>248</td>
<td>--</td>
</tr>
<tr>
<td>Iranian</td>
<td>Persian</td>
<td>909</td>
<td>Multi-Cast (Adibifar 2019)</td>
</tr>
<tr>
<td>Iranian</td>
<td>Vafsi</td>
<td>--</td>
<td>-- Silko (2018)</td>
</tr>
<tr>
<td>Iranian</td>
<td>Zaza</td>
<td>--</td>
<td>-- Silko (2018)</td>
</tr>
<tr>
<td>Armenian</td>
<td>Erzurum</td>
<td>363</td>
<td>--</td>
</tr>
<tr>
<td>Armenian</td>
<td>Lori</td>
<td>610</td>
<td>--</td>
</tr>
<tr>
<td>Armenian</td>
<td>Stepanakert</td>
<td>350</td>
<td>--</td>
</tr>
<tr>
<td>Turkic</td>
<td>Bayat</td>
<td>663</td>
<td>--</td>
</tr>
<tr>
<td>Turkic</td>
<td>Erzurum</td>
<td>530</td>
<td>--</td>
</tr>
<tr>
<td>Turkic</td>
<td>Tabriz</td>
<td>671</td>
<td>--</td>
</tr>
<tr>
<td>Semitic</td>
<td>NENA, Banwar</td>
<td>776</td>
<td>-- Silko (2018), Khan (2020)</td>
</tr>
<tr>
<td>Semitic</td>
<td>NENA, Jewish Umi</td>
<td>--</td>
<td>-- Silko (2018)</td>
</tr>
<tr>
<td>Semitic</td>
<td>NENA, Jewish Sanandaj</td>
<td>--</td>
<td>-- Silko (2018)</td>
</tr>
<tr>
<td>Semitic</td>
<td>NENA, Zakhro</td>
<td>--</td>
<td>-- Silko (2018)</td>
</tr>
<tr>
<td>Nakh-Daghestanian</td>
<td>Sanzhi Dargwa</td>
<td>500</td>
<td>Multi-Cast (Forker &amp; Schiborr 2019)</td>
</tr>
<tr>
<td>Nakh-Daghestanian</td>
<td>Tabasarancan</td>
<td>777</td>
<td>Multi-Cast (Bogomolowa, Ganenkov &amp; Schiborr 2020)</td>
</tr>
<tr>
<td>Germanic</td>
<td>English</td>
<td>3621</td>
<td>Multi-Cast (Schiborr 2015)</td>
</tr>
<tr>
<td>Tibeto-Burman</td>
<td>Jinghpaw</td>
<td>535</td>
<td>Multi-Cast (Kurabe 2020)</td>
</tr>
<tr>
<td>totals</td>
<td></td>
<td>14253</td>
<td></td>
</tr>
</tbody>
</table>

baseline OV and VO languages from outside WATZ for comparative purposes
Approx. locations of sample doculects

- **Iranian (13)**
- **Armenian (3)**
- **Turkic (3)**
- **Nakh-Daghestanian (2)**
- **Semitic (4)**
Procedure

- Texts are digitalized/recordings are transcribed, translated, and segmented into syntactically coherent utterance units
- **Non-subject, referential**, constituents are identified and coded for:

9 predictor variables:

**ORIGIN AND SETTING VARIABLES:**

- Genetic affiliation
- Place of socialization of the speaker
- Text ID
- Clause ID

**LINGUISTIC VARIABLES:**

- Pronominal/nominal
- Animacy
- Weight
- Role
- Flagging

1 dependent variable:

Position relative to governing predicate: 0=before, 1=after
Today's focus: Rates of postverbal placement, across different language families and locations in WATZ, for the following roles:

- **Copula complement:** *Martha is a doctor*
- **Direct object:** *She won a million dollars in the lottery*
- **Addressee:** *She told her partner about it*
- **Recipient:** *and gave her mother a new car*
- **Goal:** *They drove it to Newcastle.*

(and later a couple of others)

**Rationale:** Word order parameter settings are seldom categorical, but frequency graded (Levshina 2019)

Frequency profiles shift adaptively in contact situations; in order to detect these shifts, we need corpus data.
Dominant OV languages in WATZ, except Iranian
Summary OV (not Iranian):

- OV is consistent in Armenian and Turkic, less so in Nakh-Daghestanian (the Caucasus is a puzzle and does not fit particularly well with the larger-scale picture)
- Goals Last
- Not necessarily the case in OV languages outside WATZ (cf. Jinghpaw)
- Rates of post-verbal goals roughly correspond to degree of contact with Iranian

First movers towards post-verbal placement:

Goals > Recipient > Addressee > DO > Cop
Historically VO languages in WATZ (Neo-Aramaic (Semitic, Afro-Asiatic)
Summary VO languages:

- The first movers away from post-verbal placement are **copular complements**, which initially become mobile indicators for varying information structural configurations (Khan 2018).
- Then **definite direct objects**, including pronominal DO’s, which are likewise initially fronted for information structural purposes
- The most likely to remain post-verbal are recipients etc.
- Goals Last

First movers towards pre-verbal placement:

Cop > Def. DO > addressee > recipient > goal

*[However, the clause-final copular morphemes are generally historical innovations, which tend to replicate copular constructions in neighbouring OV languages, so the notion of ‘stable’ is not particularly apt here]*
Iranian languages (historically OV) in WATZ

Kumzari

Balochi Turkmenistan
Summary W. Iranian languages:

- Goals last, with rates of post-verbal goals approximately varying along a northeast-southwest cline.
- Placement of Addressees and Recipients ditto
- Remain stubbornly OV, but Kumzari has shifted pronominal objects to VO (actually in violation of Greenberg #25)

First movers towards post-verbal placement in Iranian

Goals > Recipient > Addressee > DO > Cop
Convergence to OVG: Kumzari (Iranian, grey) and Neo-Aramaic, Sanandaj (Semitic, blue)
Is it all about arguments vs. adjuncts (cf. Hawkins 2008)?

No ... **Goals Last** is still evident when we compare Goals with other peripheral arguments
Research questions revisited

- When OV languages and VO languages co-exist in a transition zone, which verbal constituents exhibit greatest adaptivity, and which the highest stability?

  For historically OV languages: Preverbal (indefinite) direct objects and copular complements are the most stable, preverbal goals the least stable.

  For historically VO languages: Postverbal goals are the most stable, definite (including pronominal) direct objects are the least stable.

This is an unexpected finding given the broad consensus on the stability of OV/VO parameter, and the apparent global preference for OV>VO, but not vice versa (Gell-Mann and Ruhlen 2011)
• To what extent do minor word order patterns align with OV/VO ordering (as predicted by UHD) in contact situations?
• Only weakly. Within WATZ, most of Iranian, and varieties of Turkic and Semitic languages in intense contact with Iranian, converge on a **disharmonic OVG word order**, possibly quite rapidly.
• Frequency patterns for (some) minor word order patterns provide a surprisingly sensitive and hitherto largely ignored diagnostic for contact effects
Prospects:

- Increase density of sample
- Ramp up the stats
- Investigate interaction with flagging (already coded)
- Test findings in other VO/OV transition zones
- Is UHD a product of gradual accretion towards the centres of relatively stable areas, which is relatively freely abandoned in contact scenarios at the peripheries?
- Why Goals Last? Sole plausible explanation is the iconic mapping of semantic endpoints to clausal endpoints as one (of several) competing motivations in determining word order (Haig, 2014), comparable, though less pervasive, to Agent First (Bickel et al 2015; Riesberg et al 2019): Check against L1 acquisition and (really) large corpus data.
Acknowledgements

• Don Stilo for many years of collaboration, inspiration, and mountains of data
• Nils Schiborr for data handling, incorporating Multi-CAST corpora, and graphics
• Mahîr Doğan for data handling and coding Erzurum Turkish
• Masoud Mohammadirad for coding Southern Kurdish
• Maryam Nourzaei for coding Koroshi Balochi
• Colleagues from the Post-Predicate-in-Iranian languages project
• DFG and Alexander-von-Humboldt Foundation for funding
References


Bickel, Balthasar & Johanna Nichols. 2006. Oceania, the Pacific Rim, and the theory of linguistic areas. BLS 32, No 2 2006. DOI: http://dx.doi.org/10.3765/bls.v32i2.3488


