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
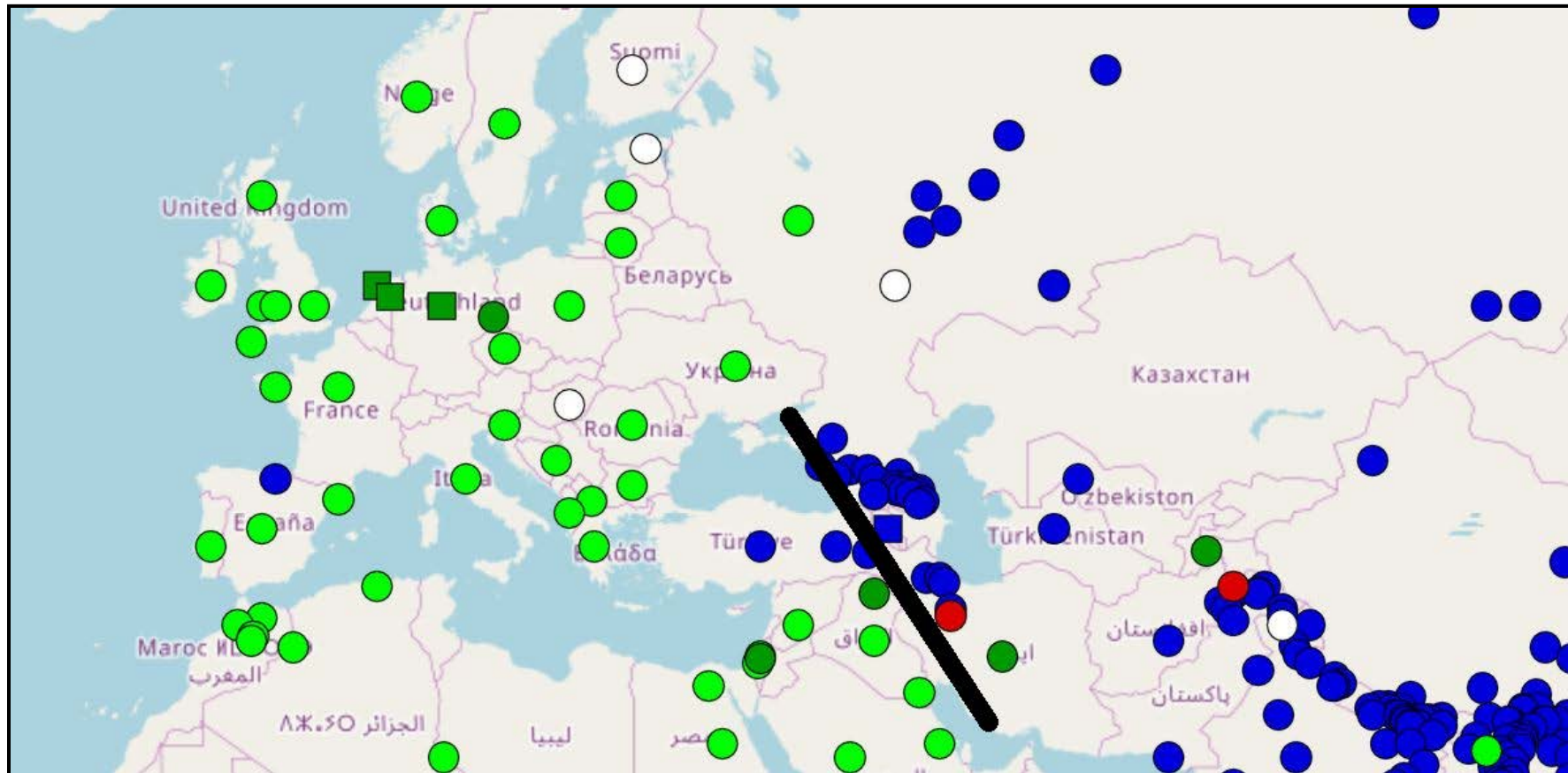
(Combination Features 83A Object/Verb, and 85A Adposition/NP, WALS)

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

The current project builds on previous research (Stilo 2005, 2006, 2012, 2018; Haig 2014, 2017, to appear; Haig & Khan 2018), compiling the largest corpus to date and adopting methodologies of Corpus Based Typology.

Transition zones could, in principle, yield four outcomes:

1. ABRUPT	Prepositions		Postpositions
2. HYBRID FORMS	Prepositions	Circumpositions	Postpositions
3. FUNCTIONAL SPLIT	Prepositions	Prepositions in context X Postpositions in context Y	Postpositions
4. FREQUENCY GRADED	Prepositions 100%	Prepositions 50% Postpositions 50%	Postpositions 100%

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, Futtrell 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

	A	B	C	D
1	**family**	**doculect**	**tokens**	**source**
2	Iranian	Balochi, Koroshi	485	
3	Iranian	Balochi, Turkmen	200	
4	Iranian	Gorani, Gawaraju	--	Stilo (2018)
5	Iranian	Kumzari	224	
6	Iranian	Mazanderani	584	
7	Iranian	N. Kurdish, Lachin	642	
8	Iranian	N. Kurdish, Muş	532	
9	Iranian	S. Kurdish, Bijar	558	
10	Iranian	Talyshi	475	
11	Iranian	Dag Kushcu Tat	248	
12	Iranian	Persian	909	Multi-CAST (Adibifar 2019)
13	Iranian	Vafsi	--	Stilo (2018)
14	Iranian	Zaza	--	Stilo (2018)
15	Armenian	Erzurum	363	
16	Armenian	Lorri	610	
17	Armenian	Stepanakert	350	
18	Turkic	Bayat	663	
19	Turkic	Erzurum	530	
20	Turkic	Tabriz	671	
21	Semitic	NENA, Barwar	776	
22	Semitic	NENA, Jewish Urmi	--	Stilo (2018), Khan (2020)
23	Semitic	NENA, Jewish Sanandaj	--	Stilo (2018)
24	Semitic	NENA, Zakho	--	Stilo (2018)
25				
26	Nakh-Daghestanian	Sanzhi Dargwa	500	Multi-CAST (Forker & Schiborr 2019)
27	Nakh-Daghestanian	Tabasaran	777	Multi-CAST (Bogomolova, Ganenkov & Schiborr 2020)
28	Germanic	English	3621	Multi-CAST (Schiborr 2015)
29	Tibeto-Burman	Jinghpaw	535	Multi-CAST (Kurabe 2020)
30				
31		totals	14253	

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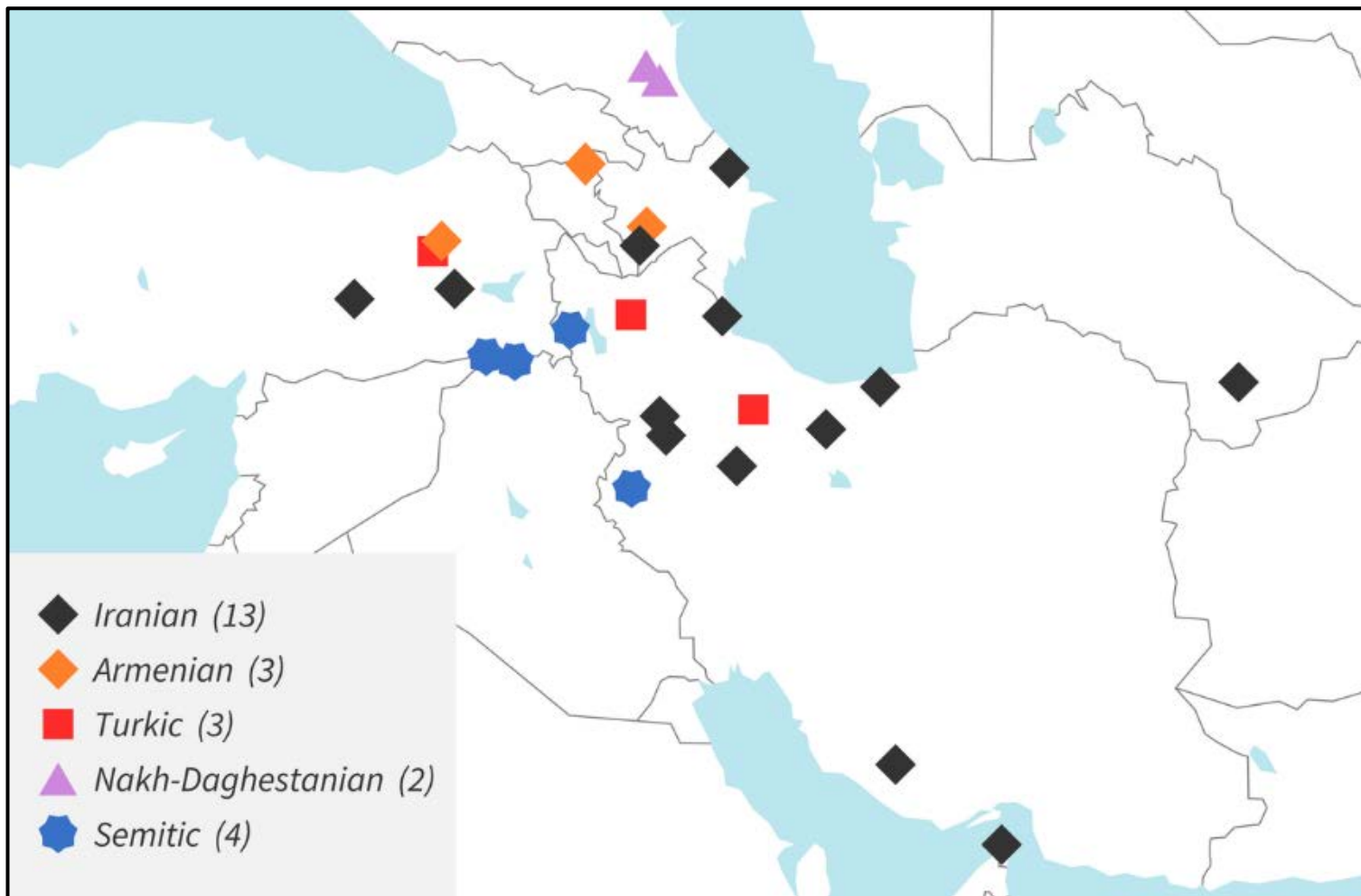
Approx. locations of sample doculects

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

Todays 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
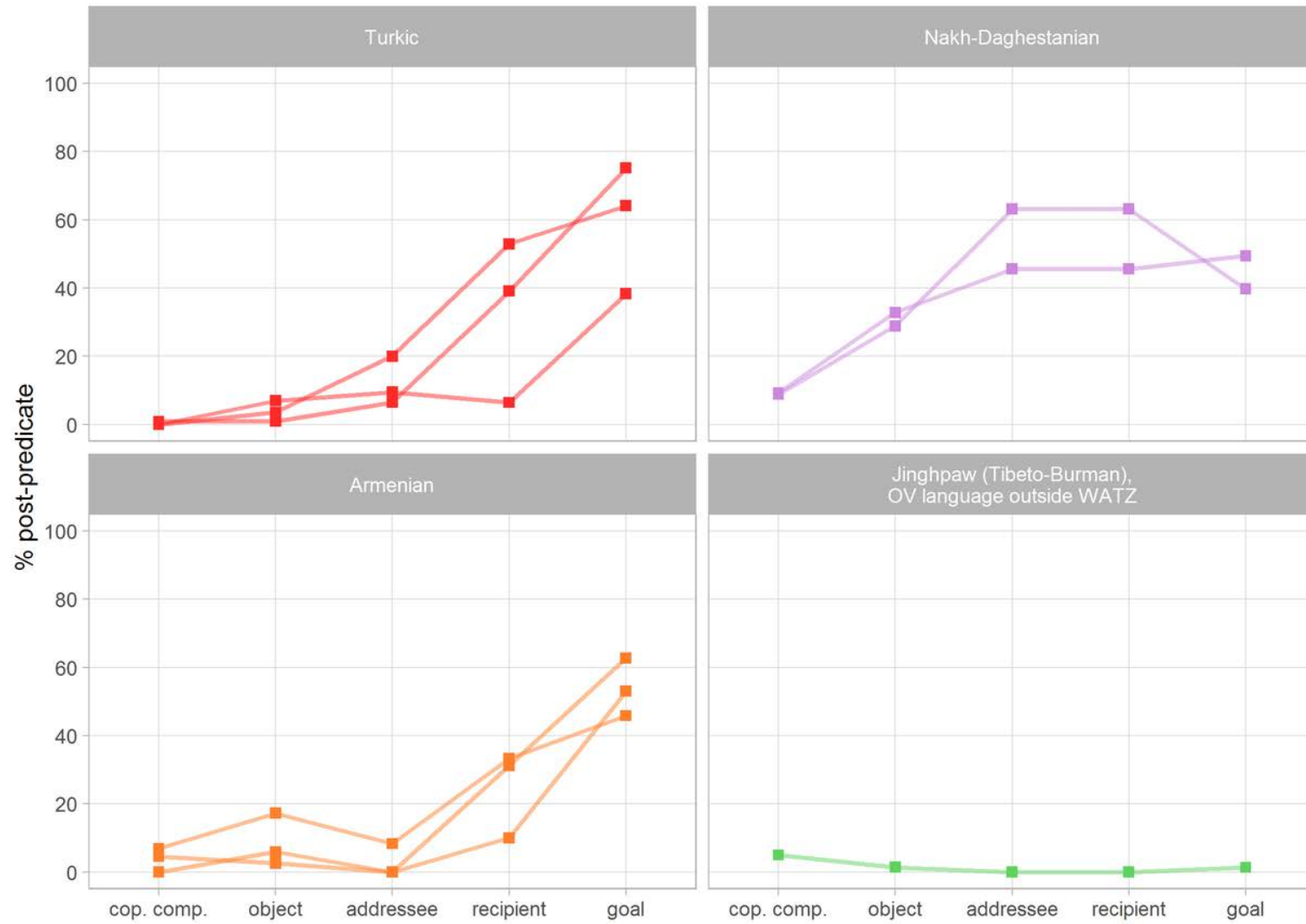
Summary OV (not Iranian):

- 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)
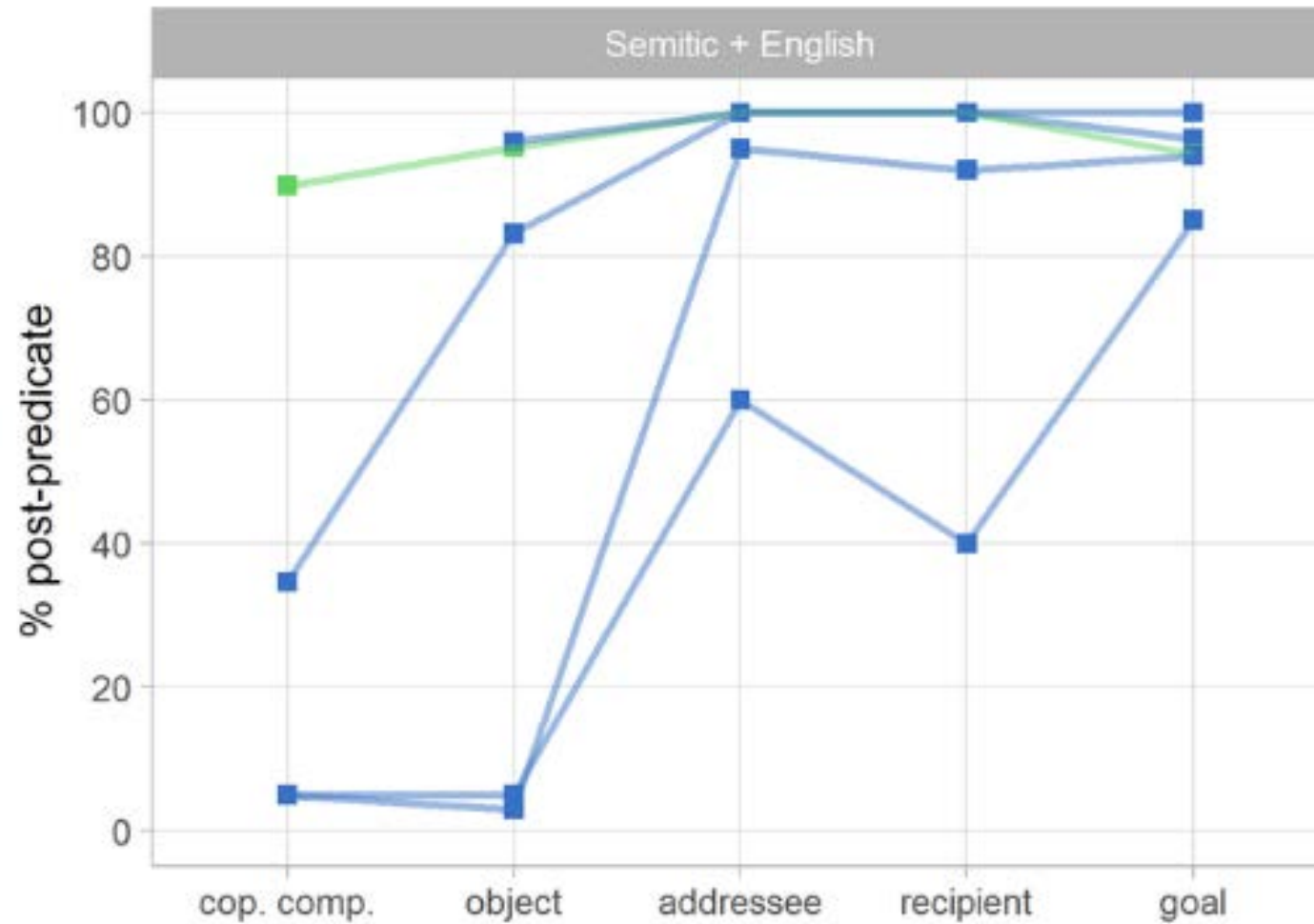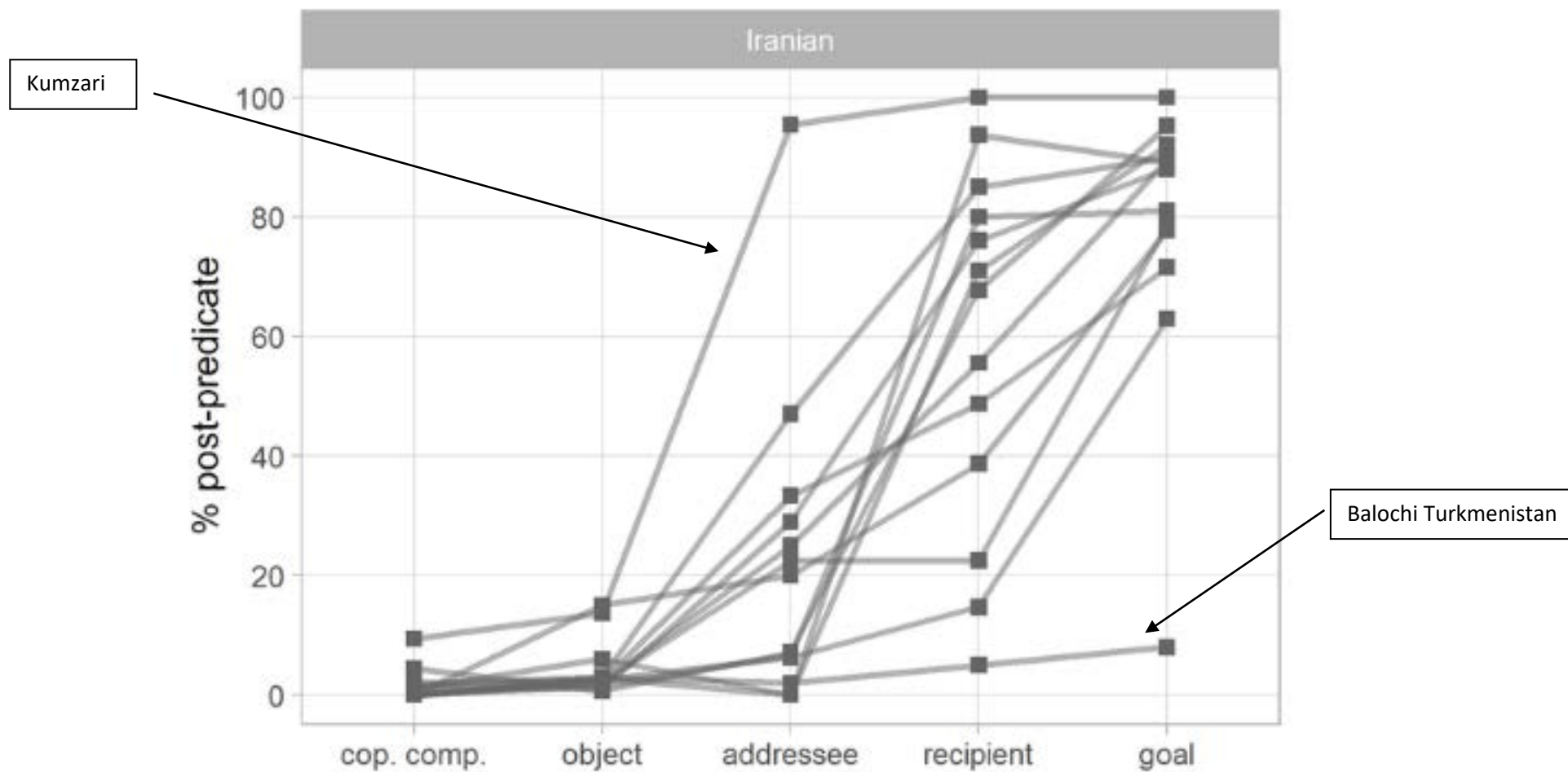
Summary VO languages:

- 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

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)
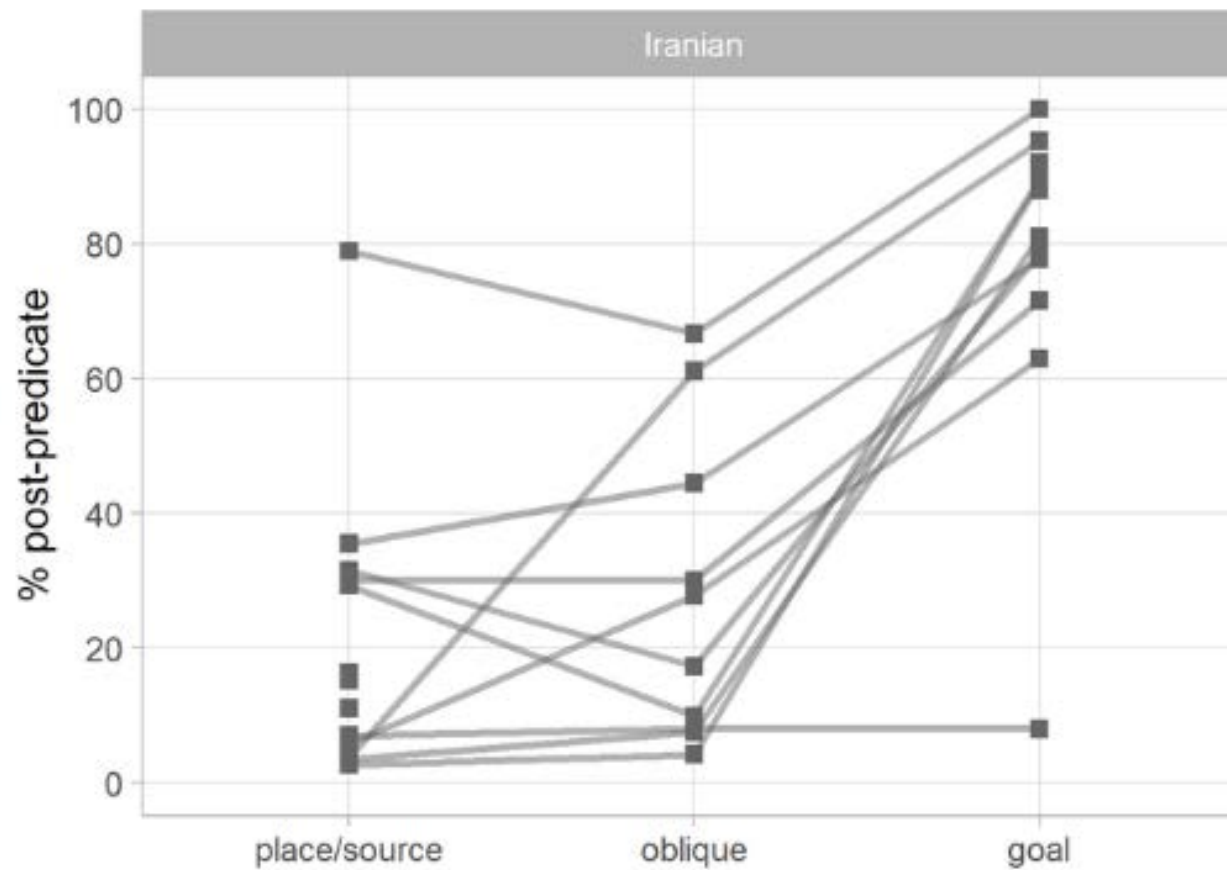
First movers towards post-verbal placement in Iranian

Goals > Recipient > Addressee > DO > Cop

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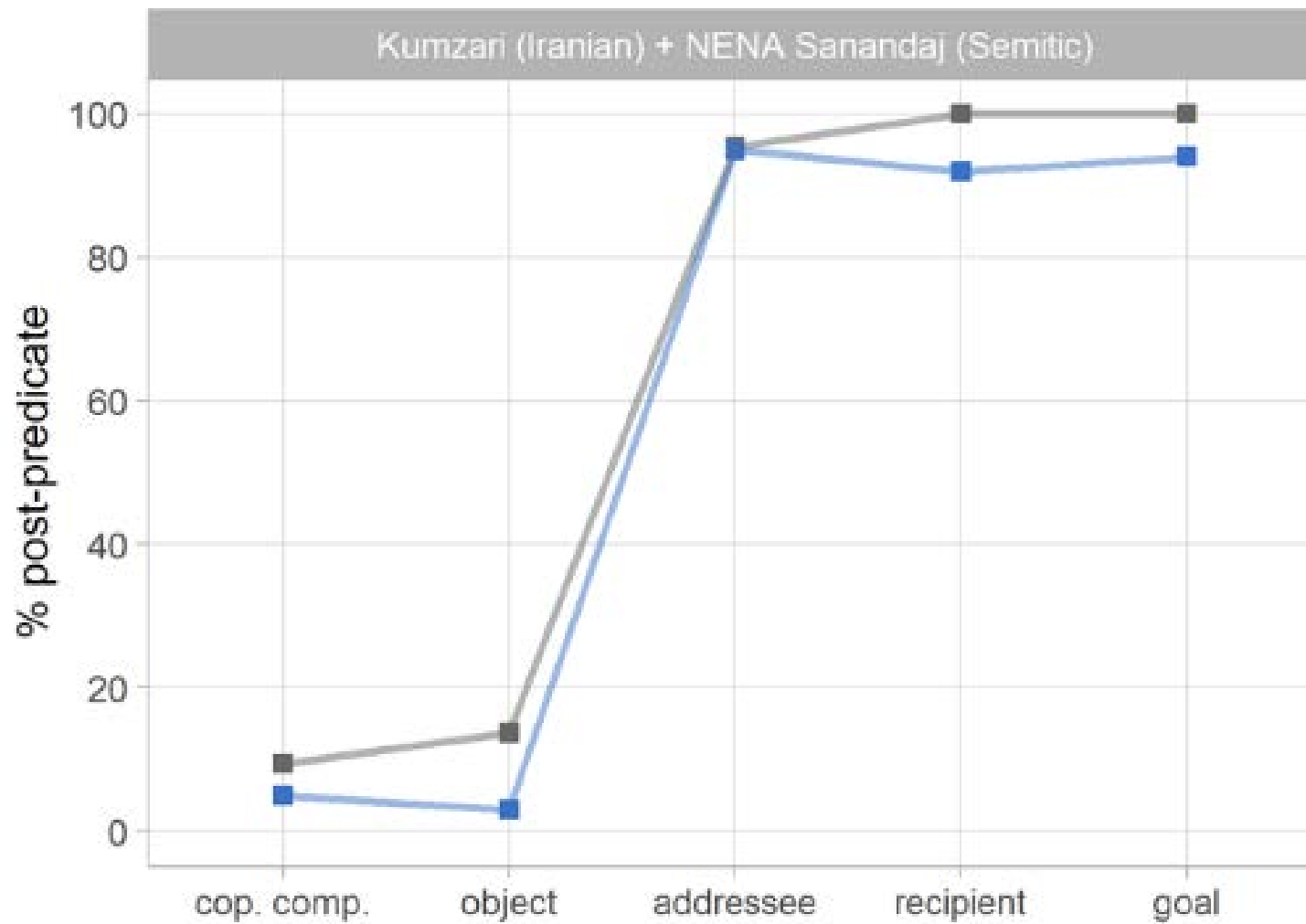
Is it all about arguments vs. adjuncts (cf. Hawkins 2008)?

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 B, Witzlack-Makarevich A, Choudhary KK, Schlesewsky M, Bornkessel-Schlesewsky I. 2015. The Neurophysiology of Language Processing Shapes the Evolution of Grammar: Evidence from Case Marking. PLoS ONE 10(8): e0132819. doi:10.1371/journal.pone.0132819

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

Cathcart, Chundra, Gerd Carling, Filip Larsson, Niklas Johansson & Erich Round. 2018. Areal pressure in grammatical evolution. An Indo-European case study. *Diachronica* 35(1), 1-34.

El Zarka, Dina & Sandra Ziagos (2020) The Beginnings of Word OrderChange in the Arabic Dialects of Southern Iran in Contact with Persian: A Preliminary Studyof Data from Four Villages in Bushehr and Hormozgan, Iranian Studies, 53:3-4, 465-488, DOI:10.1080/00210862.2019.1690433

Gell-Mann, Murray and Merrit Ruhlen. 2011. The origin and evolution of word order. PNAS. October 18, 2011. vol. 108, no. 42, 17290–17295

Dedio, Stefan, Peter Ranacher & Paul Widmer. 2019. Evidence for Britain and Ireland as a linguistic area. *Language* 95(3).498–522, *2019*.

Haig, Geoffrey. 2014. Verb-goal (VG) word order in Kurdish and Neo-Aramaic: Typological and areal considerations. In Khan, Geoffrey and Lidia Napiorkowska (eds.) *Neo-Aramaic and its Linguistic Context*, 407-425 New York: Gorgias Press.

Haig, Geoffrey. 2017a. Western Asia: East Anatolia as a transition zone. In Hickey, Raymond (ed.), *The Cambridge Handbook of Areal Linguistics*, 396-423. Cambridge: CUP.

Haig, Geoffrey. 2019. Between OV and VO: exploring word-order variation and change across the Western Asian Transition Zone (WATZ). Invited talk at OV-IS 2019 *Corrélâts de l'ordre OV et structure de l'information*. 20-21.11.2019, Inalco, Paris.

Haig, Geoffrey & Geoffrey Khan. 2018. Introduction. In: Haig, Geoffrey & Geoffrey Khan (eds.) *The languages and linguistics of Western Asia. An areal perspective*, 1-29. Berlin: Mouton de Gruyter.

Haig, Geoffrey & Stefan Schnell (eds.) 2020. *Multi-CAST: Multilingual corpus of annotated spoken texts*. (multicast.aspra.uni-bamberg.de/) (accessed 01.11.2020)

Hawkins, John. 2008. An asymmetry between VO and OV languages. The ordering of obliques. In: Corbett, Greville and Michael Noonan (eds.). *Case and Grammatical Relations: Studies in Honor of Bernard Comrie,* 167-190. Amsterdam: Benjamins.

Jeszenszky, Péter, Philipp Stoeckle, Elvira Glaser & RobertWeibel. 2018. A gradient perspective on modeling interdialectal transitions. Journal of Linguistic Geography (2018) 6, 78–99.

Meyer, Robin. Forthcoming. Iranian Syntax in Classical Armenian. Morphosyntactic alignment change and other cases of pattern replication. Oxford: Oxford University Press.

Multi-CAST: Haig, Geoffrey & Stefan Schnell (eds.) 2020. *Multi-CAST: Multilingual corpus of annotated spoken texts.* (multicast.aspra.uni-bamberg.de/) (accessed 01.11.2020)

Sonja Riesberg, Kurt Malcher, Nikolaus P. Himmelmann. How universal is agent-first? Evidence from symmetrical voice languages. *Language*, Volume 95, Number 3, September 2019, 523-561.

Schnell, Stefan & Geoffrey Haig. 2014. Assessing the relationship between object topicalisation and the grammaticalisation of object agreement. In Gawne, Lauren & Jill Vaughan (eds.) *Selected Papers from the 44th Conference of the Australian Linguistic Society, 2013,* 102-125.

Seifart, Frank. 2017. Patterns of affix borrowing in a sample of 100 languages. Journal of Historical Linguistics. 7:3, 389–431

Stilo, Donald. 2005. Iranian as buffer zone between the universal typologies of Turkic and Semitic. In Éva Ágnes Csató, Bo Isaksson & Carina Jahani (eds.), *Linguistic convergence and areal diffusion: Case studies from Iranian, Semitic and Turkic*, 35–63. London & New York: Routledge Curzon.

Stilo, Donald. 2006. Circumpositions as an areal response: The case study of the Iranian zone. In: Johanson, Lars & Christiane Bulut (eds.) *Turkic-Iranian contact areas. Historical and linguistics aspects*, 310-333. Wiesbaden: Harrassowitz.

Stilo, Donald. 2012. Intersection zones, overlapping isoglosses, and 'Fade-out/Fade-in' phenomena in Central Iran. In: Behrad Aghaei & M. R. Ghanoonparvar (eds.), *Iranian Languages and Culture: Essays in Honor of Gernot Ludwig Windfuhr,* 3-33. Costa Mesa: Mazda Publishers.

Stilo, Donald. 2014. Areal factors and postverbal Patients in spoken Azerbaijani: a corpus-based study. In Demir, Nurettin, Birsel Karakoç & Astrid Menz (eds.) Turkology and linguistics. Festschrift Éva Ágnes Csató. 417-430. Ankara: Hacettepe Üniversitesi.

Stilo, Donald. 2018. Preverbal and postverbal peripheral arguments in the Araxes-Iran linguistic area. Invited lecture at the conference *Anatolia-Caucasus-Iran: Ethnic and Linguistic Contacts,* Yerevan University, 10-12 May 2018.

Tai, James H.Y. 1985. Temporal sequence and Chinese word order. In: Haiman, John (ed.) Iconicity in syntax, 49-72. Amsterdam: Benjamins.

Wichmann, Søren and Eric Holman. 2009. Temporal stability of linguistic typological features. Lincom Europa.