

# 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

	<b>OV BASIC ORDER</b>	<b>VO BASIC ORDER</b>
<b>HARMONIC ORDER</b> (O AND X ON SAME SIDE OF V)	XOV	VOX
	OXV	VXO
<b>DISHARMONIC ORDER</b>	OVX	XVO

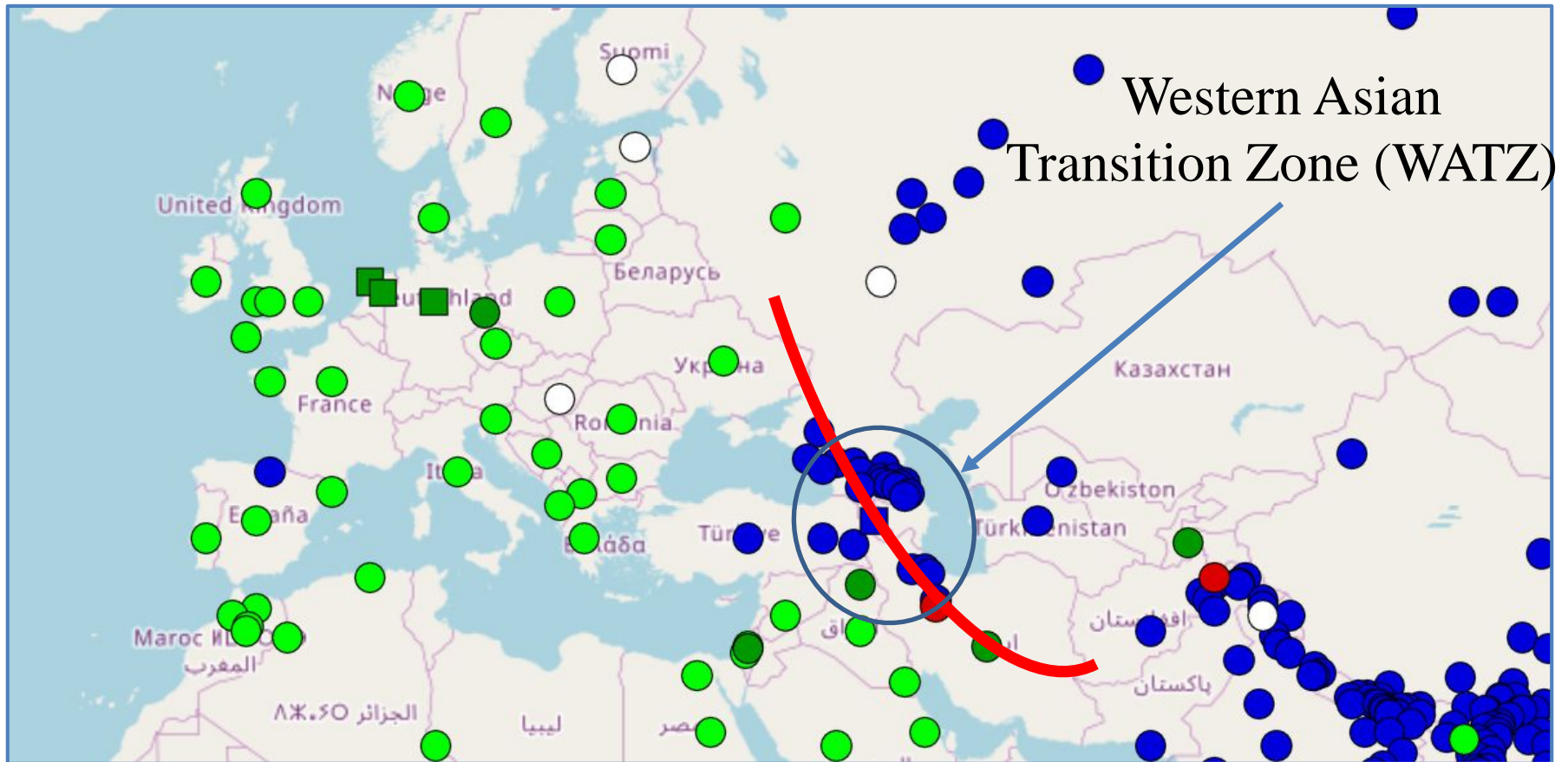
	<b>HARMONIC ORDER</b>	<b>DISHARMONIC ORDER</b>		<b>TOTAL</b>
<b>VO</b>	189 (98%)	3 (XVO)	2%	192
<b>OV</b>	68 (65%)	37 (OVX)	35%	105

(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

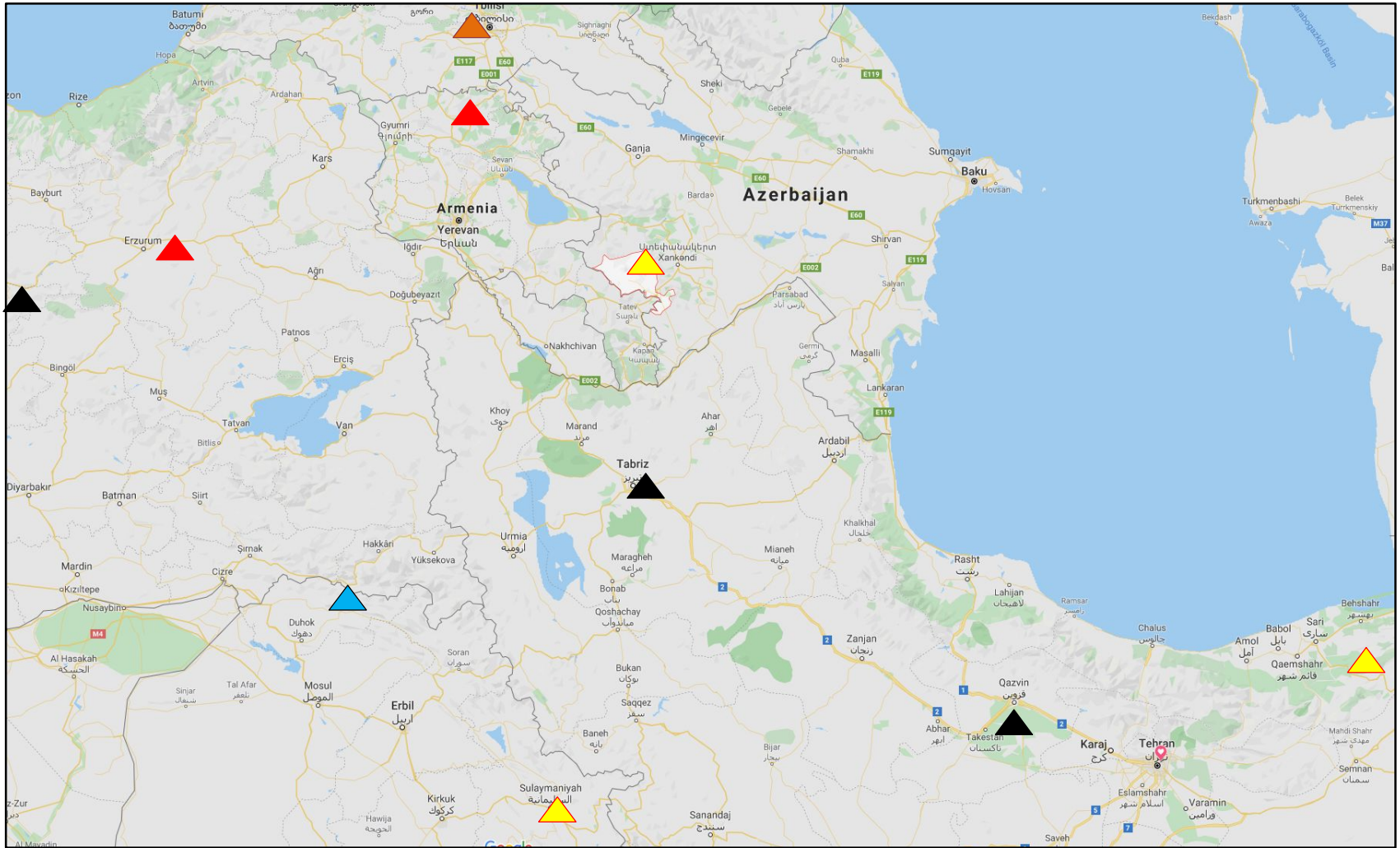
## The data sample, coding, analysis

- 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

FAMILY, GROUP	VARIETY, LOCATION	SOURCE, NOTES	N	% <u>VX</u>	%VG
Semitic, NENA	Christian Barwar, N. Iraq	Khan 2008	961	74	95
Kartvelian	Georgian, colloquial standard	Recording <u>Stilo</u>	1324	42	68
Armenian, East	Lorri group, Shnogh, Armenia	Armenian folk tales, Vol. 8	715	23	47
Armenian, West	orig. Erzurum, relocated to Georgia	Armenian folk tales, Vol. 9	395	17	53
Iranian, NW	Mazanderani	Borjian & Borjian 2008	676	16	63
Iranian, NW, Kurdish	Northern Kurdish, Lachin, Azerbaijan	Bakaev 1965	772	18	80
Turkic, Oghuz	Bayādestān, Iran	Bulut, in prep.	835	17	64
	<b>ONLY CODED FOR GOALS, <u>RECIP.</u>, <u>ADDR.</u></b>				
Turkic, Oghuz	Erzincan	Reetz 2015	316	<u>na.</u>	29
Turkic, Oghuz	Tabriz	Reetz 2015	261	<u>na.</u>	65
	<b>TOTAL NON-SUBJECT CONSTITUENTS, NOUNS &amp; PRONOUNS:</b>		<b>6255</b>		



▲ Armenian   
 ▲ Semitic   
 ▲ Turkic   
 ▲ Iranian   
 ▲ Kartvelian

## 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

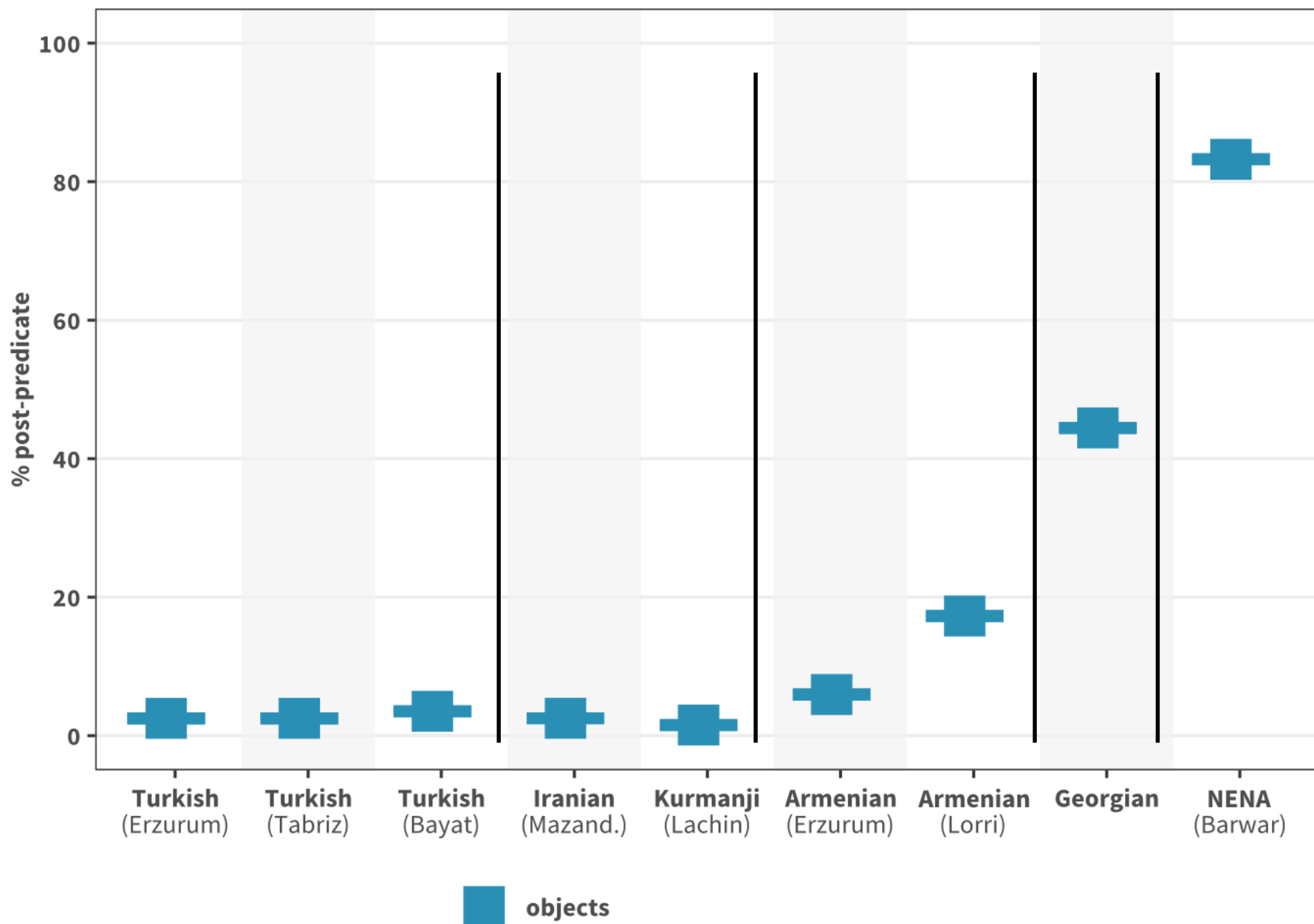
## PLACE

Any phrase indicating static location:

iran_mazand	43a	<i>εt=ta qæron pul me</i> <i>jif dæie</i>	'there was a single qeran <b>in my pocket</b> '
-------------	-----	--	--

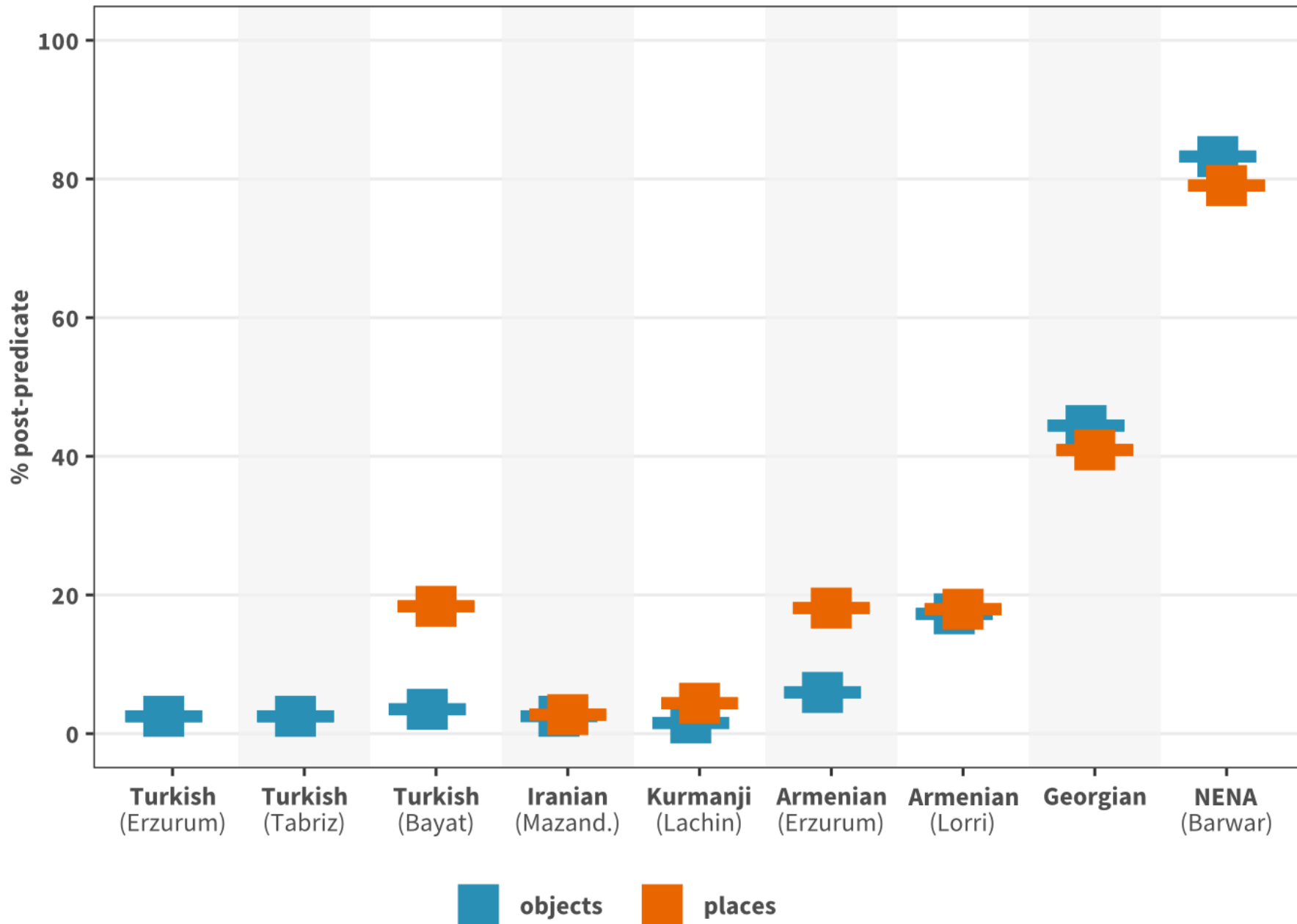
nena_barwar	2	<i>har=p'ışle gu=ʔərxe</i>	'he stayed <b>in the mill.</b> '
-------------	---	----------------------------	----------------------------------

# Post-predicate DO, nouns only

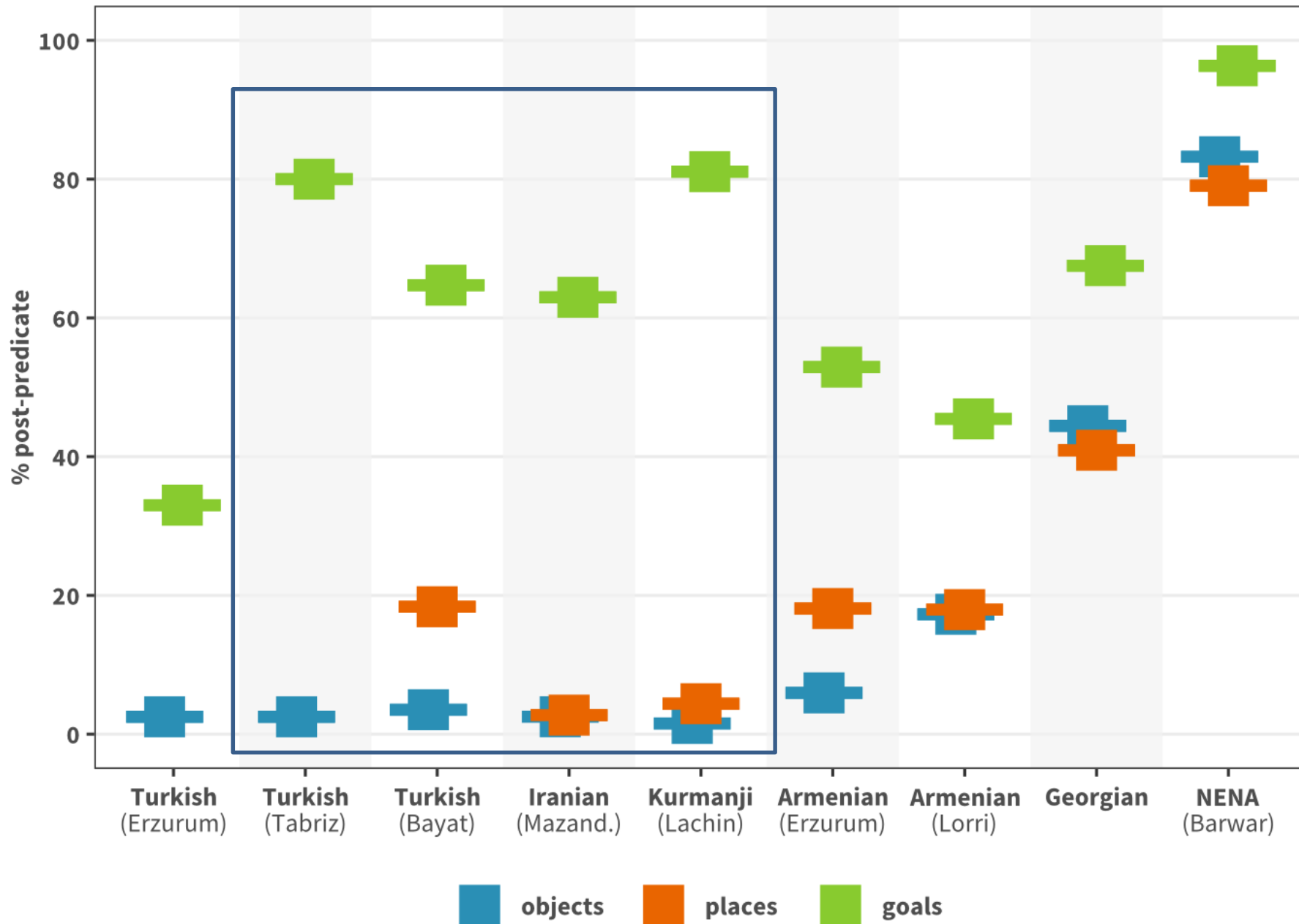


Note here and other graphics: *'Turkish (Erzurum)'* should be *'Turkish (Erzincan)'*

# Post-predicate DO and PLACE, nouns only

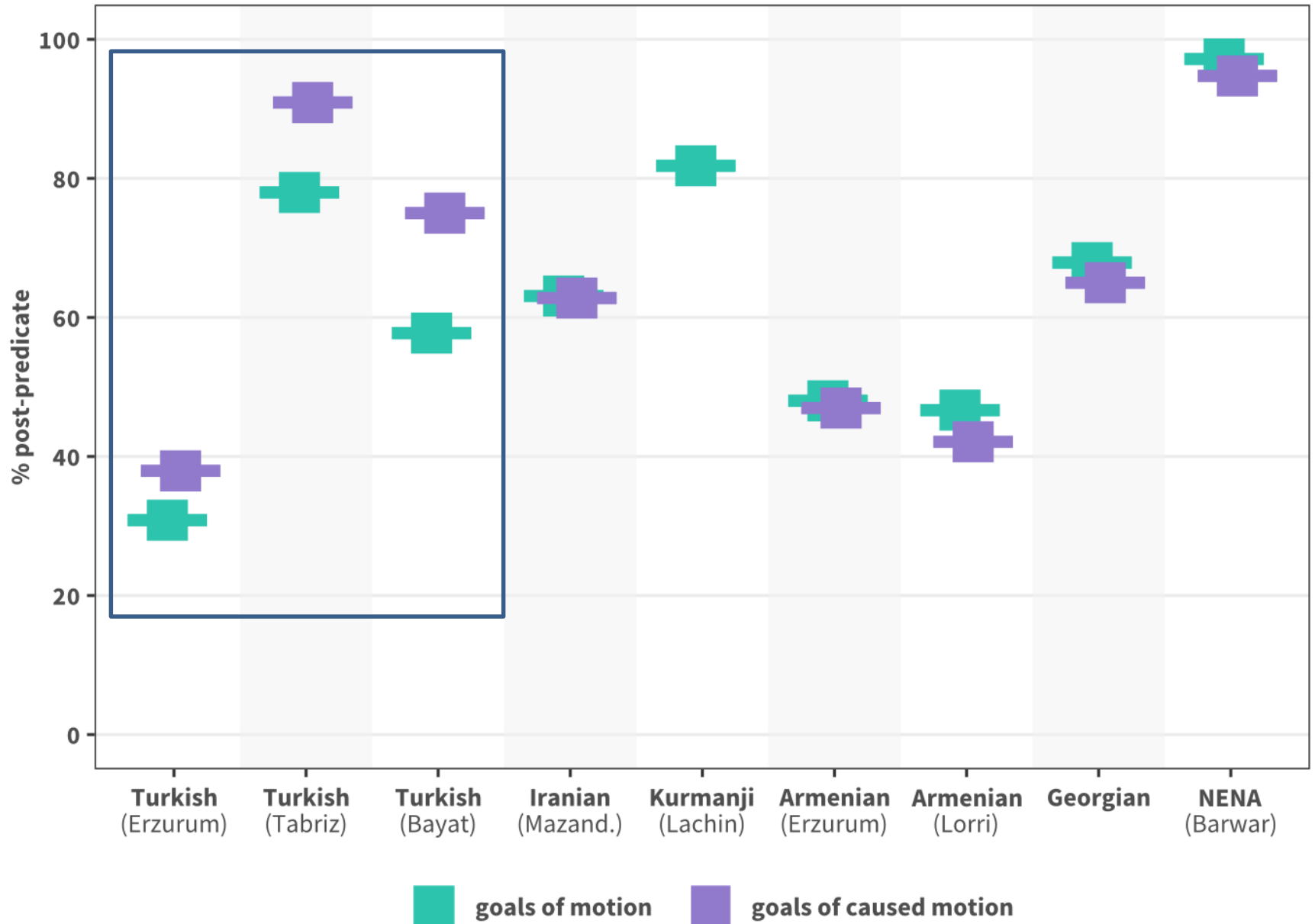


# Post-predicate DO, PLACE and GOAL, nouns only





# Motion vs. caused motion GOALS



## 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

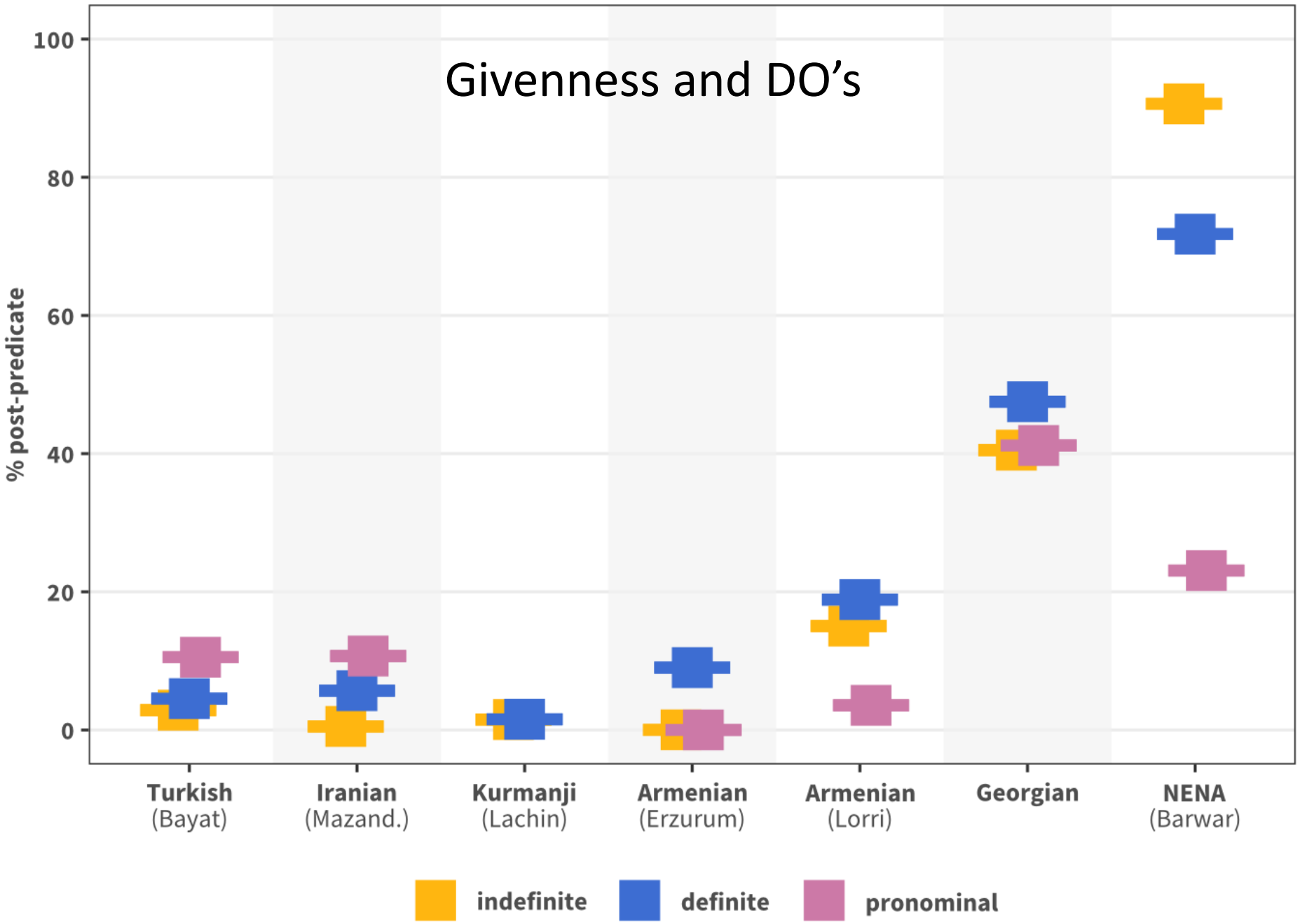
## (1) Information structure and DO's

Effects of **givenness** on position.

Three grades of givenness distinguished:

- (a) Indefinite NP
- (b) Definite NP
- (c) Anaphoric pronoun

# Givenness and DO's



## 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)



LIGHT

<i>dizidæ fik<sup>h</sup>ir dækin kö sevan žæ wi bídizin</i>	“the thieves get the idea (lit: think) to steal <b>the apples</b> from him”
<i>sevedæ wi mærika hiltæynin dærævin</i>	“they pick up <b>that man’s apples</b> (and) leave”

HEAVY

(Kurmanji Kurdish of Lachin, Azerbaijan)

# Post-predicate DO, nouns only

## Armenian Lorri

	OV	VO (postpredicate)	
LIGHT	105	12	10.3%
HEAVY	44	9	17%
	149	21	170

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

## Georgian

	OV	VO (POSTPREDICATE)	
LIGHT	125	87	41%
HEAVY	10	17	63%
	135	104	239

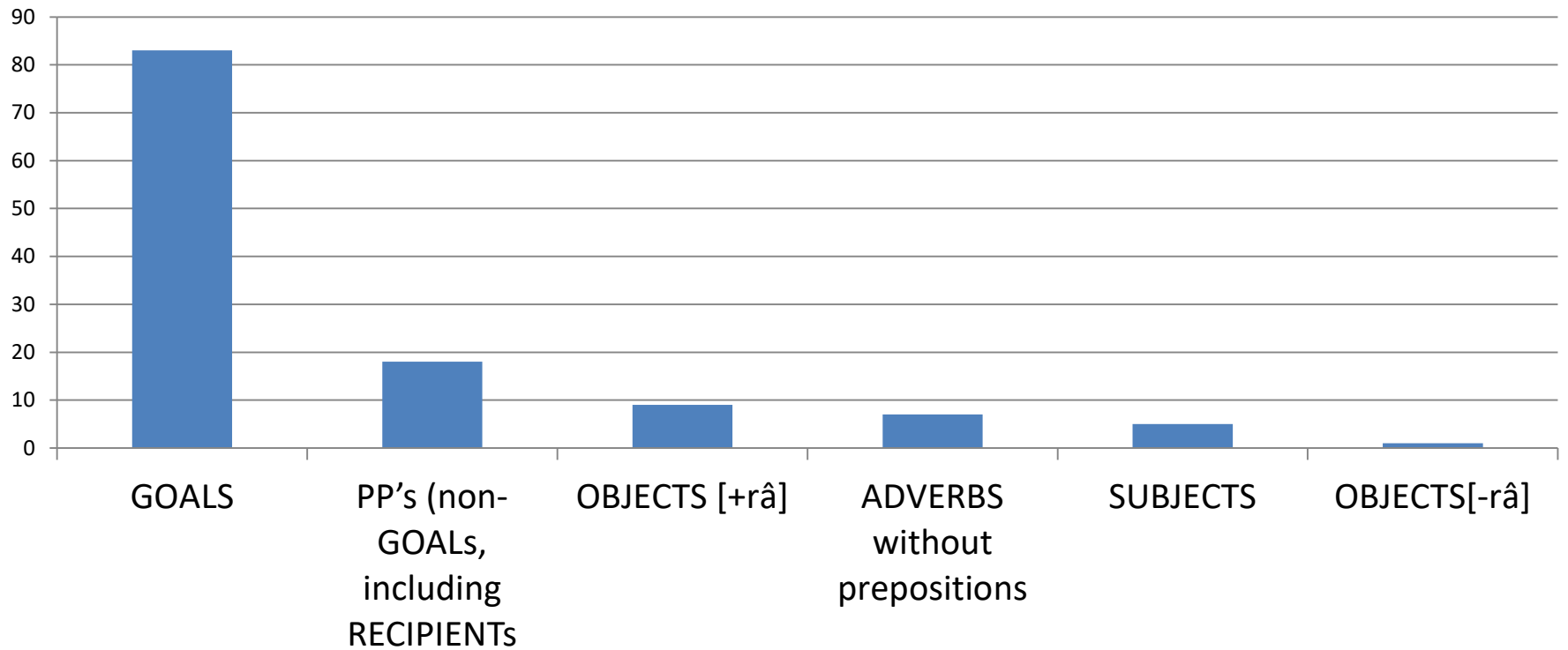
The Fisher exact test statistic value is 0.0389. The result is significant at  $p < .05$ .

## Combining role and information structure and DO

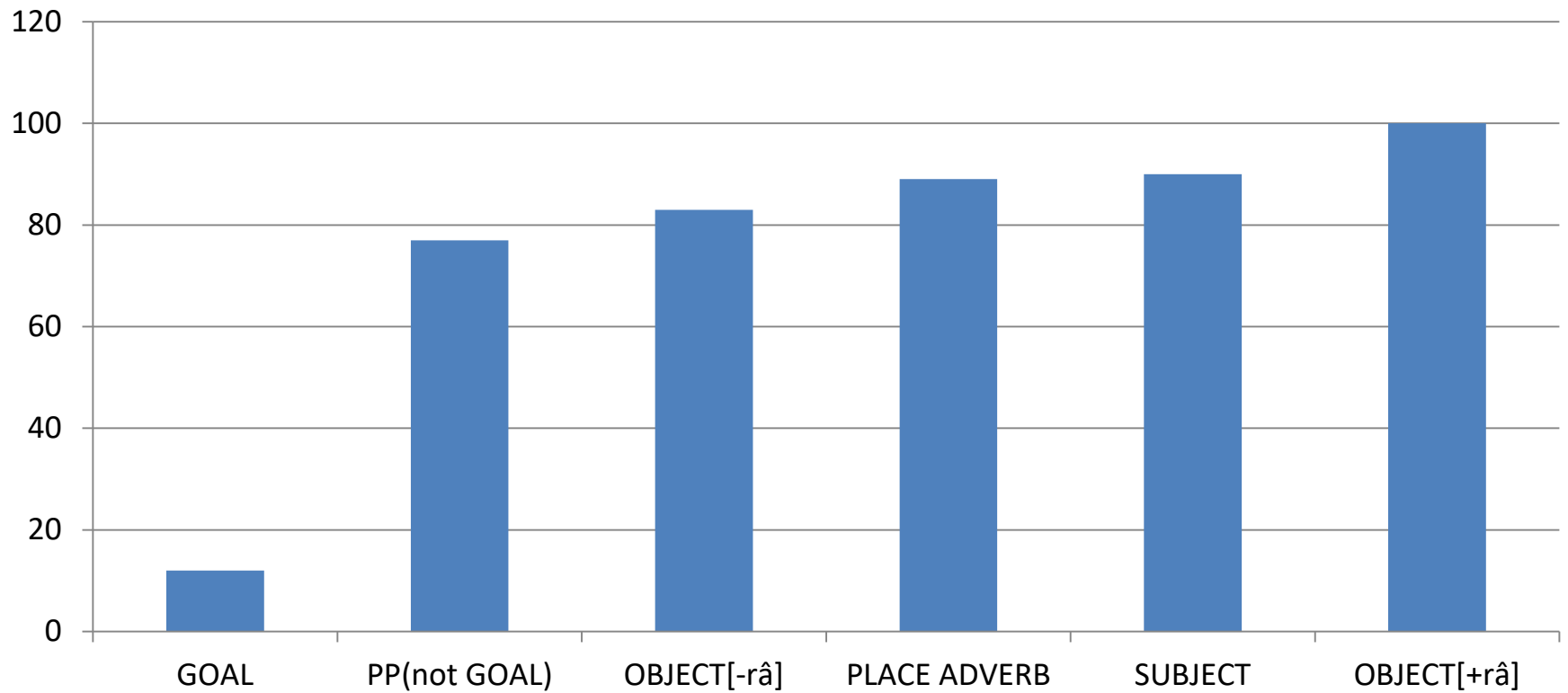
Frommer (1981) on Persian:

- 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
- Wrt 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

- Arnold, J. E., Wasow, T., Losongco, A. & Ginstrom, R. (2000). Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language* 76, 28-55.
- Bakaev, ĉ. Ch. (1965): *Jazyk azerbajdžanskich kurdov*. Moskva: Akademija Nauk.
- Borjjan, Habib & Maryam Borjjan (2008) The last Galesh herdsman: Ethno-linguistic materials from South Caspian rainforests. *Iranian Studies* 41(3): 365-402.
- Bresnan, Joan and Marilyn Ford (2010). Predicting syntax: processing dative constructions in American and Australian varieties of English. *Language* 86. 168–213.
- Bulut, Christiane. In preparation. *The Turks of Iran. Language and Culture in the Turkic Region of Bayâdestân*. Vol. 1: Texts. Wiesbaden: Harrassowitz (based on Habilitation thesis, Nov. 2006)
- Dryer, Matthew. 2013a. Order of Object and Verb. In: Dryer, Matthew S. & Haspelmath, Martin (eds.) *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://wals.info/chapter/83>, Accessed on 2019-11-11.)
- Dryer, Matthew. 2013b. Order of Adposition and Noun Phrase. In: Dryer, Matthew S. & Haspelmath, Martin (eds.) *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://wals.info/chapter/85>, Accessed on 2019-11-11.)
- Dryer, M. S. & Gensler, O. D. (2011). Order of object, oblique, and verb. In Dryer, M. S. & Haspelmath, M. (eds.), *The World Atlas of Language Structures Online*. Munich: Max Planck Digital Library, Chapter 84. Available online at: <http://wals.info/chapter/84>. (Accessed on 17/09/2011).
- Faghiri, P. & Samvelian, P. (2015). How much is determined by the syntax? An empirical approach to the position of the direct object in Persian. In Dabir-Moghaddam, M. (ed.), *Proceedings of the 9th Iranian Conference on Linguistics*. Tehran: AllameTabataba'i University Press. 1419-1435.
- Faghiri, P., Samvelian, P. & Hemforth, B. (2014). Accessibility and word order: The case of ditransitive constructions in Persian. In Müller, S. (ed.), *Proceedings of the 21st International Conference on Head-Driven Phrase Structure Grammar*. Stanford, CA: CSLI Publications. 217-237.
- Frommer, P. (1981). *Post-verbal phenomena in colloquial Persian syntax*. PhD thesis, University of Southern California.

## References (cont.)

- Haig, G. (2014). Verb-Goal(VG) word order in Kurdish and Neo-Aramaic: Typological and areal considerations. In Khan, G. & Napiorkowska, L. (eds.), *Neo-Aramaic and its linguistic context*. New York: Gorgias Press. 407-425.
- Haig, G. (2017a). Post-predicate arguments in Iranian languages. Keynote lecture, *Seventh International Conference on Iranian Languages (ICIL7)*, Lomonosov Moscow State University, 28-30th August 2017.
- Haig, G. (2017b). Western Asia: East Anatolia as a transition zone. In Hickey, R. (ed.), *The Cambridge Handbook of Areal Linguistics*. Cambridge: Cambridge University Press. 396-423.
- Haig, G. & Khan, G. (2018). Introduction. In Haig, G. & Khan, G. (eds.), *The languages and linguistics of Western Asia: An areal perspective*. Berlin, New York: Mouton de Gruyter. 1-29.
- Haig, Geoffrey and Mohammad Rasekhmahand (2019). Post-predicate elements in Iranian and neighbouring languages: Inheritance, contact, and information structure. Position paper, available at: [https://www.uni-bamberg.de/fileadmin/aspra/05\\_Events/2019\\_Post-predicate\\_elements\\_in\\_Iranian/2019\\_post-predicate\\_iranian\\_position\\_paper\\_haig\\_rasekhmahand2019.pdf](https://www.uni-bamberg.de/fileadmin/aspra/05_Events/2019_Post-predicate_elements_in_Iranian/2019_post-predicate_iranian_position_paper_haig_rasekhmahand2019.pdf)
- Haig, G. & Stefan Schnell. 2019. (eds.), *Multi-CAST: Multilingual corpus of annotated spoken texts*. (Online resource available at: [multicast.aspra.uni-bamberg.de/](http://multicast.aspra.uni-bamberg.de/)) (accessed 15.11.2019)
- Haig, G., Schnell, S. & Wegener, C. (2011). Comparing corpora from endangered language projects: Explorations in typology based on original texts. In Haig, G., Nau, N., Schnell, S. & Wegener, C. (eds.), *Documenting endangered languages: Achievements and perspectives*. Berlin, New York: Mouton de Gruyter. 55-86.
- Haig, G., Nils Schiborr, Stefan Schnell and Maria Vollmer. (in prep.) Introducing new referents: A corpus-based cross-linguistic perspective.
- Haig, Geoffrey (to appear). Post-predicate constituents in Kurdish. In: Matras, Yaron & Ergin Öpengin & Geoffrey Haig (eds.), *Structural and Typological Variation in the Dialects of Kurdish*. London: Palgrave MacMillan.
- Hawkins, J. A. (2008) An asymmetry between VO and OV languages: The ordering of obliques. In Corbett, G. & Noonan, M. (eds.), *Case and grammatical relations: Essays in honor of Bernard Comrie*. Amsterdam: Benjamins. 167-190.
- Hawkins, J. A. (2014). *Cross-linguistic variation and efficiency*. Oxford: Oxford University Press.

- Lazard, Gilbert. ([1957]2006). *Grammaire du persan contemporain*. (Nouvelle édition, avec la collaboration de Yann Richard, Rokhsareh Hechmati et Pollet Samvelian). Tehran: Institut Français de recherche en Iran. [orig. 1957].
- Levshina, Natalia (2019) Token-based typology and word order entropy: A study based on Universal Dependencies. *Linguistic Typology*, Volume 23 (1): 533–572.
- Nikitina, T. (2011). Categorical reanalysis and the origin of S-O-V-X order in Mande. *Journal of African Languages and Linguistics* 32, 251–273.
- Rasekh-Mahand, M., Alizadeh-Sahraei, M. & Izadifar, R. (2016). A corpus-based analysis of extraposition of relative clauses in Persian. *Ampersand* 3, 21-31.
- Reetz, S. (2015). *A quantitative corpus-based analysis of post-predicate goals in spoken Oghuz Turkic varieties*. MA-Thesis, Universität Bamberg.
- Schnell, S., Schiborr, N. & Haig, G. (under review). Efficiency in discourse processing: Does morphosyntax adapt to accommodate new referents?
- Stilo, D. L. (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, 10-12th May 2018..
- Wasow, T. (2002). *Postverbal behavior*. Stanford, CA: CSLI Publications.
- Wälchli, Bernhard. (2009). Data reduction typology and the bimodal distribution bias. *Linguistic Typology*, Volume 13 (1): 77–94.
- Yamashita, H. & Chang, F. (2001). ‘Long before short’ preference in the production of a head-final language. *Cognition* 81, B45–B55.
- Yamashita, H. & Chang, F. (2006). Sentence production in Japanese. In Nakayama, M., Mazuka, R. & Shirai, Y. (eds.), *Handbook of East Asian Psycholinguistics, vol. 2*. Cambridge: Cambridge University Press. 291–297.