Post-predicate elements in the Western Asian Transition Zone

A corpus-based approach to areal typology

Edited by

Geoffrey Haig

Mohammad Rasekh-Mahand

Donald Stilo

Laurentia Schreiber

Nils N. Schiborr

language science press

Contact and Multilingualism

Editors: Isabelle Léglise (CNRS SeDyL), Stefano Manfredi (CNRS SeDyL)

In this series:

1. Lucas, Christopher & Stefano Manfredi (eds.). Arabic and contact-induced change.

2. Pinto, Jorge & Nélia Alexandre (eds.). Multilingualism and third language acquisition: Learning and teaching trends.

3. Hakimov, Nikolay. Explaining Russian-German code-mixing: A usage-based approach.

4. Sánchez Moreano, Santiago & Élodie Blestel. Prácticas lingüísticas heterogéneas: Nuevas perspectivas para el estudio del español en contacto con lenguas amerindias.

5. Migge, Bettina & Shelome Gooden. Social and structural aspects of language contact and change.

6. Coler, Matt & Andrew Nevins (eds.). Contemporary research in minoritized and diaspora languages of Europe.

7. Pfadenhauer, Katrin, Sofia Rüdiger & Valentina Serreli (eds.). Global and local perspectives on language contact.

8. Haig, Geoffrey, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.). Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology.

Post-predicate elements in the Western Asian Transition Zone

A corpus-based approach to areal typology

Edited by

Geoffrey Haig

Mohammad Rasekh-Mahand

Donald Stilo

Laurentia Schreiber

Nils N. Schiborr

Freie Universität Berlin

Contents

Contents

Short biography of the editors

Geoffrey Haig is professor of linguistics at the University of Bamberg. He specializes in language typology and areal linguistics, with a focus on quantitative methodologies drawing on multi-lingual spoken-language corpora.

Mohammad Rasekh-Mahand is professor of linguistics at Bu-Ali Sina University, Hamedan, Iran. His research centres on the syntax of Persian and other new Iranian languages. He has published on issues connected with word order, clitics, definiteness, transitivity, aspect and evidentiality.

Donald L. Stilo has a B.S. in Languages from Georgetown University, and an M.A. and Ph.D. from the University of Michigan in Linguistics with a concentration in Iranian languages. He taught Persian in various universities in the United States, was a scientist at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, for ten years, and in 2005/2006 was a fellow at the Swedish Collegium for Advanced Studies in Uppsala. He has been retired since 2012 and splits his time between the United States and the Netherlands.

Laurentia Schreiber is a researcher at the Department of General Linguistics at the University of Bamberg. She focuses on language contact and multilingualism in minority languages and obtained her PhD recently with a grammar of endangered Romeyka in Turkey.

Nils Schiborr is a postdoctoral research associate at the Department of General Linguistics at the University of Bamberg. He has worked on discourse processing, referential choice, and the interface of discourse and grammar, chiefly from a corpus-based typological perspective.

Acknowledgments

Part I

Theoretical and methodological issues

Chapter 1

Post-predicate elements in the Western Asian Transition Zone: Data, theory, and methods

Geoffrey Haig[a], Mohammad Rasekh-Mahand[b], Donald Stilo[c], Laurentia Schreiber[a] & Nils Schiborr[a]

[a]University of Bamberg [b]Bu-Ali Sina University Hamedan [c]Max Planck Institute, Leipzig (retired)

This chapter spells out the conceptual and methodological foundations for the volume, summarizes previous research, illustrates the methodology and analysis, and presents the results of two case studies. We provide evidence in support of a semantically finer-grained approach to word order that distinguishes between various non-subject constituents, and illustrate how this can be leveraged to detect areal effects in syntax. We implement this approach on a sample of language corpora from 35 languages, including Turkic, Iranian, Semitic, Hellenic, Kartvelian, and Armenian from what we term the Western Asian Transition Zone (WATZ). In a first case study, we demonstrate the existence of a robust Goals Last effect across the entire database and formulate a revised hierarchy for postverbal placement. Our approach identifies the specific properties of spatial goals that distinguish them from metaphorically related roles such as recipient, addressee, and benefactive, which previous studies had conflated. In a second case study, we investigate weight effects on post-verbal placement, concluding that overall, the impact of weight is minimal, a finding reflected in several chapters of the volume. The final section summarizes the contributions to the volume, and the Appendices provide raw data summaries across the entire WOWA data set, and information on sources.

1 Theoretical preliminaries

1.1 General background

This volume represents the collaborative outcome of a team of researchers, all of whom contributed expertise and data on languages of what we loosely refer to as the Western Asian Transition Zone (WATZ, cf. Section 1.2). The companion enterprise to this volume is a portfolio of online accessible, multiply-reusable digital resources, comprising of two data-sets: WOWA (*Word Order in Western Asia*), a multi-lingual corpus containing approximately 40 data sets from a sample of languages across WATZ (Haig et al. 2022), and HamBam (*The Hamedan-Bamberg Corpus of Contemporary spoken Persian*, Haig & Rasekh-Mahand 2022). HamBam is a richly annotated corpus of a single language, colloquial spoken Persian, based on the Multi-CAST architecture (Haig 2015a, Schnell et al. 2023).[1] It is designed for finer-grained investigations of word order, prosody, and register in a single language (Persian), while WOWA is a multi-lingual database designed to investigate the transition-zone phenomena outlined in the following paragraphs. Most of the research reported here is based on WOWA.

Our research is intended to satisfy the requirements of "reproducible research," in the spirit of Berez-Kroeker et al. (2018), and the emphasis on accessibility and accountability of primary data, and maximal transparency of analysis procedures have been guiding principles throughout: research should be conducted in a manner "which allows readers to confirm claims about language structure through direct access to the original observational data" (Berez-Kroeker et al. 2018: 6). In this overview chapter, we introduce and exemplify the data sources, theoretical concepts and research questions, and illustrate the main findings with two case-studies. Figure 1 shows the location of the doculects in WOWA at the time of writing; an overview of all data sources is available in the Appendix to this chapter.

1.2 The Western Asian Transition Zone (WATZ)

The concept of "transition zone" has been discussed in various guises (e.g. "buffer zone" Stilo 2005, "intersection zone" Stilo 2009, or "typological sandwich" Szeto & Yurayong 2021, see Haig et al. In press). Here, we continue the terminology introduced in Haig & Khan (2019); we define a transition zone as a geographic area lying at the intersection of two contiguous regions characterized by diametrically opposing values for some linguistic feature. The choice of feature is essentially

---

[1]For WOWA see https://multicast.aspra.uni-bamberg.de/resources/wowa/; for HamBam see https://multicast.aspra.uni-bamberg.de/resources/hambam/.
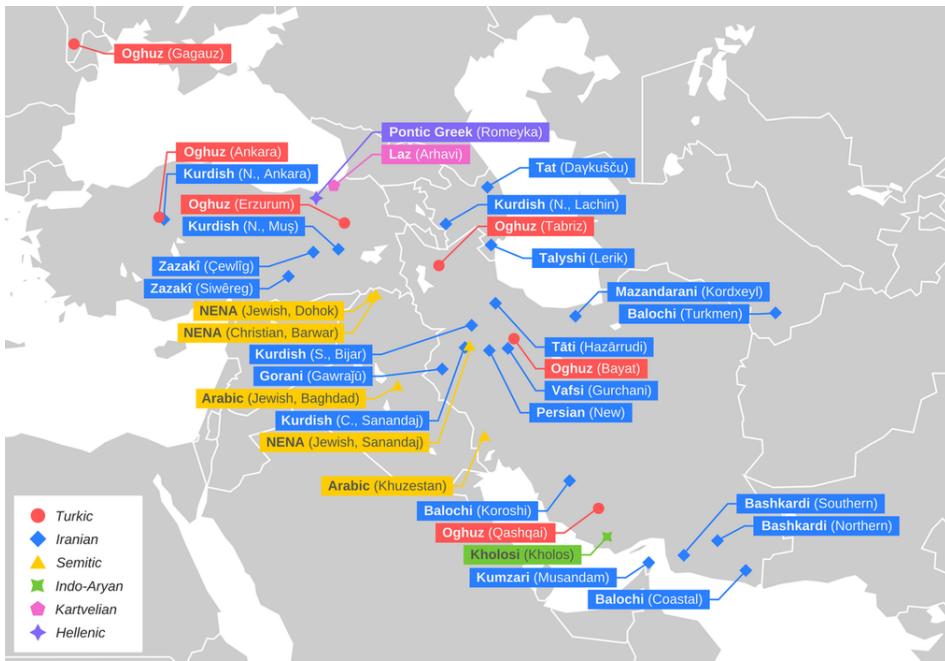
Figure 1: Locations of doculects in WOWA (November 2023)

unrestrained, and determined primarily by the research questions of the investigators, though considerations of general theoretical pertinence, availability of appropriate data, and operationalizability of the feature values concerned, will also play a role. It is clear that a transition zone, as just defined, can be identified at very different degrees of granularity: even a dialect isogloss, separating related dialects characterized by, for example, the presence versus absence of nasalized vowels, could be considered a transition zone, albeit at the micro-level of structural differentiation.

The Western Asian Transition Zone is at the other extreme of granularity. It is defined by the overlap of two areas of continental scale, which differ with regard to the following features (and a number of ancillary features, which are taken up at various points below): OV versus VO word order, and adpositional type (prepositions vs. postpositions, with some additional complications). An approximation of the global distribution of these two features can be found by considering the two maps in WALS (Dryer 2013b,c), Feature 83A (OV/VO) and Feature 85A (adposition order). On both maps, two adjacent macro-areas can be identified within Eurasia and the Indian Sub-Continent, each dominated by languages with opposing feature values: The first is the central and eastern Asian land block, dom-

inated by Robbeets' (2017) "Transeurasian languages" (Turkic, Mongolic, Tungusic, Japonic, Koreanic), which are uniformly characterized by OV and postpositions. These features flow seamlessly into the Indian Sub-Continent, dominated by Eastern Iranian, Dravidian, and Indo-Aryan. The second macro-area is to the southwest, where we find the VO and prepositional languages of North Africa, Western Europe and the Circum-Mediterranean region (Afro-Asiatic, Romance, Hellenic). The area of overlap between these two macro-areas is what we refer to as the Western Asian Transition Zone (WATZ). The core of WATZ is the lower catchment areas of the Euphrates and Tigris valleys, springing from the elevated plateau of today's Eastern Turkey and descending to the alluvial plains of northern Iraq and Syria. Today, this region is divided between four nation states, Turkey, Syria, Iraq and Iran.

Like most of the units that have been suggested in areal linguistics,[2] the geographic boundaries of WATZ cannot be precisely defined. Partly, this is due to the nature of a transition zone, which often implies fade-out phenomena (Stilo 2012), rather than abrupt transitions from one feature value to another. An example of such a fade-out phenomenon is the feature "order of spatial Goals relative to their governing predicate," as in a clause such as *the girl went to the market.* We can distinguish two options: the Goal precedes the verb (GV), or it follows the verb (VG). Mapping the frequencies of these two orders across the languages of WATZ reveals an increase in the frequency of Verb-Goal (VG) ordering as one progresses westwards. For example, Iranian languages with long-standing Semitic influence, from the southern and westernmost peripheries of WATZ, show almost 100% post-verbal Goals (in the sense of "Goals of movement"; see below on terminology), matching the figure that characterizes most of Semitic, and of the westernmost branches of Indo-European. Similarly, Turkic languages such as Qashqai, which are under heavy influence from western Iranian languages, also show high rates (>60%) of post-verbal Goals (Haig et al. In press). Note that both Iranian and Turkic are generally considered to be OV (sometimes erroneously equated with "verb final"), so the high frequency of post-verbal Goals is not expected for these languages. For OV languages situated further northward and eastward, the figures drop, for example in the Balochi variety of Turkmenistan (Nourzaei & Haig 2024 [this volume]), or Mazandarani of the Caspian region (Stilo & Haig 2022). Furthermore, initial counts from related OV languages further eastward beyond WATZ suggest they drop still further; in Indo-Iranian (Dardic) Kalasha

---

[2]Friedman & Joseph (2017: 75) discuss the issue of defining the boundaries of Sprachbünde, noting "the boundaries are as elastic as the micro-zones of convergence that add up to the larger convergence area." The boundaries, and even the set of languages and varieties involved, remain disputed even for intensely-researched linguistic areas such as the Balkan Sprachbund.

(Northern Pakistan, bordering Afghanistan), provisional counts of the texts in Petersen (2015) suggest rates of post-verbal Goals below 30%, while in the texts from Dukhan, a Turkic variety from Northern Mongolia, the figure approaches zero. If the basic hypothesis behind WATZ is correct, then we might predict similar low values of VG for unrelated OV languages toward the eastern fringe of Asia, such as spoken Japanese and Korean, but this remains to be tested.

Although our current data coverage is low-density and uneven, we tentatively hypothesize that in the OV languages of Asia, increasing rates of post-verbal Goals roughly correlate with increasing proximity to the Mesopotamian core of WATZ.[3] But we can identify no precise geographical isogloss that constitutes a categorical border separating VG from GV. Rather, we are dealing with a continuum of values, which we assume extends beyond the region defined by the sample in Figure 1.

Returning to the broader theoretical interest of transition zones, it has been suggested that regions of intense contact (a hallmark of transition zones) are havens for typologically rare constructions. Harris & Campbell (1995: 137) note that certain word-order constellations only emerge in contact situations, and our data lend some credence to this view. Furthermore, transition zones are overall smaller than the macro-areas that engender them, and are therefore likely to contain fewer languages. The probability that any random language sample includes languages from transition zones is thus lower than the probability of selecting languages from within established linguistic areas. To this extent, the contributions to this volume are thus intended to counterbalance an existing bias in language sampling. To conclude, transition zones are not definable in terms of a precisely circumscribed geographic region. Rather, they should be seen as hypotheses which demarcate a potentially fruitful set of languages located at a region of conflicting feature values, which can serve as an experimental setting for investigating the broader question: what happens when languages with opposing feature values collide?

1.3 Word order

The term "word order" is often taken as synonymous for the traditional Greenberg'ian six-way typology (S/V/O). In our work, however, we follow Dryer (1997,

---

[3]It is worth noting that areality alone does not fully account for the findings; the phylogenies of the languages concerned are also relevant, with the Iranian languages apparently the most prone to areally-induced word-order variation; see Haig et al. (In press), Bickel (2017: 42) on the interplay of areality, inheritance, universal principles in co-determining language structure, and Haig & Schiborr (In review), for arguments in favour of a universal Goals Last principle.

2013a), who proposes decomposing the traditional six-way typology into two binary sub-features, S/V and V/O. Here we focus almost exclusively on the relative ordering of direct object and verb (V/O), while setting aside the position of subjects (S); see below for empirical justification, and Dryer (1997, 2013a) for the arguments against the six-way typology. In line with recent work in corpus-based typology (see Section 1.4), we extend the typology to include the position of other verbal arguments, such as Recipients, Locatives, copula complements, Goals, Addressees, relative to the verb. Our motivation for this is entirely data-driven: a large body of research (see Section 2) on the languages of Western Asia suggests that beyond direct objects, other less prominent and often overlooked constituent types provide sensitive indicators for contact influence (Haig et al. In press). Furthermore, the default assumption that the position of direct objects (i.e. OV vs. VO) can be generalized across other verbal arguments, is falsified in many languages of the region, which exhibit consistent OV order, but simultaneously have post-verbal placement of certain non-direct objects; see Section 4 below for an overview of the relevant findings from WOWA. Word order typology has tended to either ignore non-direct objects, or to subsume them under umbrella terms (e.g. 'obliques,' Hawkins 2008, Levshina 2019, Jing et al. 2021; or 'PP' in Frommer 1981). Our research indicates that a finer-grained semantic approach to non-direct-objects is more appropriate, which we spell out in Section 4 below.

A perennial debate in word order typology concerns how, or indeed whether, one can identify some kind of "basic word order" for a language. There are a number of issues at stake, which we will address here. First, although scholars such as Mithun (1992) have argued against the universality of basic word order, it is important to note that most researchers engaged in word order typology (see e.g. Dryer 2007 and Song 2018 for summaries) have never claimed that every language has a "basic word order." Thus, the global overview of the VO vs. OV word-order parameter in (Dryer 2013c) assigns all languages in the sample to one of three types: OV, VO, and "no dominant order," with the latter comprising some 7% of the 1,518 languages in Dryer's (2013c) sample. It has always been acknowledged that it is not possible to identify a single order as "basic" for every single language. But that does not invalidate the enterprise of word order typology as a whole, any more than the fact that some languages do not have lexical tone invalidates a cross-linguistic approach to tone systems.

At least three different approaches to basic word order can be identified. First, the frequency of usage. Contrary to what is often claimed, a frequency approach does not necessarily imply a simple majority decision. Dryer (2007: 11) proposes that the basic (or dominant) order is that variant which is at least twice as frequent as the next most frequent order. In the case of a binary feature such as OV

vs. VO, that would mean one order would account for at least 66% of the relevant cases in order to count as the basic order. The frequency approach raises a host of methodological issues related to corpus size and representativity and ignores many finer nuances, some of which we take up below.

A second approach seeks to define a particular constructional sub-type which is taken as prototypical for the construction under consideration. For example, Siewierska (1988) provides a widely-cited rule-of-thumb for identifying the basic S/V/O order cross-linguistically. According to her, the basic order is that which:

> […] occurs in stylistically neutral, independent, indicative clauses with full noun phrase (NP) participants, where the subject is definite, agentive and human, the object is a definite semantic patient, and the verb represents an action, not a state or an event (Siewierska 1988: 8).

This would rule out, for example, interrogatives, subordinate clauses, or clauses in which either S or O is pronominal, and it would also exclude transitive clauses with verbs like 'see' or 'know,' which express states rather than actions and do not involve a 'semantic patient.' In principle, this is a reasonable approach and has gained some currency. However, we see no obvious justification for ruling out transitive clauses with indefinite direct objects, particularly as overall, direct objects are among the most likely argument types to host indefinite referents (Haig et al. 2021: 164, Schnell et al. 2023).

Another variant of the prototype approach is implemented by Bağrıaçık (2018) and Neocleous (2020), both investigating varieties of Asia Minor Greek.[4] These authors focus on what is variously termed "pragmatically neutral" or "unmarked" transitive clauses, in which neither argument "is associated with either a topic, or a focus reading" (Bağrıaçık 2018: 150). Such clauses only occur under very specific discourse conditions, three of which are identified by Bağrıaçık as follows: (i) answer to an all focus question (*What happened?*); (ii) introductory clauses of narratives, where both subject and object are newly introduced into the world of discourse; (iii) generic statements (*the Earth orbits around the Sun, Sam knows Tibetan*), (Bağrıaçık 2018: 151–154). Neocleous (2020) adopts a similar approach, though with a different formalization of the concept of "pragmatically unmarked." Note that Bağrıaçık (2018) concept of pragmatically neutral transitive clause is different from Siewierska's (1988). Siewierska stipulates that in a

---

[4]Although both authors work within a Minimalist framework, thus approach "word order" from a rather different perspective to the typologically-oriented approach of Siewierska (1988), they nevertheless attempt to define a "basic" clause type on the basis of certain pragmatic and surface morphosyntactic properties, from which other orders are derived.

basic transitive clause, both subject and object are required to be definite, while Bağrıaçık's "introductory clauses," for example, imply that neither would be definite.

Along with prototype, and frequency-based approaches, there are two other connotations of "basic word order." The first is the concept of inherited, or historical word order. By this, we mean the word order that is reconstructible for the assumed proto-language of the languages under investigation. For example, there is little reason to doubt that both Turkic and Iranian languages had OV word order at the oldest period of attestation, or that proto-Semitic was VO, so these languages can be reasonably classified as historically OV and VO respectively. However, claiming that Semitic is historically VO does not equate to a claim that this is the "basic order" for all modern Semitic languages; basic order can change.[5]

Finally, a basic word order may be motivated by theory-internal considerations. This is the case for German, for which the basic word order is often claimed to be OV, with VO considered as secondarily derived via (some version of) verb movement. If we strictly applied the criteria of Bağrıaçık (2018), which relies on the concept of pragmatically neutral clause, or Neocleous (2020: 116), which invokes (among other things) word order in main declarative clauses, we would obtain a different result for German, because OV order in German is actually found in subordinate clauses. The fact that different criteria yield different results is reflected in the classification of German as "no dominant order" in Dryer (2013c).

Having briefly considered various interpretations of "basic word order," we turn to a methodological issue in connection with word order and small corpora. Both Siewierska (1988) and Bağrıaçık (2018) require a transitive clause to have two nominal (as opposed to pronominal) arguments. However, for research based on small corpora of spoken language, often without recourse to native speakers' judgements, this approach runs into an immediate problem. Cross-linguistically, in natural discourse, very few transitive clauses contain two overt lexical (as opposed to pronominal) arguments. Du Bois (2003: 62–63) provides data from five spoken language corpora indicating that the overall frequency of clauses with two lexical core arguments lies between two and seven percent, and similar findings are reported in the literature (see contributions in Du Bois 2003). Additional restrictions, such as requiring both S and O to be "pragmatically neutral"

---

[5]It may seem superfluous to labour this point, but it is nevertheless misunderstood in Asadpour (2022a: 42), who interprets references in the literature to "historical" and "inherited" word order in Neo-Aramaic as claims regarding "basic word order" in contemporary Neo-Aramaic. These are clearly separate claims.

(Bağrıaçık 2018), or both to be "definite" (Siewierska 1988), would further reduce the pool of valid tokens. It is no surprise that Bağrıaçık (2018) investigation of word order in Pharasiot Greek is largely informed by the elicitation of grammaticality judgements, rather than a quantitative analysis of naturalistic data.

Cross-linguistic research of spontaneous spoken discourse demonstrates that transitive subjects overwhelmingly express given information (>90%, Haig et al. 2021: 165), and are consequently predominantly either zero, or pronominal in form, rather than lexical NPs (cf. Du Bois' 1987 'Avoid Lexical A'-constraint; see discussion in Haig & Schnell 2016). If we were obliged to exclude all clauses with pronominal or zero subjects, we would vastly decrease the number of potentially analysable tokens in the sample. For objects, however, the likelihood of lexical NP expression is very much higher, and the population of analysable tokens correspondingly larger. This is a further motivation to eschew the six-way SVO-typology, and to focus on the relative ordering of verb and direct object only. In the future, we may expand our investigation to include the position of subjects in the WOWA data, which are fully accessible and amenable for additional coding of subjects (see also Molin 2022, Rasekh-Mahand et al. 2024 [this volume], Forker 2024a [this volume],b [this volume]). Currently, however, we continue to work with the binary feature OV/VO.

1.4 Corpus-based approaches to word order

Methodologically, we apply a corpus-based typological approach (Wälchli 2009, Levshina 2019, Futrell et al. 2020, Gerdes et al. 2021, Haig et al. 2021, Schnell & Schiborr 2022, among many others). Within such an approach, the emphasis shifts away from assigning a 'basic word order' to a particular "language." Rather, in corpus-based approaches, statements on word order refer to frequency distributions derived from actual corpora, and are probabilistic in nature. Strictly speaking, corpus-based approaches to word order yield a characterization of a specific corpus (a "doculect"), rather than "a language," though we continue to use the over-simplified terminology here.

Corpus-based typological approaches to word-order typology are dominated by research on large, written corpora of languages with a pre-existing orthographic norm, for which copious quantities of pre-digitalized text are available (see in particular Universal Dependencies (UD)-consortium, Nivre et al. 2020).[6] Consequently, there is a bias towards standardized (and mostly Eurasian) languages, and more importantly, towards written language. Cross-linguistic re-

---

[6]https://universaldependencies.org/

search based on spoken language corpora, on the other hand, is still in its infancy (see among others Schnell et al. 2021, Mettouchi & Vanhove 2021, Schnell & Schiborr 2022, Seifart et al. 2022, Levshina et al. 2023). Frommer's (1981) research had already demonstrated significant differences between formal written and spoken Persian, indicating that written language data cannot be assumed to reliably reflect the structures of spoken language. Any research agenda that purports to investigate the impacts of processing and production constraints on language structure would be well advised to focus on the mode of language production where these constraints are operative in real time — and that would not be written language (see Schnell et al. 2021 for summary arguments). In Section 5, we present an investigation into the role of weight as a predictor for word order, based on our spoken language corpora, which illustrates the importance of controlling for modality.

As mentioned above, we focus on various non-subject constituents, and their position relative to the governing predicate, for example direct object and verb, or Addressee and verb, and so on. Our database is thus designed to answer the following questions:

(1)   a.   What is the probability that non-subject argument A, in doculect X, belonging to language family Y, spoken at location Z, occurs after its governing predicate?
      b.   Which variables influence this probability?

From the answers to 1a we could infer word-order "types," by setting some predefined quantitative boundaries. For example, Levshina (2019: 559) provisionally classifies a language (=corpus) with greater than 80% VO as "VO," with less than 20% VO as "OV," and 20–80% as "mixed." However, this is a matter of heuristic interpretation of the raw data, rather than principled classification of "Type." It should be obvious that two doculects with values of 79% and 81% VO respectively are not necessarily exemplars of fundamentally distinct types (see Wälchli 2009 on "data reduction typology"). The variables that were tested for question 1b are presented in Section 3.2. Questions 1a-1b; they can be investigated both at the level of individual doculects, or by applying appropriate statistical methods to the entire sample or some sub-set thereof. In Sections 4 and 5 below we present case studies for the impact of semantic role (Section 4), and of weight (Section 5). Having outlined the theoretical background and the research questions, we illustrate the structure of the data sets and the methodology in the following sections.

2 Previous research on Western Asia, and terminological issues

The assumptions and aims of the WOWA project were inspired by insights gained over many years of previous research, and it is appropriate to briefly outline the main currents of that research. Typologists have long been aware of the "mixed typology" of Iranian languages, e.g. Comrie (1989: 19) on Persian, a language with OV in the VP, but head-initial NPs, prepositions, and clause-initial complementizers (see Dabir Moghaddam 2018 for a recent summary). Don Stilo developed the idea that the mixed typology of Persian was shared to differing degrees by other West Iranian languages, and that the degree and nature of West Iranian mixed typologies followed an approximate areal distribution. Stilo's claim was that West Iranian was sandwiched between the opposing typologies of Semitic (consistently head-initial) and Turkic (head-final), with different West Iranian languages synchronizing with the profile of their respective geographic neighbours. These ideas were fleshed out with a survey of adpositional types in Stilo (2005, 2006, 2009), and developed in a number of other publications (2012, 2018b, 2018a).

Frommer (1981) noted a further non-harmonic aspect of West Iranian syntax: the post-verbal positioning of certain kinds of non-direct-object arguments. Frommer (1981) focussed on the syntax of "informal Persian" (IP), including both spoken and written samples from different registers. This was, in fact, the crucial breakthrough: formal written Persian, the more usual object of study, is rather consistently "verb final," hence post-verbal elements are a fringe phenomenon that had not been systematically investigated.[7] Frommer (1981) was the first systematic analysis of post-predicate elements in different registers of informal Persian;[8] his findings can be summed up as follows: (i) across the different registers of informal Persian, there is a cline of formality such that a lower degree of formality correlates with an increase in post-predicate elements; in-group domestic conversational Persian exhibited the highest levels; (ii) semantic role is crucial,

---

[7]Lazard (1957) had already noted the predominance of post-verbal Goals in informal spoken Persian, but did not systematically investigate the topic.

[8]Frommer did not actually investigate formal written Persian, and we still lack a systematic study. Parizadeh's (Parizadeh & Rasekh-Mahand 2024 [this volume]) study of Early Classical New Persian (11–14th Century CE) demonstrates near 100% verb finality in these written texts, which matches the native speaker intuition of one of our authors regarding contemporary formal written (e.g. academic prose) Persian. More recent corpus-based approaches to written Persian (e.g. Faghiri et al. 2018) investigate the relative ordering of pre-verbal constituents, while post-verbal constituents lie outside the purview of this research. Formal written Persian is thus essentially considered to be a "verb final" language.

with goals of motion ("destinations" in Frommer's terminology) as the leaders in post-verbal placement, across all registers; (iii) information status is relevant for post-verbal placement (focal versus non-focal) of direct objects, but appears to be irrelevant for goals of motion; (iv) there is a stronger tendency for post-verbal constituents to lack overt flagging. By and large, these findings have been confirmed on more recent corpora of spoken Persian (Rasekh-Mahand et al. 2024 [this volume]).

From a comparative Iranian perspective, the remarkable aspect of Frommer's findings is that they closely align with findings from the lesser-researched and generally non-standardized West Iranian languages documented in this volume and elsewhere. What this suggests is that the phenomena which Frommer identified were not merely irregularities specific to informal spoken Persian, but in fact reflected word order traits of considerable antiquity, which characterize most (perhaps all) West Iranian languages (see Nourzaei & Haig 2024 [this volume], Korn 2024 [this volume], Nourzaei 2024 [this volume], Rasekh-Mahand et al. 2024 [this volume], Parizadeh & Rasekh-Mahand 2024 [this volume], Mohammadirad 2024 [this volume]). From this perspective, it is the strictly verb-final, formal written Persian that is the exception when it comes to West Iranian word order. This has considerable implications for the diachronic study of word order, which is largely reliant on written language sources.

Research on post-predicate elements in other Iranian languages began with Kurdish (Haig 2015b, Haig & Thiele 2014, Haig 2022d), and has since expanded to neighbouring languages (Haig 2015a, 2017, Stilo 2018a, Jahani 2018, Asadpour 2022a, and contributions to this volume). Most of this research is based on corpora of spoken narrative texts (see below), though increasingly enhanced with experimental data (see Skopeteas 2024 [this volume]). It has emerged that while all West Iranian languages investigated to date (with the exception of Kumzari, Anonby 2015) are consistently OV (see below), like spoken Persian, they are not "verb final" because a significant number of non-direct object arguments regularly follow the verb. A similar pattern can also be observed in Turkic languages in contact with Iranian (Schreiber et al. 2021, Stilo 2021a). A second point that quickly emerged from the earlier studies is that the nature, and systematicity, of post-verbal arguments follows an approximate areal distribution, along the lines of Stilo's (2009) suggestions. Among Iranian OV languages, the highest frequencies, and greatest variety of post-verbal argument types, are attested among varieties of northern Kurdish spoken in Iraqi Kurdistan and adjacent regions of southeastern Turkey, Syria and Iran, a region we provisionally refer to as Mesopotamia. Mesopotamia is of course also home to a number of historically VO Semitic languages, (Neo-)Aramaic and Arabic, which have co-existed with

Kurdish and other OV languages for centuries, and indeed for millennia in the case of Aramaic. In these VO languages, it is universally the case that other non-direct object arguments also follow the verb, and it seems plausible to assume that the syntax of these languages had some impact on the Iranian languages with which they shared territory for at least 2000 years, and ultimately also on Turkic (perhaps via Iranian in many cases).

In a pilot study, Haig & Thiele (2014) compared word order in naturalistic texts from a sample of languages mostly from Mesopotamia.[9] The authors identified four types of arguments that are predominantly post-verbal in these languages, cited in the original formulation as follows:

- Recipients of verbs of transferred possession (e.g. GIVE)

- Destination or direction of verbs of movement (e.g. GO, RUN, FALL)

- Destination or direction of verbs of caused motion
  (e.g. PUT, PLACE, TAKE)

- Addressees of verbs of speech (e.g. SAY, SPEAK, PROMISE)

Examples illustrating these four types, from Badini Kurdish (from the Gullī and Akre dialects of Iraqi Kurdistan, from Haig & Thiele 2014, citing MacKenzie 1962) are provided in (2-5):

(2) Recipient
     Northern Kurdish Akre (MacKenzie 1962)
     *min   kič-ā   xo dā   **ta***
     1SG.OBL daughter-EZ.F REFL give.PST.3SG 2SG.OBL
     'I have given my daughter **to you**.'

(3) Addressee
     Northern Kurdish Akre (MacKenzie 1962)
     *sultān-ī   got-a   **ahmad halwāčī***
     Sultan-OBL.M say.PST.3SG-DRCT Ahmad Halwachî
     'the Sultan said **to Ahmad Halwachî**.'

---

[9]The varieties were Northern Kurdish from Iraqi Kurdistan, Northern Kurdish from Midyat, Southeastern Turkey, Northern Kurdish from Muš and Erzurum; Northeastern Neo-Aramaic (Jewish) from Urmi, West Iran, and from Koy Sanjaq, Iraqi Kurdistan, and Turkish from Erzincan, Turkey. For comparison, they also included corpus data from a dominant VO language, Cypriot Greek. It was already apparent from this small data set that outside of Mesopotamia, Recipients, Addressees, and Goals of motion do not necessarily pattern alike.

(4) Goal of simple motion
     Northern Kurdish Akre (MacKenzie 1962)
     harduk rābon,   hāt-in-a   bāžar-ī
     both   get_up.PST.3PL come.PST-3PL-DRCT town-M.OBL
     'Both of them got up and came **to the town**.'

(5) Goal of caused motion
     Northern Kurdish Gullī (MacKenzie 1962)
     kir   t=sēnīk-ā   dayk-ā   xwa dā
     do.PST.3SG ADP=tray-F.EZ mother-F.EZ REFL ADP
     '(He) put (it) **on his mother's tray**.'

In the Mesopotamian languages investigated in Haig & Thiele (2014), all four semantic types exhibited broadly similar rates of post-verbal placement, which motivated the authors to define a macro-role, labeled "Goal," that would encompass all four types (and some further types such as final state of a change-of-state verb, see below). In retrospect, this terminological decision proved injudicious, for two reasons. First, it introduced ambiguity to the term "Goal," which could either be understood in the narrower sense of "Goal of verb of motion," or in the broader sense that would include Recipient, Addressee, etc. Second, it has become increasingly evident that many languages of WATZ do not lump Addressees, Recipients, and Goals of verbs of motion together (Section 4 below for data), thus casting doubt on the validity of a macro-category altogether. A broadly similar macro-category was subsequently adopted by Asadpour (2022a,b), who relabels it as "Target," and this terminology has been used in the contributions to Asadpour & Jügel (2022). While the re-labeling alleviates the ambiguity problem, it does not resolve the empirical problem that outside of some varieties of Kurdish with deep historical ties to Semitic languages, Addressees, Recipients, and spatial Goals do not pattern alike among the OV languages of WATZ, so the motivation for assuming a priori a meta-category is questionable.[10] In an effort to restore clarity, we therefore eschew the macro-category sense of "Goal" in this volume, reserving the term "Goal" strictly in the sense of "Goal or endpoint of a predicate of motion or caused motion." See end of Section 4.1 for discussion of "Recipient" vs. "Goal," and (11) for an overview of roles distinguished in WOWA.

One of the observations in the earlier literature concerned the syntax of 'final state' constituents, defined here as expressions indicating the final state of a

---

[10]There is also a lack of consensus about the nature and number of categories that are included under "Target"; some researchers include variously Benefactives, and Final States of change-of-state predicates, rendering comparison across different publications difficult.

change-of-state ('become') predicate. Haig (2017, 2022d) noted that in much of Kurdish, the final states are significantly more likely to be post-verbal than the complements of copular expressions that do not imply a change of state, even when the lexical verb is the same;[11] compare (6) (change-of-state) and (7) (static state).

(6) Central Kurdish Sanandaj (Mohammadirad 2022b: I, 1016)
   bū-m=a   ***wirdafirūš***
   be.SUBJ-1SG=DRCT peddler
   '(I will) become (a) **peddler**.'

(7) Southern Kurdish Bijar (Mohammadirad 2022c: D, 0282)
   *aware **kur** bī*   ...
   if   boy be.SUBJ.3SG
   'If it were **a boy** ...'

Similar phenomena have been noted for unrelated OV languages in close contact with Kurdish. For example, in the Northeastern Neo-Aramaic (NENA) dialect of the Jewish speech community from Urmi (West Iran), "the complement of the verb *qlb* 'turn into' is invariably placed after it" (Khan 2008a: 323). Although post-posing of complements of change-of-state predicates is widespread in the region, it is not grammaticalized to the same extent in all languages. Having outlined some of the main currents in earlier research and clarified terminology, in the remaining sections, we describe the design of the database and present two case studies illustrating cross-corpus results.

3 Design of the WOWA (Word Order in Western Asia) database

The WOWA sample includes data sets from 35 languages and varieties, based on monological, unscripted, spoken texts (see Figure 1 and the Appendices for details). As most of the project was conducted during the 2020–2022 pandemic, it was not possible to systematically select locations and languages in which to conduct dedicated fieldwork; rather, we have been obliged to rely on pre-existing

---

[11]Interestingly, post-verbal placement of a change-of-state complement is much more likely when the complement is nominal (e.g. 'she became (a) teacher'), rather than adjectival (e.g. 'she became rich').

resources. The result is that we were unable to compile a geographically or phylogenetically balanced sample of varieties. Nevertheless, the present sample represents the largest and most systematic data source currently available for investigating word order across the region.

The data sets stem from a range of distinct research contexts. Some are based on texts extracted from the published output of scholars working within individual philologies (e.g. the Neo-Aramaic texts of Barwar, Northern Iraq, originally published in Khan 2008b, a sub-set of which is analysed for a WOWA data set, Stilo 2021a). Other data-sets are taken from published sources of national language academies in the framework of dialect surveys (e.g. the Erzurum dialect of Turkish, which feed into Dogan 2021a), while others stem from contemporary language documentation projects, such as the Hazarrudi Tat texts used in Izadifar (2022) and the Qashqai texts in Schreiber (2021a).

For most data sets, the most widely represented genre is traditional narrative, but some data sets also include stimulus-based narratives (e.g. Pear story (Chafe 1980) retellings). The texts have been transcribed according to the academic tradition of the original researcher (we have not attempted to impose a common transcription scheme), and translated into English (in one case, into German). Generally, each data set includes more than one text, in most cases from different speakers; the composition of each data set is described in the accompanying metadata, and the source of each token (i.e. the individual text, and speaker) is recoverable. The main criteria for inclusion of a dataset in WOWA are a minimum yield of 500 codable tokens, reliable and authentic spoken data, and no restrictions on data accessibility.

3.1 Content of each data set

The list below provides the downloadable resource types included in WOWA. The first three are available for all data sets, while the other three are accessible to varying degrees, depending on the nature of the source data:

1. All files: Complete data set in a single ZIP-directory.

2. Coded values: The actual coded data (see below), in Excel and TSV format.

3. Metadata: A text document containing information on sources, references, speaker metadata, links, and other relevant information.

4. Source texts: Contains an orthographic rendering of the entire text with a translation, often from a published source, or provided by the contributor. In some cases, the source texts include additional morphological glossing or other information.

5. Sound files: The original sound files (where available), in .WAV and .MP3 formats.

3.2 Segmentation and token coding

The basic units of the database are tokens of prosodically independent (rather than bound), referential, non-subject constituents. Creating the database thus involves identifying the relevant tokens, and coding them for a series of features (see below). Texts selected for inclusion into the corpus are first segmented into strings that correspond approximately to meaningful utterances (in many cases this corresponds to a clause), termed utterance units. Each utterance unit is accompanied by a translation into English (column "utterance_translation" in Table 1 below). Utterance units are consecutively numbered and entered as single rows in the database, initially implemented in an Excel spreadsheet.

In a second step, all tokens of referential, non-subject constituents are identified and entered into a distinct cell (token) aligned with its source utterance unit. Note that clausal constituents (complement clauses, etc.) are not included as tokens. If an utterance unit contains more than one relevant token, that row of the data is repeated. If an utterance unit contains no relevant token (for example, a simple intransitive clause often does not contain any overt non-subject constituent, see 0006 in Table 1 below), then the token column remains empty. Note that these non-coded utterances remain in the data set, which thus preserves the overall unity of the original text, and maintains its re-usability for future research.

The basic structure is illustrated in Table 1, from the Hazarrudi Tat data set (Izadifar 2022). The utterance in 0006 does not contain a relevant token, thus the token columns are empty. The utterance in 0007, on the other hand, contains two relevant tokens ('some prey,' and 'there'). Each receives its own ID (0007 and 0008), and the utterance unit is repeated, enabling tokens to be systematically associated with their contexts across all analysis steps. WOWA currently contains approximately 20,000 analyzed tokens in context.

Table 1: Fragment of Hazarrudi Tat data set (Izadifar 2022)

token ID	utterance unit	utterance translation	token	token_translation
0001	*bale čemā rustā de i nefar ve*	yes, there was a person in our village	*čemā rustā de*	in our village
0002	*šekārči ve*	he was a hunter	*šekārči*	hunter
0003	*ševi šekār*	he had gone for hunting	*šekār*	hunting
0004	*ševi šekār čemā kua de*	he had gone hunting in our mountains	*šekār*	hunting
0005	*ševi šekār čemā kua de*	he had gone hunting in our mountains	*čemā kua de*	in our mountains
0006	*i jangali ve*	there was a forest	(no token)	
0007	*berā de i šekāri bezzeše*	he killed (hit) some prey there	*i šekāri*	some prey
0008	*berā de i šekāri bezzeše*	he killed (hit) some prey there	*berā de*	there

Once identified, each token is coded for a number of features, which fall into the following three types:

Doculect-related features:
– Genetic affiliation (e.g. Iranian, southwestern)
– Doculect location (latitude, longitude)

Context-related features
– Text and speaker identification (unique identifiers are assigned, which are described in the accompanying metadata document)

Linguistic features of the token and immediate context

– Classifiable versus non-classifiable (if an utterance unit contains either no relevant token or none that can be unambiguously classified). Non-classifiable tokens are not included in statistical analyses

– Pronominal versus nominal form

– Animacy

– Definiteness (only applied to direct objects)

– Weight

– Role (see (11))

– Flag (adposition, or case-marking)

– Position relative to the governing predicate (the dependent variable): before (0) vs. after (1)

– Comments (free text entry)

Obviously, the set of linguistic features could easily be extended to include, for example, main versus subordinate clause, finer-grained metrics of topicality, and so on. The final decision on which features to include was a compromise determined by the partially conflicting demands of theoretical relevance, and practical concerns such as economy of time and resources, simplicity of implementation across multiple languages with multiple coders, and replicability and transparency of coding-decisions. Previous research has pointed to the importance of pronominal versus nominal (e.g. Gerdes et al. 2021), animacy, weight (references in Section 5), informativity (Faghiri & Samvelian 2020), flagging and role (see Section 2 above), and these are also features that best satisfy the practical constraints just mentioned. Note that the raw data are available for coding additional features in the future. For each linguistic feature, coders select from a pre-defined set of options, which are explained in the Coding Guidelines.[12]

Coders work with the project coordinators, and problematic issues are resolved collaboratively to maximize cross-coder consistency. The coding scheme was presented and discussed collectively at two workshops (2019, 2020), and continued to evolve over the course of the project, before a final version was adopted in 2020. It should be evident that in a project of this nature, with multiple contributors working on multiple languages, compromise is inevitable. We have strived

---

[12]https://multicast.aspra.uni-bamberg.de/resources/wowa/data/_docs/guidelines/wowa_ coding-guidelines.pdf

to maintain the fine line between maximal simplicity and generality (limiting the number of coding options), while maintaining sufficient flexibility for capturing the range of cross-language variation contained in the data. Nevertheless, some degree of coding indeterminacy is inevitable, and for this reason, we include the coding option "other" in all linguistic categories to capture those instances where the analyst cannot decide among the available options. The full list of coding options is available in the Coding Guidelines; by way of illustration, we demonstrate in Table 2 the linguistic coding of the eight items from Table 1 above.

Table 2: Coding the linguistic values for the tokens in Table 1

token	token translation	pro	anim	weight	weight2	role	flag	position
čemā rustā de	in our village		inan	2	11	loc	postp	0
šekārči	hunter		hum	1	7	cop	bare	0
šekār	hunting		inan	1	5	Goal	bare	1
šekār	hunting		inan	1	5	Goal	bare	1
čemā kua de	in our mountains		inan	2	9	loc	postp	1
(no token)								
i šekāri	a prey		inan	2	7	do	bare	0
berā de	there		adv	1	6	loc	postp	0

The "pro" column is empty in Table 2, because there are no pronominal tokens in this stretch of discourse. The "weight" column records orthographic words, except function words solely employed as flagging devices (e.g. simple adpositions). The "weight2" column is a finer-grained weight metric that is automatically generated, based on the number of characters contained in the transcription of the token (thus the first token consists of 11 characters); it provides a rough proxy for the number of phonological segments in each token. The column "position" is the dependent variable, and offers a binary option of <0> (pre-verbal) versus <1> (post-verbal).

The proprietary spreadsheet format used for data entry was dictated by practical considerations; most contributors use MS Excel (or equivalent) and were able to enter their data into the template that we provided. For the actual analysis, data are exported to R, a powerful and flexible programming language and platform for statistical computing.

With regard to the pronoun category, we have included only prosodically independent pronouns as tokens. For languages that make extensive use of clitic object pronouns, this means that the number of classifiable object tokens in these languages may be very low, which has a detrimental impact on the statistical analysis (authors have the option of noting the presence of clitic pronouns in the comments column (e.g. Schreiber 2021b), so the information is available for future analyses.) There are sound empirical reasons for distinguishing free and clitic pronouns, illustrated below from spoken Persian (Rasekh-Mahand et al. 2024 [this volume]): around 95% of nominal direct objects precede the verb (OV), as in (8). Clitic object pronouns, on the other hand, frequently right-attach to the verb, and indeed must do so if the verb is the sole available host, as in (9) and (10):

(8)　Colloquial New Persian (Izadi 2022: C, 0263)
　　　doz=râ　　*bord*　　　*bâlâ*
　　　dosage=ACC carry.PST.3SG upwards
　　　'(He) increased **the dosage**.'

(9)　Colloquial New Persian (Izadi 2022: V, 2375)
　　　*mi-šenâs-im=**ešân***
　　　INDIC-know.PRS-1PL=3PL
　　　'We know **them**.'

(10)　Colloquial New Persian (Haig & Rasekh-Mahand 2022: oh_f_accident_0166)
　　　*be-bar-id=**aš***
　　　IMPER-take.PRS-2PL=3SG
　　　'Take **him**!'

It would make little sense to count constructions such as (9) and (10) as 'VO,' apparently in contrast to the OV of (8). Examples (9) and (10) illustrate a language-specific rule of cliticization, which permits no variability of object placement in these examples. Clitic placement is a fascinating issue in its own right but of limited relevance for the principles operating in the linearization of independent phrases in syntax. Consequently, clitic pronouns hosted by the predicate are not included in calculations of pre- versus post-verbal argument placement. Clitic pronouns hosted by an item distinct from the predicate, on the other hand, are coded as "bound," and the normal coding procedures applied. Depending on the analysis, pronominal tokens may be filtered out of a given sample.

4 The impact of semantic role: the "Goals Last" effect

4.1 Background

For the majority of languages in the sample, the variable "Role" turned out to be the most influential factor in determining pre- versus post-verbal position. The category "Role" in WOWA distinguishes the 19 categories shown in (11).

(11)   Role categories recognized in WOWA (see Coding Guidelines, Section 3.2)

ABL   source of motion ('she came out of the house')

ADDR   addressee of a verb of speech ('they spoke to him/asked her/begged the King')

BECM   'become,' i.e. the final state of a change-of-state (inchoative), predicate, such as 'become X,' 'turn into X'

BECM-C   final state of a caused change-of-state predicate ('they made him King,' 'she turned him to stone')

BEN   benefactive; a person who benefits, or is disadvantaged, by an event without being directly impinged on by the action

COM   comitative; a person who accompanies another participant in some action, or state ('I went to the market with my father')

COP   complement of a copular expression ('they were farmers')

COP-LOC   locational complement of a copular expression ('she was in the car')

DO   direct object, which needs to be identified on language-specific criteria such as typical case marking properties

DO-DEF   definite direct object (which will include most pronouns), i.e. an item whose identity is recoverable from the context through previous mention or assumed deictic reference ('she took that cup')

GOAL   endpoint or destination of a verb of motion ('it fell on the table')

GOAL-C   endpoint or destination of a verb of caused motion ('he put it on the table')

INSTR   instrument for carrying out an action

LOC   static location (with no implication of movement) of a participant or event

OTHER   none of the available categories

POSS   possessed in a clause expressing possession 'she had two brothers,' unless the language has a HAVE verb and expresses the possessed in the same way as a direct object (do)

<small>REC</small>   recipient of a theme in an event of transfer, typically GIVE

<small>REC-BEN</small>   recipient-benefactive. This is included for contexts in which it is unclear whether a particular token is the recipient, or a benefactive of an action ('he bought the apples for us' — recipient or benefactive?)

<small>STIM</small>   stimulus, typically of verbs of emotion, perception, desire — if they are not coded as direct objects (English 'she was afraid of the snake' (stim), but not 'she hates snakes' (coded as <do>))

The data reveal very divergent token frequencies of different roles. In fact, some are so infrequent that they offer little leverage for statistical purposes. Figure 2 below provides the respective proportions of different roles, whereby we have lumped together those role types that occur only marginally.
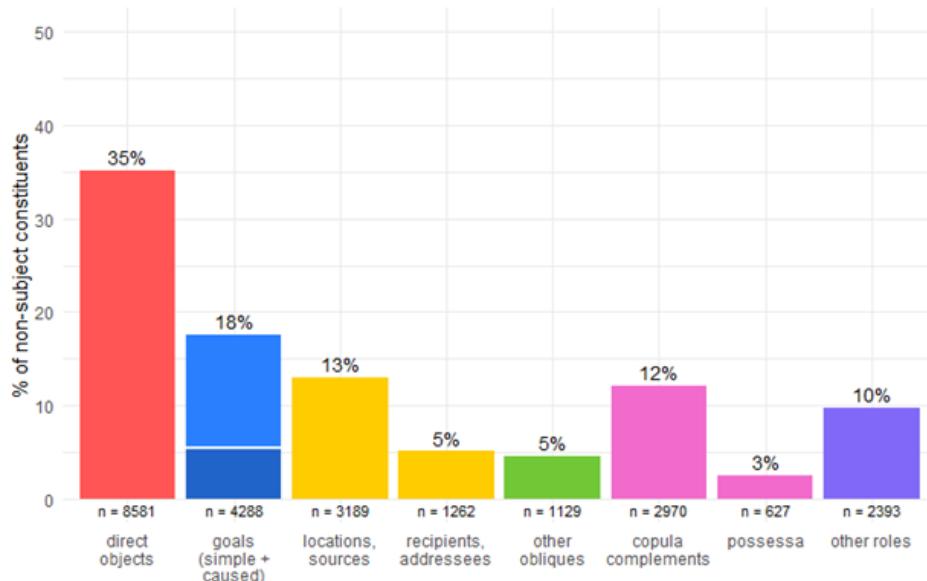
Figure 2: The respective frequency of non-subject roles across the WOWA sample

The most frequent role type in our data are direct objects, which account for around one third of all tokens. The next most frequent are goals of motion, including both simple motion and caused motion (distinguished in Figure 2 by different shades of blue). Recipients and addressees are relatively infrequent, and are therefore lumped together here (see below). Similarly, roles such as instrument, comitative, stimulus, and benefactive do not occur in sufficient numbers

to permit meaningful quantitative analysis, so have been combined under "other obliques." The "other role" category includes tokens that were not classifiable under any of the available role-categories.

Turning now to the respective frequencies of post-verbal placement, Table 3 visualizes the general trend discernible across the data set. Four categories are distinguished: Goals (including caused goals), Recipient+Addressee, other obliques, and direct objects. For each of those four categories, we have colour-coded any frequency values for post-verbal placement which exceed 66%.

The data in Table 3 permit a number of generalizations, which to our knowledge have hitherto not been recognized. First of all, the data underscore the exceptional status of spatial Goals when compared to any other role. More than two thirds of the sample languages have dominant post-verbal placement of Goals (>66%), and for the majority of these languages, dominant post-verbal placement is restricted only to Goals. In several of these languages, the frequency of post-verbal Goals is twice as high as the frequency of post-posing any other constituent. In other words, Goals are different. We can formulate this as a potential universal in (12):

(12)   If a language postposes any role with a greater than two-thirds frequency, it will postpose Goals.

Note that Table 3 combines the two roles Recipient and Addressee, due to the low absolute numbers of tokens in these categories. However, this obscures the fact that Recipients and Addressees do not actually pattern consistently across all data sets. In fact, three distinct patterns can be identified in the sample, bearing in mind that the relevant absolute numbers are small:

  (i)   both Addressee and Recipient occur before the verb (e.g. Oghuz Erzurum, Zazaki Cewlig, modern spoken Persian);

 (ii)   both occur after the verb (e.g. NENA Jewish Duhok, Central Kurdish Sanandaj, Kumzari);

(iii)   Recipients occur after the verb, but Addressees before the verb (e.g. Vafsi Gurchani, Northern Kurdish Muš, or Armenian of Agulis).

Crucially, no language is attested in the sample that has post-verbal Addressees, but pre-verbal Recipients. Bearing this in mind, we can formulate the following

Table 3: Frequency of post-verbal placement for different roles; shaded values indicate frequencies above 66%. Goals and direct objects include only nominal expressions; Recipients, Addressees, and other oblique roles (Locations, Sources, Instrumentals, Benefactives, and Comitatives) also include pronominal expressions. Two doculects with less than 8 observations in each role category are excluded. See Appendix B for raw data. (NENA = Northeastern Neo-Aramaic).

	Doculect	% post-verbal Goals	Recipients + Addressees	Other Obliques	Direct Objects
⬠	*Laz (Arhavi)*	4	6	1	3
◆	*Persian (New, Early Classical)*	5	0	9	2
●	*Oghuz (Ankara)*	7	19	7	2
●	*Oghuz (Erzurum)*	38	8	9	7
◆	*Balochi (Turkmen)*	48	14	6	2
◆	*Kurdish (Northern, Ankara)*	59	13	6	0
✗	*Kholosi (Kholos)*	62	39	1	1
◆	*Mazandarani (Kordxeyl)*	63	12	8	3
◆	*Balochi (Coastal)*	63	28	18	7
◆	*Bashkardi (Northern)*	63	56	31	27
●	*Oghuz (Bayat)*	64	43	13	4
◆	*Zazaki (Cewlig)*	91	15	0	5
●	*Oghuz (Tabriz)*	75	21	13	1
◆	*Persian (New)*	84	25	19	5
◆	*Kurdish (Southern, Bijar)*	97	27	11	2
◆	*Tat (Daykuscu)*	78	34	30	15
◆	*Kurdish (Northern, Muš)*	89	36	5	3
◆	*Bashkardi (Southern)*	80	38	36	11
◆	*Vafsi (Gurchani)*	88	38	11	2
◆	*Kurdish (Northern, Lachin)*	81	40	4	2
◆	*Balochi (Koroshi)*	90	42	27	2
◆	*Talyshi (Lerik)*	72	45	26	2
●	*Oghuz (Gagauz)*	73	47	38	51
◆	*Zazaki (Siwereg)*	100	52	8	5
◆	*Tati (Hazarrudi)*	92	53	16	3
●	*Oghuz (Qashqai)*	71	88	16	8
◆	*Kurdish (Central, Sanandaj)*	94	95	22	1
◆	*Gorani (Gawraju)*	96	82	36	5
▲	*NENA (Jewish, Sanandaj)*	92	81	40	5
◆	*Kumzari (Musandam)*	100	97	87	7
▲	*NENA (Christian, Barwar)*	96	100	74	83
▲	*NENA (Jewish, Dohok)*	99	98	93	90
▲	*Arabic (Jewish, Baghdad)*	100	100	85	97

implicational hierarchy indicating the conditions under which a particular role type may occur as dominant (at least 66%) postverbal.[13]

(13)   Goals > Recipients > Addressees > Other > Direct object

The hierarchy in (13) is to be understood as an implicational universal, which can be formulated as follows: based on token frequency in corpora of spontaneous spoken language, and frequencies of nominal as opposed to pronominal constituents, if a language postposes any of the roles in (12) with greater than two-thirds frequency (dominant), then it will also dominantly postpose all higher roles on the hierarchy. Thus for the languages in the WATZ sample — regardless of genetic affiliation — there is no language that has, for example, dominant post-verbal Addressees, but not dominant post-verbal Goals. There are, however, a considerable number that only have dominant post-verbal Goals, and no other roles. Finally, if any language has dominant post-verbal direct objects, then all other roles are likewise dominant post-verbal. Thus there are no languages that, for example, combine dominant post-verbal objects with pre-verbal Goals.

Currently it is impossible to say with certainty whether the regularities illustrated in Table 3 and expressed in (12) and (13) represent a peculiarity of the languages in and around WATZ, or whether they reflect a deeper trait of connected spoken language, which should surface in spoken-language corpora of any language. We suggest there are grounds for assuming that (13) does reflect, at least in part, universal tendencies. Haig & Schiborr (In review) compare the two ends of the hierarchy (direct objects and Goals) across a more extensive sample of spoken language corpora and report that in no corpus do frequencies of postverbal objects exceed frequencies of postverbal Goals, regardless of language type or genetic affiliation.

Even if only parts of (13) should turn out to be valid outside of WATZ, this would have considerable implications for understanding, for example, diachronic change in word order. Essentially, (13) predicts an ordered sequence in a shift from OV to VO and vice versa. If a language is, for example, consistently verb-final (i.e. no role is dominant post-verbal), then (13) predicts that any change towards less verb-finality would occur first with Goals, and proceed down the

---

[13]The hierarchy in (13) is similar to others that have been formulated in the literature (e.g. Frommer 1981, Haig & Khan 2019, Stilo 2018b), some of which add additional roles, or employ somewhat different terminology. One role that has often been included is Benefactive; however, our data for this role are sparse, and were unfortunately not always consistently coded, rendering interpretation of the results difficult. Currently they are included under 'other'; this requires more research.

hierarchy, with direct objects the last to shift. For the other direction, i.e. a language that is consistently verb-initial in the VP (like English), the prediction is that if any argument type shifts across the verb, it will be direct objects first, and Goals last. There is thus an asymmetry in the way VO and OV languages can be expected to move closer to one another. Preliminary observation of word-order change in WATZ suggests that this holds, regardless of whether the shift is considered internally motivated, or contact induced.

At this point, the relationship between recipient and Goal roles merits discussion. In earlier work (Haig 2022d), it was argued, on the basis of Kurdish data, that these roles share a common semantic component, defined as "event endpoint," which motivated the shared post-verbal syntax in Kurdish. However, as we have seen, for the majority of other languages in our sample, recipients and goals do not pattern alike. On the assumption that both share endpoint semantics, the question arises as to what inhibits post-verbal placement of recipients? To understand this, it is important to recall that word order is the product of competing motivations, of which iconicity is but one. These include verb-object adjacency (the tendency for direct objects and verbs not to be separated by other constituents), weight, animacy, and agency considerations. Thus word order in any given context is the product of multiple factors, including information structure, semantics, and configurational constraints. Recipients differ strikingly from goals in several dimensions relevant here: they are overwhelmingly human, with high frequency pronominal and first or second person, and are treated syntactically as direct objects in many languages (Haspelmath 2015). Thus we suggest that endpoint semantics are simply overridden by other factors in the ordering of recipients. The distinction between goals of verbs of motion, and recipients is clearly maintained in Haspelmath's (2015) concept of "ditransitive construction," which presupposes an element of transfer of possession, while goals of verbs of caused-motion, such as 'put,' lack such an entailment and are thus outside the purview of the typology of ditransitive constructions. We might add that conversely, transfer of possession does not necessarily entail movement: it is possible to give someone a house, or a piece of land, which involves no actual change of location of the "theme." Our revised conclusion is thus that although shared endpoint semantics mean goals and recipients may pattern alike in some languages, the overall weight of evidence suggests that a distinction should be maintained (see Haig & Schiborr In review for a more detailed discussion).

5 The impact of weight on post-posing

5.1 Background

It is fair to say that in both experimental and corpus-based approaches to word-order typology, considerations of weight (however formalized) have attracted more attention than any other single factor (e.g. Faghiri & Samvelian 2020, Schnell & Schiborr 2022, Wasow 2022: 5–10). However, as Yao (2018) points out, most of the relevant research considers weight as a factor in determining the relative order of **constituents occurring on the same side of the predicate**, for example the relative ordering of the two PP's in (14) (from Wasow 2022: 6).

(14)   *The gamekeeper looked [through his binoculars] [into the blue but slightly overcast sky].*

For languages such as English, which regularly place objects and other non-subject verbal dependents after the verb (the "post-verbal domain," Yao 2018), it seems that short constituents tend to precede longer constituents ("short before long"), as illustrated in (14). But this trend is apparently reversed for the pre-verbal domain in head-final languages like Japanese, where long constituents reportedly preferably precede short (Yamashita & Chang 2001). However, it is less clear which prediction would hold in languages which permit constituents to occur on either side of the predicate ("cross-domain NP shift," Yao 2018). Yao (2018) investigates cross-domain NP shift for object placement in Mandarin, which varies between a post-verbal (VO) and a pre-verbal (*bǎ*-OV) option. Interestingly, this detailed study reveals no linear correlation between VO versus OV, and NP-length. Levshina (2019: 560) notes a significant effect of length only for VO languages, and most notably for clausal rather than nominal constituents. Research on diachronic syntax does consider cross-domain shift for direct objects, for example pragmatically driven object fronting in VO languages, and "heavy NP shift" in OV languages (Faarlund 2010: 205). However, for other kinds of constituent there is a general lack of research that would guide the expectations of a length effect for constituent order relative to the verb.

5.2 Method and results

In the absence of a clear hypothesis from the literature, here we present an initial exploration of length effects on pre- versus post-verbal placement of different constituents in the WOWA sample. It is important to note that overall, the leverage of the weight factor is significantly reduced in our spoken-language WOWA

data, when compared to the written-language or experimental corpora that form the basis of most previous research. Schnell & Schiborr (2022: 178) observe that in the spoken language corpora from the Multi-CAST collection (Haig & Schnell 2023), almost 90% of all NPs in the data consist of maximally three words, with the majority being two words or less. In written Universal Dependency corpora on the other hand, 36% of NPs contain four or more words (Schnell & Schiborr 2022: 178–179). Note furthermore that for WOWA, we have not included clausal constituents in our analyses, which rules out the kinds of very long tokens that figure in written language corpora (complement clauses, NPs with embedded relative clauses, etc.).

For the WOWA project, we operationalized "weight" with two measures: (i) length of the object phrase in words; and (ii) a finer-grained metric using characters, which provides a proxy for phonological weight. The data shows a strong power law-like distribution of weights, with 64% of analyzed tokens in certain roles (direct objects, Goals, Recipients and Addressees, as well as Locations, Sources, Instrumentals, Benefactives, and Comitatives) consisting of a single word and 91% of two or fewer words. Similarly, for weight in characters, 67% of these tokens contain 8 or fewer characters, and 91% contain 13 or fewer.

Two analyses were conducted both based on the finer-grained character-based weight metric. The first includes 33 data sets in WOWA, split by role; we exclude those data sets that have fewer than 8 observations in a role or display no variation in positioning. Table 4 shows the mean correlation (Pearson) between pre- and post-verbal positioning (0, 1) and weight.

Table 4: Pearson correlation of weight with position

Roles	R-value	SD	Observations
direct objects	+0.007	0.117	8364
goals	+0.022	0.161	4172
recipients+addressees	−0.046	0.251	1340
other obliques	+0.005	0.103	4444

All correlation coefficients hover around zero, with no substantial variation between data sets. Only a small handful of data sets have a coefficient exceeding a value of ±0.4 in any of the roles, three of which are for Recipients/Addressees, which due to their comparative rarity unfortunately offer the least robust results in general.

The second analysis takes into consideration the basic word order of each doculect, because as Wasow (2022: 11) notes, different predictions hold for head final versus other languages. Consequently, we follow Levshina (2019) and divide the sample doculects into three groups, based on the frequency of nominal post-verbal direct objects in the corpora (the sample is not balanced across these three groups, due to the dominance of OV Iranian languages): "OV" (<25% VO), N=13; "mixed" (25–75% VO), N=16; and "VO" (>75% VO), N=4. Figure 3 shows the mean weight in characters for four role types, distinguishing pre- and post-verbal placement. For each role type, we present the findings split according to the three word-order types mentioned above (<25% VO; 25–75% VO; >75% VO).

Figure 3: Mean weight in characters of pre- and post-verbal constituents for four role types, split according to word-order type of the doculect

Figure 3 suggests a weak correlation between post-verbal placement and higher weight, for direct objects (bottom right panel) across all word order types,

but none of the individual differences reach significance. Any claim for a correlation can only therefore draw on the fact that the minimal differences are in the same direction for each language type. For other roles, no consistent pattern can be identified. Turning to word-order type, it is only the >75% VO languages that exhibit a weak but consistent association of weight and likelihood of post-verbal placement, with the strongest effect occurring for Goals in dominantly VO languages; this would merit closer investigation, but note the low absolute figures (n=15) for pre-verbal Goals in the four VO doculects.

In sum, our investigation of the effect of weight reveals only weak effects for only some roles, and some language types. This is partly attributable to the aforementioned narrow envelope of variation for weight in spoken language, where the overwhelming majority of tokens consist of only one to two words. As noted, clausal constituents were not considered in our data. Equally, we emphasize that our investigation considers weight as a factor in cross-domain shift, i.e. shift across the predicate, as opposed to relative position of constituents on the same side of the predicate, which has been the focus of most existing research (see Wasow 2022). As Yao (2018) notes, research on cross-domain weight effects is scarcely available, and their own results, like ours, reveal no clear weight effects of weight. We provisionally conclude that weight effects noted in the literature do not carry over to cross-domain word order variation in spoken language. This is definitely an area that would merit further research; see Skopeteas 2024 [this volume], on the contrasting prosodic properties that hold in the pre- and post-verbal domain respectively.

6 Summary and prospects, residual issues

In sections one to three above, we have outlined the rationale, research context, and methodologies implemented in the WOWA project. In Sections 4 and 5, we illustrate two use cases for exploring the entire database. The results for semantic role provide abundant evidence for the special role of spatial Goals in word order variability, in particular of the dominantly OV languages of Western Asia. The findings for weight, however, do not yield a simple picture, suggesting that cross-domain word-order variation (Yao 2018) requires a distinct set of explanations to those that have been proposed for same-domain word order, which focusses on the respective ordering of constituents occurring on the same side of the verb (Wasow 2022).

We hope that our research will stimulate further research in this direction, and that in the future, the hitherto neglected effects of semantic role are afforded due

consideration. In Section 5 we demonstrate that role provides the best overall predictor of post-verbal placement, with Goals outstripping any other roles by a considerable margin. We formulated our findings in the form of an implicational universal (13), which embodies a number of testable hypotheses for future work on spoken-language corpora, and also has considerable implications for understanding word order change.

Our findings also lend broad support for the concept of Transition Zone, indicating a gradual shift towards higher frequencies of verb-final constituents in the westward regions of WATZ. However, we require a more balanced and denser sample of doculects to develop a more robust framework for mapping structural variation to geospatial features, and to control for phylogenetic distance. Other issues that remain to be considered are measures of corpus-internal variation (see Craevschi 2022 for provisional findings), co-argument effects, and the role of additional morphosyntactic features such as agreement, clause type (modality, negation, subordination, etc.), and more nuanced controlling for information structure (see Hodgson et al. 2024 [this volume], Noorlander 2024a [this volume],b [this volume]).

Finally, our data point to the potential impact of register and modality (spoken versus written language) on word order. While the overwhelming majority of data in WOWA represents informal spoken language, in those data sets where data from more formal registers are available, they indicate some striking differences in word order (Nourzaei & Haig 2024 [this volume], Rasekh-Mahand et al. 2024 [this volume] and Parizadeh & Rasekh-Mahand 2024 [this volume] on Persian, Hodgson et al. 2024 [this volume] on East Armenian). These differences invite further research, but in the meantime we urge caution when comparing cross-linguistic data, and emphasize the necessity for controlling for modality and register.

7 The organization of the volume

The volume consists of 16 chapters, divided into four sections, each of which is introduced below:

    I Theoretical and methodological issues (Chapters 1–3);

   II Case studies from Iranian and Indo-Aryan languages (Chapters 4–9);

  III Case studies from the Caucasus and Black Sea (Chapters 10–13);

  IV Case studies from Semitic languages (Chapters 14–16).

7.1 Section I: Theoretical and methodological issues

Section I includes the current introductory chapter, and two further chapters. Chapter 2 by Kateryna Iefremenko investigates elements in the post-verbal domain of young adult bilingual speakers of Kurmanji and Turkish in Ankara in comparison with Turkish in Erzurum and under consideration of the sociolinguistic dichotomy between Turkish as dominant national language and Kurmanji as regional language. Although the findings on elements in the post-verbal domain in the two languages are generally in line with previous research, the results show that the Turkish dialect of Erzurum tends to have more frequent post-verbal Goals than other varieties of Turkish, which only apply post-verbal positions based on information structure and weight considerations. The higher rates of post-verbal Goals in the Erzurum dialect may plausibly reflect contact influence from neighbouring Kurmanji Kurdish dialects, which exhibit typical Iranian post-verbal Goals of motion and caused motion. In one particular construction *erdê ketin* 'to fall on the ground' the Goal nevertheless appears predominantly pre-verbal among bilingual Kurmanji Kurdish speakers; it may be potentially modelled on Tr. *yere düşmek* 'to fall on the ground.' The methodology of this study is unique in the context of the volume, in that explicitly bilingual data were analysed that were elicited from the speakers by means of video prompts.

In Chapter 3, Stavros Skopeteas investigates prosodic structure in the preverbal and post-verbal domain in a sample of (primarily OV) languages that includes Turkish, Georgian, Caucasian Urum, Eastern Armenian, and Persian. Skopeteas identifies three main types of OV languages, distinguished according to the nature of constraints that determine whether objects may occur in the post-verbal domain. In some languages, post-verbal objects are very restricted, and are only permitted if they are outside the focus domain of the clause, i.e. express given information, or afterthoughts (e.g. Standard Turkish). In other languages, post-verbal objects are permitted as part of broad sentence focus (e.g. Persian), while in others, even objects with narrow focus are also permitted (e.g. Georgian). In a sense, then, these three types represent increasing levels of tolerance for the integration of focal material into the post-verbal domain. The author reviews extant research on these languages and reports experimental results that illustrate the typology, and explores the interaction of prosodic and syntactic phrasing. This line of research complements the corpus-based approach of most contributions, which capture frequency patterns of linear ordering in naturalistic discourse, but leaves little space for systematic investigation of prosodic structure.

7.2 Section II: Case studies from Iranian and Indo-Aryan languages

This is the largest section in the book and contains six chapters, each of which deal with one (or a group of) Iranian or Indo-Aryan languages. In Chapter 4, Maryam Nourzaei and Geoffrey Haig present an overview of word order across three varieties of Balochi, each from areally diverse locations. The results provide further confirmation of the overall trend identified in this volume, that proximity to Mesopotamia correlates with an overall increase in post-verbal constituents; the westernmost variety of Balochi (Koroshi) exhibits both overall higher frequencies of post-verbal constituents, but also a greater range of role types permitted in this position, when compared to the two more easterly varieties. In Chapter 5, Agnes Korn presents data from two varieties of Bashkardi in southern Iran. The data stem from legacy materials recorded in the 1950's, providing a rare opportunity to explore the possibility of recent changes in the language, but also for considering micro-variation across the two varieties.

Chapter 6 (Maryam Nourzaei) illustrates the only Indo-Aryan language in the sample, Kholosi, a language island in southern Iran that has preserved aspects of Indo-Aryan morphosyntax, but has adapted in word order to conform with the post-verbal placement of spatial Goals that characterizes all of its currently neighbouring languages. In Chapter 7, Mohammad Rasekh-Mahand and co-authors provide an in-depth study of spoken Persian, comparing the recent HamBam data with the results of Frommer (1981). For the least formal registers of Persian, they report stable values over the 40-year time span with regards to most aspects of post-verbal syntax, but note a shift in register distribution in the modern data when compared to the older sample. Chapter 8 is the sole chapter based on written data, and investigates a sample of Early New Persian texts (10–13th Century CE). The texts reveal some internal variation, but an overall remarkably consistent degree of verb-finality, with little evidence for the post-verbal syntax that characterizes all spoken western Iranian languages investigated so far. These findings raise questions regarding the diachronic development of post-verbal syntax in West Iranian, but also regarding the relationship between the spoken and written languages; it is possible (and we believe plausible) that the Early New Persian texts are not representative of the spoken language of the time, any more than today's formal written Persian texts are representative of contemporary spoken Persian. In Chapter 9, Masoud Mohammadirad takes a comparative look at three varieties from the Zagros region (Gorani Gawraǰu; Central Kurdish Sanandaj; Southern Kurdish Bijar). The findings are suggestive of Gorani substrate effects in southernmost dialects of Central Kurdish.

7.3 Section III: Case studies from the Caucasus and Black Sea

In Chapters 10 and 11, Diana Forker investigates word order in Kartvelian and East Caucasian and Adyghe respectively. The data come from several sources, mostly outside the WOWA framework, but can be interpreted within the same framework. Role effects (Goals) are noticeable, though considerably less prevalent than in the Iranian languages and other languages of Mesopotamia. In Chapter 12, Laurentia Schreiber and Mark Janse investigate word order patterns in Romeyka in bilingual speakers under language shift to Turkish. While information structure and phrase type are the most relevant factors determining the dominant word orders in Romeyka, significant inter-speaker variation indicates the ongoing process of language shift. Chapter 13 presents original spoken-language data from East Armenian (Katherine Hodgson, Victoria Khurshudyan and Pollet Samvelian). This research adds a new perspective to the growing literature on word order in East Armenian, complementing existing research based on experimental and written-language data. The authors confirm a definiteness effect on direct object ordering, with definite direct objects showing greater word-order flexibility with respect to the verb (higher frequency of VO ordering), while indefinite objects remain overwhelmingly OV. They also identify inter-speaker variation and the effect of register. The Goals Last effect documented for most of the language of WATZ (Section 4 above) is also confirmed in these data, though in somewhat weaker magnitude than in the Iranian languages of Mesopotamia.

7.4 Section IV: Case studies from Semitic languages

This section includes three contributions on Semitic languages. In Chapter 14, Bettina Leitner describes the basic word order profile of Khuzestani Arabic, a linguistic island of Arabic in Iran, and discusses reasons for deviations from the default word order VX, such as language contact and internal change. In Chapter 15, Paul Noorlander discusses Neo-Aramaic dialects in Iran and northeastern Iraq, which include at least one dialect that has undergone a complete shift from VO to OV (Jewish Urmi). While the impetus for the shift is almost certainly language contact, Noorlander illustrates how internal factors, in particular information structure, shape the way these changes have played out. Paul Noorlander also contributes Chapter 16 on Arabic and Neo-Aramaic in Eastern Anatolia, a region of high linguistic diversity. Noteworthy findings include the variability in copula construction s, which contrasts with the otherwise fairly regular presence of clause-final copula elements in most of WATZ.

References

Anonby, Christina van der Wal. 2015. *A grammar of Kumzari: A mixed Perso-Arabian language of Oman*. Leiden: Leiden University. (Doctoral dissertation).

Asadpour, Hiwa. 2022a. *Typologizing word order variation in northwestern Iran*. Frankfurt am Main: Goethe-Universität. (Doctoral dissertation).

Asadpour, Hiwa. 2022b. Word order in Mukri Kurdish - The case of incorporated targets. In Hiwa Asadpour & Thomas Jügel (eds.), *Word order variation: Semitic, Turkic and Indo-European languages in contact*, 63–87. Berlin: De Gruyter Mouton.

Asadpour, Hiwa & Thomas Jügel (eds.). 2022. *Word order variation: Semitic, Turkic and Indo-European languages in contact*. Berlin: De Gruyter Mouton.

Bağrıaçık, Metin. 2018. *Pharasiot Greek: Word order and clause structure*. Ghent: Ghent University. (Doctoral dissertation).

Berez-Kroeker, Andrea L, Lauren Gawne, Susan Smythe Kung, Barbara F Kelly, Tyler Heston, Gary Holton, Peter Pulsifer, David I Beaver, Shobhana Chelliah, Stanley Dubinsky, et al. 2018. Reproducible research in linguistics: A position statement on data citation and attribution in our field. *Linguistics* 56(1). 1–18.

Bickel, Balthasar. 2017. Areas and universals. In Raymond Hickey (ed.), *The Cambridge handbook of areal linguistics*, 40–54. Cambridge: Cambridge University Press.

Chafe, Wallace. 1980. *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production*. Norwood, NJ: Ablex.

Comrie, Bernard. 1989. *Language universals and linguistic typology*. 2nd edn. Chicago: University of Chicago Press.

Craevschi, Alexandru. 2022. *Historical contingency and typological tendencies in languages of Western Asia: A quantitative study of word order of non-subject constituents*. Bamberg: University of Bamberg. (MA thesis).

Dabir Moghaddam, Mohammad. 2018. Typological approaches and dialects. In Anousha Sedighi & Pouneh Shabani-Jadidi (eds.), *The Oxford handbook of Persian linguistics*, 52–88. Oxford: Oxford University Press.

Demir, Netîce & Mahîr C. Doğan. 2021a. Zazakî (Çewlîg). In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation*. Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

Demir, Netîce & Mahîr C. Doğan. 2021b. Zazakî (Siwêreg). In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in*

word order variation. Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

Dogan, Mahîr. 2021a. Oghuz (Erzurum). In Geoffrey Haig, Donald Stilo, Nils N. Schiborr & Mahîr Dogan (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* (accessed on October 20, 2022). Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa.

Dogan, Mahîr. 2021b. Oghuz (Gagauz). In Geoffrey Haig, Donald Stilo, Nils N. Schiborr & Mahîr Dogan (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

Dogan, Mahîr. 2022. Vafsi (Gurchani). In Geoffrey Haig, Donald Stilo, Nils N. Schiborr & Mahîr Dogan (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

Dryer, Matthew S. 1997. On the six-way word-order typology. *Studies in Language* 21(1). 69–103.

Dryer, Matthew S. 2007. Word order. In Timothy Shopen (ed.), *Language typology and syntactic description*, vol. 1, 61–131. Cambridge: Cambridge University Press.

Dryer, Matthew S. 2013a. Against the six-way order typology, again. *Studies in Language* 37. 267–301.

Dryer, Matthew S. 2013b. Order of adposition and noun phrase. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online.* Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info/chapter/85.

Dryer, Matthew S. 2013c. Order of object and verb. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online.* Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info/chapter/83.

Du Bois, John. 1987. The discourse basis of ergativity. *Language* 63(4). 805–855. DOI: 10.2307/415719.

Du Bois, John. 2003. Argument structure: Grammar in use. In John Du Bois, Lorraine Kumpf & William Ashby (eds.), *Preferred argument structure*, 11–60. Amsterdam: Benjamins.

Faarlund, Jan Terje. 2010. Word order. In Silvia Luraghi & Vít Bubeník (eds.), *Continuum companion to historical linguistics*, 201–211. London/New York: Continuum.

Faghiri, Pegah & Pollet Samvelian. 2020. Word order preferences and the effect of phrasal length in SOV languages: Evidence from sentence production in Persian. *Glossa: A journal of general linguistics* 5(1). DOI: 10.5334/GJGL.1078.

Faghiri, Pegah, Pollet Samvelian & Barbara Hemforth. 2018. Is there a canonical order in Persian ditransitive constructions? Corpus based and experimental studies. In Agnes Korn & Andrej Malchukov (eds.), *Ditransitive constructions in a cross-linguistic perspective*, 165–185. Wiesbaden: Reichert.

Forker, Diana. 2024a. Post-predicate elements in Adyghe. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 309–336. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266351.

Forker, Diana. 2024b. Post-predicate elements in Kartvelian and East Caucasian. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 283–307. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266349.

Friedman, Victor A & Brian D Joseph. 2017. Reassessing Sprachbunds: A view from the Balkans. In Raymond Hickey (ed.), *The Cambridge handbook of areal linguistics*, 55–87. Cambridge: Cambridge University Press.

Frommer, Paul. 1981. *Post-verbal phenomena in colloquial Persian syntax*. Los Angeles: University of Southern California. (Doctoral dissertation).

Futrell, Richard, Roger P Levy & Edward Gibson. 2020. Dependency locality as an explanatory principle for word order. *Language* 96(2). 371–412.

Gerdes, Kim, Sylvain Kahane & Xinying Chen. 2021. Typometrics from implicational to quantitative universals in word order typology. *Glossa: A journal of general linguistics* 6(1). 17.1–31.

Haig, Geoffrey. 2015a. Verb-goal (VG) word order in Kurdish and Neo-Aramaic: Typological and areal considerations. In Geoffrey Khan & Lidia Napiorkowska (eds.), *Neo-Aramaic and its linguistic context*, 407–425. Piscataway, NJ: Gorgias Press.

Haig, Geoffrey. 2015b. VG-word order in Kurdish and Neo-Aramaic: Typological and areal factors. *Presentation at the Cambridge Neo-Aramaic Conference, Cambridge University, 6-8th July 2011*. Cambridge.

Haig, Geoffrey. 2017. Western Asia: East Anatolia as a transition zone. In Raymond Hickey (ed.), *The Cambridge handbook of areal linguistics*, 396–423. Cambridge: Cambridge University Press.

Haig, Geoffrey. 2022a. Balochi (Turkmenistan). In Geoffrey Haig, Donald Stilo, Nils N. Schiborr & Mahîr Dogan (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* Bamberg: University of Bamberg. multicast . aspra . uni - bamberg . de / resources/wowa/.

Haig, Geoffrey. 2022b. Kumzari (Musandam). In Geoffrey Haig, Donald Stilo, Nils N. Schiborr & Mahîr Dogan (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* Bamberg: University of Bamberg. multicast . aspra . uni - bamberg . de / resources/wowa/.

Haig, Geoffrey. 2022c. Kurdish (Northern, Muş). In Geoffrey Haig, Donald Stilo, Nils N. Schiborr & Mahîr Dogan (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* Bamberg: University of Bamberg. multicast . aspra . uni - bamberg . de / resources/wowa/.

Haig, Geoffrey. 2022d. Post-predicate constituents in Kurdish. In Yaron Matras, Geoffrey Haig & Ergin Öpengin (eds.), *Structural and typological variation in the dialects of Kurdish*, 335–377. Cham: Springer International Publishing. DOI: 10.1007/978-3-030-78837-7_8.

Haig, Geoffrey & Geoffrey Khan. 2019. Introduction. In Geoffrey Haig & Geoffrey Khan (eds.), *The languages and linguistics of Western Asia: An areal perspective*, 1–29. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110421682-001.

Haig, Geoffrey, Paul M. Noorlander & Nils N. Schiborr. In press. Which word order features are stable in a contact setting? Corpus-based evidence from the Western Asian transition zone. In Jeroen Darquennes, Joe Salmons & Wim Vandenbussche (eds.), *Language contact: An international handbook*, vol. 2. Berlin: De Gruyter Mouton.

Haig, Geoffrey & Mohammad Rasekh-Mahand. 2022. *HamBam: The Hamedan-Bamberg corpus of contemporary spoken Persian.* multicast.aspra.uni-bamberg. de/resources/hambam/ (10 July, 2023).

Haig, Geoffrey & Nils N. Schiborr. In review. Goals last in grammar and discourse.

Haig, Geoffrey & Stefan Schnell. 2016. The discourse basis of ergativity revisited. *Language* 92(3). 591–618. DOI: 10.1353/lan.2016.0049.

Haig, Geoffrey & Stefan Schnell. 2023. *Multi-CAST: Multilingual corpus of annotated spoken texts. (Version 2311).* Bamberg. multicast.aspra.uni-bamberg.de.

Haig, Geoffrey, Stefan Schnell & Nils N. Schiborr. 2021. Universals of reference in discourse and grammar: Evidence from the multi-CAST collection of spoken corpora. In Geoffrey Haig, Stefan Schnell & Frank Seifart (eds.), *Doing corpus-*

based typology with spoken language data: State of the art, 141–177. Honolulu: University of Hawai'i Press.

Haig, Geoffrey, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr. 2022. *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* Bamberg. multicast.aspra.uni-bamberg. de/resources/wowa/.

Haig, Geoffrey & Hannah Thiele. 2014. Post-predicate Goals in Northern Kurdish and neighbouring languages: A pilot study in quantitative areal linguistics languages. Paper presented at the *Second International Conference on Variation and Change in Kurdish, mardin artuklu university, turkey,* 8-9 October 2014.

Harris, Alice C. & Lyle Campbell. 1995. *Historical syntax in cross-linguistic perspective.* Cambridge: Cambridge University Press.

Haspelmath, Martin. 2015. Ditransitive constructions. *Annual Review of Linguistics* 1(1). 19–41. DOI: 10.1146/annurev-linguist-030514-125204.

Hawkins, John A. 2008. An asymmetry between VO and OV languages: The ordering of obliques. In Greville Corbett & Michael Noonan (eds.), *Case and grammatical relations: Studies in honor of Bernard Comrie*, 167–190. Amsterdam: Benjamins. DOI: doi.org/10.1075/tsl.81.08ana.

Hodgson, Katherine, Victoria Khurshudyan & Pollet Samvelian. 2024. Post-predicate arguments in Modern Eastern Armenian. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 375–410. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266355.

Iefremenko, Kateryna. 2021a. Kurdish (Northern, Ankara). In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* (accessed on October 20, 2022). Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa.

Iefremenko, Kateryna. 2021b. Oghuz (Ankara). In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* Bamberg: University of Bamberg. multicast.aspra.uni-bamberg. de/resources/wowa/#turkic.

Izadi, Elham. 2022. Persian (New). In Geoffrey Haig, Donald Stilo, Nils N. Schiborr & Mahîr Dogan (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/ wowa/.

Izadifar, Raheleh. 2022. Tāti (Hazārrudi). In Geoffrey Haig, Donald Stilo, Nils N. Schiborr & Mahîr Dogan (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

Jahani, Carina. 2018. Post-verbal arguments in Balochi. Paper presented at *the International Conference on Anatolia-Caucasus-Iran: Ethnic and linguistic contacts (ACIC),* Yerevan University, 10–12 May.

Jing, Yingqi, Paul Widmer & Balthasar Bickel. 2021. Word order variation is partially constrained by syntactic complexity. *Cognitive Science* 45(11). 1–25. DOI: 10.1111/cogs.13056.

Khan, Geoffrey. 2008a. *The Jewish Neo-Aramaic dialect of Urmi.* Piscataway, NJ: Gorgias Press.

Khan, Geoffrey. 2008b. *The Neo-Aramaic dialect of Barwar.* Leiden: Brill. DOI: 10.1163/ej.9789004167650.i-2198.

Korn, Agnes. 2024. Notes on word order in Bashkardi. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 155–173. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266339.

Korn, Agnes & Ilya Gershevitch. 2023a. Bashkardi (Northern): Recordings by Ilya Gershevitch (1914–2001), transcription and analysis by Agnes Korn. In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

Korn, Agnes & Ilya Gershevitch. 2023b. Bashkardi (Southern): Recordings by Ilya Gershevitch (1914–2001), transcription and analysis by Agnes Korn. In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

Lazard, Gilbert. 1957. *Grammaire du Persan contemporain.* Paris: Klincksieck.

Leitner, Bettina. 2021. Arabic (Khuzestan). In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* Bamberg. http://multicast.aspra.uni-bamberg.de/resources/wowa/ (6 March, 2023).

Levshina, Natalia. 2019. Token-based typology and word order entropy: A study based on universal dependencies. *Linguistic Typology* 23(3). 533–572. DOI: 10. 1515/lingty-2019-0025.

Levshina, Natalia, Savithry Namboodiripad, Marc Allassonnière-Tang, Mathew Kramer, Luigi Talamo, Annemarie Verkerk, Sasha Wilmoth, Gabriela Garrido Rodriguez, Timothy Michael Gupton, Evan Kidd, Zoey Liu, Chiara Naccarato, Rachel Nordlinger, Anastasia Panova & Natalia Stoynova. 2023. Why we need a gradient approach to word order. *Linguistics* 61(4). 825–883. DOI: 10.1515/ling-2021-0098.

MacKenzie, David N. 1962. *Kurdish dialect studies Vol. II.* London/New York: Oxford University Press.

Mettouchi, Amina & Martine Vanhove. 2021. Prosodic segmentation and cross-linguistic comparison in corpafroas and cortypo: corpus-driven and corpus-based approaches. In Geoffrey Haig, Stefan Schnell & Frank Seifart (eds.), *Doing corpus-based typology with spoken language corpora: State of the art (language documentation & conservation special publication 25)*, 59–113. Honolulu, HI: University of Hawai'i Press.

Mithun, Marianne. 1992. Is basic word order universal? Grounding and coherence in discourse. In Doris Payne (ed.), *The pragmatics of word-order flexibility*, 15–61. Amsterdam: Benjamins.

Mohammadirad, Masoud. 2022a. Gorani (Gawraǰū). In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* Bamberg: Bamberg University Press. multicast.aspra.uni-bamberg.de/resources/wowa/.

Mohammadirad, Masoud. 2022b. Kurdish (Central, Sanandaj). In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

Mohammadirad, Masoud. 2022c. Kurdish (Southern, Bijar). In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* Bamberg: Bamberg University Press. multicast.aspra. uni-bamberg.de/resources/wowa/.

Mohammadirad, Masoud. 2024. Zagros region: The Kurdish-Gorani continuum. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian*

Transition Zone: A corpus-based approach to areal typology, 245–279. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266347.

Molin, Dorota. 2022. NE Neo-Aramaic (Jewish Dohok). In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa (5 July, 2023).

Neocleous, Nicolaos. 2020. *Word order and information structure in Romeyka: A syntax and semantics interface account of order in a minimalist system.* Cambridge: University of Cambridge. (Doctoral dissertation).

Nivre, Joakim, Marie-Catherine de Marneffe, Filip Ginter, Jan Hajič, Christopher D Manning, Sampo Pyysalo, Sebastian Schuster, Francis Tyers & Daniel Zeman. 2020. Universal dependencies v2: An evergrowing multilingual treebank collection. In *Proceedings of the 12th language Resources and Evaluation Conference*, 4034–4043. Marseille: European Language Resources Association.

Noorlander, Paul M. 2022a. Arabic (Jewish, Baghdad). In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA – Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

Noorlander, Paul M. 2022b. NE Neo-Aramaic (Jewish Sanandaj). In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA – Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

Noorlander, Paul M. 2024a. Arabic and Aramaic in eastern Anatolia. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 471–514. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266361.

Noorlander, Paul M. 2024b. Neo-Aramaic in Iran and northeastern Iraq. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 431–469. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266359.

Nourzaei, Maryam. 2021a. Balochi (Coastal). In Geoffrey Haig, Donald Stilo, Nils N. Schiborr & Mahîr Dogan (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order vari-*

ation. Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

Nourzaei, Maryam. 2021b. Balochi (Koroshi). In Geoffrey Haig, Donald Stilo, Nils N. Schiborr & Mahîr Dogan (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

Nourzaei, Maryam. 2022. Kholosi (Kholos). In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

Nourzaei, Maryam. 2024. Kholosi. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 175–195. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266341.

Nourzaei, Maryam & Geoffrey Haig. 2024. Balochi: A cross-dialect investigation of post-verbal elements. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 121–154. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266337.

Parizadeh, Mehdi. 2022. Persian (Early new). In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

Parizadeh, Mehdi & Mohammad Rasekh-Mahand. 2024. Post-predicate elements in Early New Persian (10–13th Century CE). In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 231–244. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266345.

Petersen, Jan Heegård. 2015. Kalasha texts with introductory grammar. *Acta Linguistica Hafniensia* 47(S1). 1–275. DOI: {10.1080/03740463.2015.1069049}.

Rasekh-Mahand, Mohammad, Elham Izadi, Mehdi Parizadeh, Geoffrey Haig & Nils Schiborr. 2024. Post-predicate elements in modern colloquial Persian: A multifactorial analysis. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in*

the Western Asian Transition Zone: A corpus-based approach to areal typology*, 197–230. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266343.

Robbeets, Martine. 2017. The Transeurasian languages. In Raymond Hickey (ed.), *The Cambridge handbook of areal linguistics*, 586–626. Cambridge: Cambridge University Press.

Schnell, Stefan, Geoffrey Haig, Nils N. Schiborr & Maria Vollmer. 2023. Are referent introductions sensitive to forward planning in discourse? Evidence from Multi-CAST. In Simone Mattiola & Alessandra Barotto (eds.), *Discourse phenomena in typological perspective*, 231–268. Amsterdam: Benjamins.

Schnell, Stefan, Geoffrey Haig & Frank Seifart. 2021. The role of language documentation in corpus-based typology. In Geoffrey Haig & Stefan Schnell (eds.), *Doing corpus-based 38 typology with spoken language data: State of the art,* 1–28. Honolulu: University of Hawai'i Press.

Schnell, Stefan & Nils N. Schiborr. 2022. Crosslinguistic corpus studies in linguistic typology. *Annual Review of Linguistics* 8(1). 171–191. DOI: 10.1146/annurev-linguistics-031120-104629.

Schreiber, Laurentia. 2021a. Oghuz (Qashqai). In Geoffrey Haig, Donald Stilo, Nils N. Schiborr & Mahîr Dogan (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

Schreiber, Laurentia. 2021b. Pontic Greek (Romeyka). In Geoffrey Haig, Donald Stilo, Nils N. Schiborr & Mahîr Dogan (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

Schreiber, Laurentia, Mortaza Taheri-Ardali, Geoffrey Haig & Erik Anonby. 2021. Contact-induced change in the morphosyntax of Turkic in Boldaji, Chahar Mahal va Bakhtiari province, Iran. *Turkic Languages* 25(2). 210–242. DOI: 10.13173/TL.25.2.210.

Seifart, Frank, Ludger Paschen & Matthew Stave (eds.). 2022. *Language documentation reference corpus (DoReCo).* Berlin/Lyon: Leibniz-Zentrum Allgemeine Sprachwissenschaft & Laboratoire Dynamique Du Langage. DOI: 10.34847/nkl.7cbfq779.

Siewierska, Anna. 1988. *Word order rules.* London: Croom Helm.

Skopeteas, Stavros. 2024. Post-predicate prosody in OV languages. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A*

corpus-based approach to areal typology, 87–117. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266335.

Song, Jae Jung. 2018. *Linguistic typology*. Oxford: Oxford University Press.

Stilo, Donald. 2005. Iranian as a buffer zone between the universal typology of Turkic and Semitic. In Éva 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 Lars Johanson & Christiane Bulut (eds.), *Turkic-Iranian contact areas: Historical and linguistic aspects*, 310–333. Wiesbaden: Harrassowitz.

Stilo, Donald. 2009. Circumpositions as an areal response: The case study of the Iranian zone. *Turkic Languages* 13. 3–33.

Stilo, Donald. 2012. Intersection zones, overlapping isoglosses, and "Fade-out/Fade-in" phenomena in Central Iran. In Mohammad R. Ghanoonparvar & Behrad Aghaei (eds.), *Iranian languages and culture: Essays in honor of Gernot Ludwig Windfuhr*, 134–155. Costa Mesa, CA: Mazda Publisher.

Stilo, Donald. 2018a. Investigating shared features in the Araxes-Iran linguistic area and its subareas. In Christiane Bulut (ed.), *Linguistic minorities in Turkey and Turkic-speaking minorities of the periphery*, 427–452. Wiesbaden: Harrassowitz.

Stilo, Donald. 2018b. Preverbal and postverbal peripheral arguments in the Araxes-Iran linguistic area. Paper presented at the *Conference Anatolia-Caucasus-Iran: Ethnic and Linguistic Contacts* 10-12 May 2018. Yerevan University, Yerevan, Armenia.

Stilo, Donald. 2021a. Oghuz (Tabriz). In Geoffrey Haig, Donald Stilo, Nils N. Schiborr & Mahîr Dogan (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation*. Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

Stilo, Donald. 2021b. Talyshi (Lerik). In Geoffrey Haig, Donald Stilo, Nils N. Schiborr & Mahîr Dogan (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation*. Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

Stilo, Donald. 2022a. Kurdish (Northern, Lachin). In Geoffrey Haig, Donald Stilo, Nils N. Schiborr & Mahîr Dogan (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation*. Bamberg: University of Bamberg. multicast . aspra . uni - bamberg . de / resources/wowa/.

Stilo, Donald. 2022b. NE Neo-Aramaic (Christian) Barwar. In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA – Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation*. Bamberg: Bemberg University. multicast.aspra.uni-bamberg.de/resources/wowa/.

Stilo, Donald & Geoffrey Haig. 2022. Mazandarani (Kordxeyl). In Geoffrey Haig, Donald Stilo, Nils N. Schiborr & Mahîr Dogan (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation*. Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

Stilo, Donald & René Lacroix. 2021. Laz (Arhavi). In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation*. Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/ (27 July, 2023).

Szeto, Pui Yiu & Chingduang Yurayong. 2021. Sinitic as a typological sandwich: Revisiting the notions of altaicization and taicization. *Linguistic Typology* 25(3). 551–599.

Wälchli, Bernhard. 2009. Data reduction typology and the bimodal distribution bias. *Linguistic Typology* 13(1). 77–94. DOI: 10.1515/LITY.2009.004.

Wasow, Thomas. 2022. Factors influencing word ordering. In Hiwa Asadpour & Thomas Jügel (eds.), *Word order variation: Semitic, Turkic and Indo-European languages in contact*, 1–14. Berlin: De Gruyter Mouton.

Yamashita, Hiroko & Franklin Chang. 2001. "Long before short" preference in the production of a head-final language. *Cognition* 81(2). B45–B55. DOI: 10.1016/S0010-0277(01)00121-4.

Yao, Yao. 2018. NP weight effects in word order variation in Mandarin Chinese. *Lingua Sinica* 4(5). 1–31. DOI: 10.1186/s40655-018-0037-8.

Appendix A Data sources and raw figures

Table 5: Data sources: 35 doculects in WOWA (April 2024). Legend for abbreviations: OV = object-verb word order; VO = verb-object word order; NENA = North Eastern Neo-Aramaic. "Unpubl." indicates data-sets which are fully annotated, but due to accessibility issues cannot be published online.

doculect	affiliation	source
Oghuz (Ankara)	🔴 Turkic	Iefremenko 2021b
Oghuz (Bayat)	🔴 Turkic	Unpubl.
Oghuz (Erzurum)	🔴 Turkic	Dogan 2021a
Oghuz (Gagauz)	🔴 Turkic	Dogan 2021b
Oghuz (Qashqai)	🔴 Turkic	Schreiber 2021a
Oghuz (Tabriz)	🔴 Turkic	Stilo 2021a
Balochi (Coastal)	🔷 Iranian, western	Nourzaei 2021a
Balochi (Koroshi)	🔷 Iranian, western	Nourzaei 2021b
Balochi (Turkmen)	🔷 Iranian, western	Haig 2022a
Bashkardi (Northern)	🔷 Iranian, western	Korn & Gershevitch 2023a
Bashkardi (Southern)	🔷 Iranian, western	Korn & Gershevitch 2023b
Gorani (Gawrajū)	🔷 Iranian, western	Mohammadirad 2022a
Kumzari (Musandam)	🔷 Iranian, western	Haig 2022b
Kurdish (Central, Sanandaj)	🔷 Iranian, western	Mohammadirad 2022b
Kurdish (Northern, Ankara)	🔷 Iranian, western	Iefremenko 2021a
Kurdish (Northern, Lachin)	🔷 Iranian, western	Stilo 2022a
Kurdish (Northern, Muš)	🔷 Iranian, western	Haig 2022c
Kurdish (Southern, Bijar)	🔷 Iranian, western	Mohammadirad 2022c
Mazandarani (Kordxeyl)	🔷 Iranian, western	Stilo & Haig 2022
Persian (New)	🔷 Iranian, western	Izadi 2022
Persian (New, Early Classical)	🔷 Iranian, western	Parizadeh 2022
Talyshi (Lerik)	🔷 Iranian, western	Stilo 2021b
Tat (Daɣuščču)	🔷 Iranian, western	Unpubl.
Tati (Hazārrudi)	🔷 Iranian, western	Izadifar 2022
Vafsi (Gurchani)	🔷 Iranian, western	Dogan 2022
Zazakî (Çewlîg)	🔷 Iranian, western	Demir & Doğan 2021a
Zazakî (Siwêreg)	🔷 Iranian, western	Demir & Doğan 2021b
NENA (Christian, Barwar)	🔺 West Semitic	Stilo 2022b
NENA (Jewish, Dohok)	🔺 West Semitic	Molin 2022
NENA (Jewish, Sanandaj)	🔺 West Semitic	Noorlander 2022b
Arabic (Jewish, Baghdad)	🔺 West Semitic	Noorlander 2022a
Arabic (Khuzestan)	🔺 West Semitic	Leitner 2021
Kholosi (Kholos)	✖ Indo-Aryan	Nourzaei 2022
Laz (Arhavi)	⬣ Kartvelian	Stilo & Lacroix 2021
Pontic Greek (Romeyka)	✦ Hellenic	Schreiber 2021b

Table 6: Raw figures for the WOWA data sets, corpus size and mean token weights in words and characters

doculect	texts	words	valid tokens	token weight in words		token weight in characters	
				mean	SD	mean	SD
Oghuz (Ankara)	28	4145	587	1.42	0.68	8.80	5.18
Oghuz (Bayat)	1	3037	835	1.46	0.72	7.62	4.17
Oghuz (Erzurum)	3	3860	636	1.35	0.58	7.63	3.77
Oghuz (Gagauz)	2	5220	594	1.39	0.65	7.63	4.12
Oghuz (Qashqai)	5	2915	557	1.52	0.82	7.84	4.56
Oghuz (Tabriz)	13	3468	851	1.47	0.71	8.37	4.88
Balochi (Coastal)	3	6768	1535	1.42	0.60	8.11	4.79
Balochi (Koroshi)	2	3083	573	1.53	0.73	9.11	4.63
Balochi (Turkmen)	4	4323	580	1.60	0.84	8.25	4.94
Bashkardi (Southern)	5	947	234	1.35	0.63	6.85	3.37
Bashkardi (Northern)	6	2744	596				
Gorani (Gawraju)	7	8782	1015	1.35	0.64	7.21	4.12
Kumzari (Musandam)	2	4496	592	1.25	0.58	5.49	3.04
Kurdish (Central, Sanandaj)	11	8502	1180	1.37	0.62	7.48	3.91
Kurdish (Northern, Ankara)	30	4728	507	1.45	0.60	8.03	4.31
Kurdish (Northern, Lachin)	28	3714	773	1.84	0.76	7.37	4.25
Kurdish (Northern, Mus)	2	2711	693	1.47	0.64	4.84	1.99
Kurdish (Southern, Bijar)	8	7251	1150	1.45	0.72	7.53	4.61
Mazandarani (Kordxeyl)	7	3193	676	1.56	0.70	7.30	4.16
Persian (New)	30	12564	1628	1.65	0.84	9.71	6.02
Persian (New, Early Classical)	3	6751	1278	1.59	0.86	9.40	6.67
Talyshi (Lerik)	3	2872	650	1.76	0.73	7.16	3.80
Tat (Daykuscu)	1	1316	320	1.38	0.54	7.38	3.52
Tati (Hazarrudi)	8	4068	665	1.37	0.68	7.00	4.08
Vafsi (Gurchani)	10	4751	733	1.56	0.76	7.89	3.75
Zazaki (Cewlig)	1	2444	410	1.43	0.66	5.89	3.28
Zazaki (Siwereg)	1	1972	352	1.39	0.59	5.99	3.19
NENA (Christian, Barwar)	5	3517	963	1.38	0.66	7.83	4.68
NENA (Jewish, Dohok)	11	3295	514	1.26	0.54	6.85	3.56
NENA (Jewish, Sanandaj)	4	7166	1184	1.25	0.52	6.74	3.05
Arabic (Jewish, Baghdad)	4	3057	490	1.39	0.68	9.01	5.14
Arabic (Khuzestan)	6	6391	546	1.33	0.65	7.90	3.94
Kholosi (Kholos)	2	3171	516	1.54	0.77	8.72	4.89
Laz (Arhavi)	11	1389	400	1.22	0.51	7.72	4.43
Pontic Greek (Romeyka)	5	2946	501	1.66	0.72	8.37	3.76

Table 7: Raw figures of the WOWA data sets, rates of post-verbal placement of nominal direct objects and goals

doculect	nominal direct objects			nominal goals		
	n(post)	n(all)	%(post)	n(post)	n(all)	%(post)
Oghuz (Ankara)	2	88	2	9	123	7
Oghuz (Bayat)	10	283	4	75	117	64
Oghuz (Erzurum)	16	229	7	46	120	38
Oghuz (Gagauz)	78	154	51	64	88	73
Oghuz (Qashqai)	12	147	8	58	82	71
Oghuz (Tabriz)	2	219	1	88	117	75
Balochi (Coastal)	23	338	7	71	112	63
Balochi (Koroshi)	4	182	2	77	86	90
Balochi (Turkmen)	3	192	2	20	42	48
Bashkardi (Southern)	8	73	11	41	51	80
Bashkardi (Northern)	50	182	27	58	92	63
Gorani (Gawraju)	13	275	5	233	243	96
Kumzari (Musandam)	8	115	7	83	83	100
Kurdish (Central, Sanandaj)	3	295	1	267	283	94
Kurdish (Northern, Ankara)	0	81	0	70	119	59
Kurdish (Northern, Lachin)	3	197	2	90	111	81
Kurdish (Northern, Muš)	6	217	3	107	120	89
Kurdish (Southern, Bijar)	7	298	2	272	281	97
Mazandarani (Kordxeyl)	8	319	3	68	108	63
Persian (New)	19	377	5	218	258	84
Persian (New, Early Classical)	4	257	2	1	21	5
Talyshi (Lerik)	4	164	2	73	102	72
Tat (Daykuscu)	15	100	15	35	45	78
Tati (Hazarrudi)	4	153	3	111	121	92
Vafsi (Gurchani)	4	257	2	146	166	88
Zazaki (Cewlig)	4	85	5	90	99	91
Zazaki (Siwereg)	4	86	5	46	46	100
NENA (Christian, Barwar)	262	315	83	105	109	96
NENA (Jewish, Dohok)	188	210	90	105	106	99
NENA (Jewish, Sanandaj)	18	331	5	171	185	92
Arabic (Jewish, Baghdad)	159	164	97	77	77	100
Arabic (Khuzestan)	267	308	87	77	81	95
Kholosi (Kholos)	2	138	1	34	55	62
Laz (Arhavi)	4	128	3	2	54	4
Pontic Greek (Romeyka)	116	175	66	62	78	7

Table 8: Raw figures of the WOWA data sets, rates of post-verbal placement of pronominal direct objects and goals

	pronominal direct objects			pronominal goals		
doculect	n(post)	n(all)	%(post)	n(post)	n(all)	%(post)
Oghuz (Ankara)	1	14	7	2	19	11
Oghuz (Bayat)	6	70	9	4	8	50
Oghuz (Erzurum)	4	54	7	1	11	9
Oghuz (Gagauz)	26	64	41	7	11	64
Oghuz (Qashqai)	1	28	4	6	11	55
Oghuz (Tabriz)	6	59	10	11	16	69
Balochi (Coastal)	27	99	27	1	2	50
Balochi (Koroshi)	0	18	0	0	1	0
Balochi (Turkmen)	2	55	4	0	5	0
Bashkardi (Southern)	0	9	0	0	1	0
Bashkardi (Northern)	2	23	9	1	5	20
Gorani (Gawraju)	0	32	0	3	4	75
Kumzari (Musandam)	52	81	64	40	40	100
Kurdish (Central, Sanandaj)	0	24	0	11	13	85
Kurdish (Northern, Ankara)	1	11	9	5	9	56
Kurdish (Northern, Lachin)	0	34	0	3	5	60
Kurdish (Northern, Muš)	2	41	5	4	12	33
Kurdish (Southern, Bijar)	0	45	0	2	2	100
Mazandarani (Kordxeyl)	6	62	10	8	13	62
Persian (New)	1	63	2	1	3	33
Persian (New, Early Classical)	0	63	0	0	4	0
Talyshi (Lerik)	2	73	3	8	14	57
Tat (Daykuscu)	7	17	41	4	4	100
Tati (Hazarrudi)	4	60	7	13	17	76
Vafsi (Gurchani)	3	46	7	2	2	100
Zazaki (Cewlig)	4	41	10	11	13	85
Zazaki (Siwereg)	3	30	10	3	7	43
NENA (Christian, Barwar)	15	44	34	14	17	82
NENA (Jewish, Dohok)	21	31	68	6	6	100
NENA (Jewish, Sanandaj)	1	48	2	17	21	81
Arabic (Jewish, Baghdad)	13	18	72	0	2	0
Arabic (Khuzestan)	5	10	50	0	1	0
Kholosi (Kholos)	2	16	12	1	1	100
Laz (Arhavi)	0	35	0	0	5	0

Table 9: Raw figures for the WOWA data sets, rates of post-verbal placement of nominal and pronominal addressees/recipients and various other obliques (locations, sources, instruments, benefactives, comitatives)

doculect	recipients/addressees			other obliques		
	n(post)	n(all)	%(post)	n(post)	n(all)	%(post)
Oghuz (Ankara)	3	16	19	13	199	7
Oghuz (Bayat)	21	49	43	20	152	13
Oghuz (Erzurum)	4	52	8	10	116	9
Oghuz (Gagauz)	7	15	47	43	114	38
Oghuz (Qashqai)	7	8	88	13	81	16
Oghuz (Tabriz)	11	53	21	24	190	13
Balochi (Coastal)	31	112	28	25	138	18
Balochi (Koroshi)	10	24	42	17	64	27
Balochi (Turkmen)	3	21	14	7	110	6
Bashkardi (Southern)	3	8	38	5	14	36
Bashkardi (Northern)	23	41	56	15	49	31
Gorani (Gawraju)	27	33	82	37	102	36
Kumzari (Musandam)	93	96	97	61	70	87
Kurdish (Central, Sanandaj)	21	22	95	44	200	22
Kurdish (Northern, Ankara)	2	15	13	12	209	6
Kurdish (Northern, Lachin)	22	55	40	9	244	4
Kurdish (Northern, Muš)	16	45	36	7	139	5
Kurdish (Southern, Bijar)	11	41	27	17	148	11
Mazandarani (Kordxeyl)	6	50	12	9	119	8
Persian (New)	17	67	25	51	262	19
Persian (New, Early Classical)	0	41	0	16	176	9
Talyshi (Lerik)	23	51	45	37	142	26
Tat (Daykuscu)	12	35	34	17	57	30
Tati (Hazarrudi)	9	17	53	33	212	16
Vafsi (Gurchani)	14	37	38	8	72	11
Zazaki (Cewlig)	4	27	15	0	49	0
Zazaki (Siwereg)	15	29	52	4	49	8
NENA (Christian, Barwar)	13	13	100	148	200	74
NENA (Jewish, Dohok)	41	42	98	50	54	93
NENA (Jewish, Sanandaj)	58	72	81	57	141	40
Arabic (Jewish, Baghdad)	11	11	100	70	82	85
Arabic (Khuzestan)	5	5	100	32	34	94
Kholosi (Kholos)	7	18	39	1	84	1
Laz (Arhavi)	2	36	6	1	87	1
Pontic Greek (Romeyka)	3	5	60	43	114	38

Chapter 2

Word order in the speech of Kurmanji-Turkish bilinguals

Kateryna Iefremenko

Leibniz-Centre General Linguistics & University of Potsdam

The paper investigates word order, particularly the domain of post-predicate position, in Turkish and Kurmanji as two languages located in the Western Asian Transition Zone that are in an intense and long-term contact with each other. Both languages are OV; however, each of them allows placement of constituents in post-predicate position. The results of the analysis show that there is an effect of flagging and semantic role in Kurmanji, which is in line with previous research on word order in Kurmanji, and an effect of weight in Turkish on the employment of post-predicate elements. At the same time, the qualitative analysis showed that there are instances of noncanonical placement of case-flagged goals in Kurmanji that occur in one particular construction *erdê ketin* 'to fall on the ground.'

1 Introduction

In this paper, I investigate a possible structural convergence in the word order, namely the post-predicate domain, of Turkish and Kurmanji as two languages located in the Western Asian Transition Zone (for a definition, see Haig et al. 2024 [this volume]). Both languages are OV languages, but each of them employs the post-predicate position to some extent. In Turkish, word order is determined by information structural requirements, and thus the post-predicate position is reserved for background information. In Kurmanji, the post-predicate position is driven by verb semantics, i.e., Goal arguments are placed in the post-predicate position.

1.1 Factors of language change

In situations of language contact, it is typical for languages to influence each other. According to Thomason (2001), linguistic factors of contact-induced language change determine what will change once social factors have created the situation where something will change. However, linguistic factors can be overridden by social factors, and therefore Thomason & Kaufman (1988) define the intensity of contact as the most important social factor for the prediction of a contact-induced language change. Thomason (2001) provides a four-stage borrowing scale depending on the intensity of contact. In casual contact, when borrowers are not fluent in the source language, only content words are borrowed. In a slightly more intense contact setting when borrowers are fluent bilinguals but form a minority in the community, function words as well as content words can be borrowed at the lexical level, and some minor structural borrowing is possible. In a moderately intense contact setting when there are more borrowers than in the previous stage and social factors favor borrowing, more change can be expected on the structural level; word order features, clause-combining strategies, or inflectional affixes might be borrowed. And finally, in an intense contact setting when there is extensive bilingualism among the speakers and social factors strongly favor borrowing, anything can be borrowed, resulting even in major typological change of the borrowing language.

The intensity of contact is determined by the duration of the contact and the number of speakers in the community. The contact between Turkish and Kurmanji has begun in the Ottoman Empire and intensified with the establishment of the Turkish Republic where Turkish became the country's sole official language (Yağmur 2001). Nowadays, Turkish is dominant and even the only language in official spheres as education, government, business, while Kurmanji is used mostly in the families. As for the number of speakers in the community, the size of the Kurmanji-speaking communities varies depending on the region: in the eastern and south-eastern regions of Turkey, Kurmanji is spoken by the majority of the population, while in the western regions, Turkish prevails even in informal contexts. Thus, on the borrowability scale, the contact between Kurmanji and Turkish in Turkey can be referred to as stage 3 (a setting of moderately intense contact) where structural changes, including word order alternations, can be expected.

Another social factor, which is not provided in Thomason & Kaufman (1988) as one of the most important factors for predicting the outcome of the contact, but which I find relevant for my particular research is the societal status of a language: minority and majority language. It is believed that typically a more prestigious donor language influences a less prestigious recipient language (Johanson

2002). As has been noted above, Turkish in Turkey is the dominant language in such spheres of life as education, business, media, and in the western regions of Turkey it prevails also in the informal settings. Thus, Turkish is clearly the majority language of the society. As for Kurmanji in Turkey, its societal status heavily depends on the region: while in the eastern and south-eastern regions Kurmanji is the dominant language of the majority of speakers (though only in informal settings), in the western regions, it is the minority language. For my study, the data from Kurmanji-Turkish bilinguals in Turkey were collected in Ankara, the capital of Turkey, where Turkish is the dominant language, and Kurmanji is a minority language in this context.

1.2 Word order in language contact situations

A number of studies (Thomason & Kaufman 1988; Thomason 2001; Heine 2008) pointed that word order is prone to change in language contact scenarios. Following this, there are studies that exemplify this claim for Turkic languages. For example, the word order in Karaim changed from OV to VO due to the contact with Slavic and Baltic languages (Csató 2002). Similar to Karaim, Gagauz, which has stayed in a long term contact with Slavic languages, underwent change that resulted in a VO order becoming dominant (Menz 1999). Another study by Keskin (2023) investigated word order across numerous Turkic varieties in the Balkans where Turkish has been in contact mostly with Indo-European languages for centuries. Analysis of the post-predicate domain in these Turkic varieties showed that the further the Turkic variety is located from the borders of Turkey, the higher is the frequency of the VX order in this variety. On the other hand, studies that investigated comparatively recent contact of Turkish with Indo-European languages, i.e., Turkish as a heritage language in the Netherlands, Germany and the U.S., did not find a shift from OV to VO order in these varieties (Doğruöz & Backus 2007; Schroeder et al. In press). Finally, as for research on word order in Turkish in contact with another language in a context where Turkish is the majority language of the society, there are no studies, at least to my knowledge, apart from those done by our research group based on the same data set as the one presented in this paper (Iefremenko et al. 2023; Iefremenko Submitted).

With respect to word order in Kurmanji in Turkey, there are studies that investigate the post-predicate position across different dialects of Kurmanji (Haig 2015; Haig 2019; Gündoğdu 2019), but the focus is more on word order typology and its diachronic change due to contact with other languages. Besides, Asadpour (2022) investigated the placement of goals in Mukri Kurdish spoken in Iran and in the contact languages (such as Armenian, Azeri Turkic and Northeastern

Neo-Aramaic) and found that in fact the combination of several factors such as information structure, semantics and morphosyntax explains the placement of constituents in the post-predicate position in the analyzed languages.

Thus, this study will try to fill a gap first by investigating possible changes in the word order in Turkish in contact with another language in a context where Turkish is the majority language of the society; second, by investigating word order, namely the post-predicate position, in Kurmanji, that is in intense contact with Turkish and is the minority language of the society. The analysis is based on the data that come from 30 Kurmanji-Turkish adult bilingual speakers residing in the Turkish-dominant region in Turkey (Iefremenko 2021a,b).

2 Word order in Kurmanji

Kurdish is a macro-language that consists of a continuum of closely related languages spoken by Kurds over a large geographic area spanned across several countries, such as Turkey, Iran, Iraq, Syria, among others (Sheyholislami 2015). Northern Kurmanji is one of the Kurdish languages, predominantly spoken in southeast Turkey, northwest and northeast Iran, northern Iraq, and northern Syria. It is classified as a member of the northwest Iranian branch of the west Iranian languages, within the Iranian branch of Indo-European language family (Haig & Matras 2002).

Like other West Iranian languages, Kurmanji is an OV language, although it is not always verb-final (Haig 2015). Kurmanji indeed systematically places certain elements after the verb. According to Haig & Thiele (2014), the post-predicate position is reserved for "goals", where it is a cover term for:

a) locational goals of verbs of motion (e.g., *go, run, fall*) and caused motion (e.g., *put, place, take*)

(1) Northern Kurdish Yavuzeli (Matras et al. 2016: K002)
     ez   çû-m-e   mal-ê
     1SG go.PST-1SG-DRCT house-OBL.F
     'I went home.'

(2) Karakoçan (Matras et al. 2016: K075)
     jinik   qutîk-ek anî   mal-ê
     woman box-INDEF bring.PST.3SG house-OBL.F
     'The woman brought a box into the house.'

b) recipients of verbs of transfer (e.g., *give*)

(3) Northern Kurdish Elbistan (Matras et al. 2016: K022)
we ew ne-dê   mi
2PL this NEG-give.PST.3SG 1SG.OBL
'You didn't give it to me.'

c) addressees of verbs of speech (e.g., *say, speak, promise*)

(4) Northern Kurdish Siirt (Matras et al. 2016: K008)
min   got-e   wî
1SG.OBL say.PST.3SG-DRCT 3SG.OBL
'I said it to him.'

In other words, the word order of Kurmanji is not a pure OV, but rather OVX. Goals are systematically yet not consistently placed in the post-predicate position in Kurmanji. Haig (2022) states that the position of post-predicate elements in Kurmanji is syntactically fixed and is not the result of pragmatically driven scrambling or stylistic variation, i.e., factors that account for example for the post-predicate position in Turkish. Factors that influence the position of goals are flagging and regional variation (Haig 2015; Gündoğdu 2019). On the other hand, a recent study by Asadpour (2022) on Mukri Kurdish spoken in Iran showed that information structure also plays a role in the placement of Goal arguments. Namely, accessible inferable information occurs in the post-predicate position, while topicalized goals are placed in the preverbal position. Nonetheless, considering that Kurdish varieties are spoken across several countries and as a result come into contact with different languages, the variations observed in the studies could potentially stem from distinctions between these varieties.

I will now elaborate on the relation between the word order in Kurmanji and the type of flagging of Goal arguments. Goals in Kurmanji can be flagged by case (as demonstrated in 5) or by an adposition (see examples 7–9).

(5) Northern Kurdish Tatvan (Matras et al. 2016: K024)
min   xarin anî   od-ê
1SG.OBL food bring.PST.3SG room-OBL.F
'I brought the food to the room.'

Case-flagged Goal arguments are always placed right after the predicate and cannot be separated by an adverb or any other argument. Importantly, the Goal

argument cannot be placed immediately before the verb. As Gündoğdu (2019: 110) emphasizes, "in a Kurmanji clause, at most two case-flagged NPs (subject and direct object) are licensed in the preverbal position". Thus, example 6 below where the argument *odê* (to the room) is placed in the immediate pre-predicate position would be considered noncanonical.

(6)   Northern Kurdish Tatvan (constructed example)
      min   xarin od-ê   anî
      1SG.OBL food room-OBL bring.PST.3SG
      'I brought the food to the room.'

The second means of marking goals is by the help of an adposition. In Kurmanji, there are several types of adpositions, namely basic prepositions, locational nouns which can be used together with a preposition (see example 7), postpositions, and circumpositions.

(7)   Northern Kurdish Pertek (Matras et al. 2016: K028)
      lawik-ê   qiçik di-bez-e   ber_bi   dî-ya   xwe
      boy-EZ.M little PRS-run.3SG-DRCT towards mother-EZ.F own
      'The little boy is running to his mother.'

In general, the position of Goal arguments flagged by adpositions is more flexible compared to those flagged by case. The only exception are locational nouns that are not preceded by a preposition: they are always placed in post-predicate position. The reason for such position is that this type of adpositions historically evolved from nouns. Hence, similar to case-flagged goals, locational nouns are placed in the post-predicate position (Haig 2015; Haig & Thiele 2014). As for the position of the other types of adpositions, it is largely dependent on dialect. As it is noted in Haig (2015) and Haig & Thiele (2014), initially the OVX word order of Kurmanji emerged due to the contact with early Aramaic/Iranian languages. As a result, there are different preferences in placing goals across the modern dialects of Kurmanji: namely, goals are predominantly post-predicative in the southernmost dialects where the language stayed in a long-lasting contact with Neo-Aramaic, while in the northern and western dialects goals are overwhelmingly pre-predicative due to their extensive contact with Armenian and Turkish. Besides, in the south, there is a tendency for post-predicative goals to be accompanied by a preposition or a circumposition. At the same time, in the dialects of Central Anatolia, the combination of a post-predicative Goal accompanied by a preposition is very restricted.

3 Word order in Turkish

Turkish is considered to have a relatively free word order, with basic word order being (S)OV, which means that even though word order variation is possible, in some instances word order must stay fixed. In Turkish, variation in word order serves pragmatic purposes such as signaling topics and distinguishing between old and new information (Erguvanlı 1984). Hence, word order in Turkish is strongly motivated by information structure: a link to the previous context or topicalized information appears sentence-initially, new information occurs immediately before the verb, and backgrounded information can be placed post-predicatively (Erguvanlı 1984; Kornfilt 1997). Thus, (8a) has pragmatically neutral unmarked order, whereas (8b–8f) are pragmatically marked.

(8)   a.   Turkish (self-constructed examples)
          SOV
          Murat araba-yı sat-tı.
          Murat car-ACC   sell-PST.3SG
          'Murat sold the car.'

      b.   OSV
          Araba-yı Murat sat-tı.
          car-ACC   Murat sell-PST.3SG
          'It is Murat who sold the car.'

      c.   SVO
          Murat sat-tı   araba-yı.
          Murat sell-PST.3SG car-ACC
          'Murat sold the car.' (emphasis on the verb)

      d.   OVS
          Araba-yı sat-tı   Murat.
          car-ACC   sell-PST.3SG Murat
          'It is the car that Murat sold.'

      e.   VSO
          Sat-tı   Murat araba-yı.
          sell-PST.3SG Murat car-ACC
          'It (the car) was sold by Murat.'

      f.   VOS
          Sat-tı   araba-yı Murat.
          sell-PST.3SG car-ACC Murat
          'The car was sold (by Murat).'

Furthermore, it is important to note that there are differences in the employment of post-predicate structures in spoken and written modes as well as formal and informal language. Particularly, there seem to be restrictions on the use of post-predicate structures in formal written situations. For example, in legal documents as well as news items (both written in newspapers and journals as well as read on radio or TV), post-predicate structures are extremely rare (Erguvanlı 1984: 67). Exceptions are columns in papers when authors deliberately choose more informal style, or literature pieces where an author uses their own style and may employ backgrounding techniques for specific pragmatic reasons (Erguvanlı 1984: 67). On the other hand, elements placed in post-predicate position frequently occur in informal spoken mode, which is usually characterized as spontaneous and unplanned (Schroeder 1995). Thus, spoken language is full of (self-)corrections and afterthoughts, while the same process in written language can be employed with the help of editing and corrections. Another reason of a higher number of post-predicate structures in spoken language compared to written one is that utterances are limited in size due to speaker's awareness of the listener's capacity limitations (Chafe 1985). Thus, as Schroeder (1995: 206) emphasizes, the employment of the post-predicate position helps the listener to keep track of the topical development and the deictic framework in which the predication holds.

4 Research questions

To summarize the preceding discussion, both languages in contact that are under research here have OV word order, but both employ the post-predicate position in a different way. While Kurmanji systematically places Goal arguments — particularly those flagged with case — in the post-predicate position, word order in Turkish is determined by information structural requirements, and the post-predicate position is reserved for background information, regardless of the semantic role of the elements.

As discussed in Section 1, the contact between the two languages has lasted for centuries and is fairly intense, and according to Thomason & Kaufman's (1988) borrowability scale, we can expect structural changes, including the ones concerning word order. In the paper, I will investigate potential convergences in word order, in particular in the post-predicate domain, in Turkish and Kurmanji in Turkey, basing the analysis on the variables encoded in WOWA (see Haig et al. 2024 [this volume]). For the reason that WOWA categories do not incorporate information structure as one of the variables, I do not explore its effect on the word order changes in the investigated languages. This is a limitation of the current paper.

5 Methodology

5.1 Participants

The data for this study come from 30 Kurmanji-Turkish bilinguals (9 females and 21 males). The participants were exposed to Kurmanji from birth or an early age in their family and started acquiring Turkish mainly when they entered school (though some participants were already exposed to Turkish in their families). The place of birth of the participants varies: the majority of the speakers were born and raised in the east and the south-east of Turkey, but some were born and grew up in the western cities of Turkey. At the time of data collection, all participants were living in Ankara, where the dominant language of the society is Turkish. In an urban city, such as Ankara, speakers of different dialects interact between each other on a daily basis; hence, such contact may lead to dialect levelling. Therefore, my data were not controlled for the dialect of Kurmanji. The age of the participants varies between 23 and 37 years, with the mean age being 28.1; thus, all the participants are young adults. All 30 participants had a high school degree, and most of them completed bachelor's or master's degree. So, all the speakers are highly proficient in Turkish since it is the language of education in Turkey. As for the education in Kurmanji, 14 participants stated that they had taken at least a several-month language course in Kurmanji, one participant had earned a master's degree from a Kurmanji language and literature department, and several participants indicated that they actively used Kurmanji for business purposes. However, the results of the section on the self-rated proficiency in the questionnaire showed that on average the speakers rate their proficiency in Kurmanji lower ($\bar{x}$ = 17.53 out of 20 possible) compared to their proficiency in Turkish ($\bar{x}$ = 19.32 out of 20 possible).

5.2 Data collection

The data used for this study were collected with the help of the "Language Situations" method (Wiese 2020). This method combines controlled elicitation with spontaneous data, and thus is suitable for systematic comparisons across contact-linguistic constellations as well as different languages. It captures quasi-naturalistic productions across different communicative situations, including informal versus formal, and written versus spoken communicative situations. The elicitation comprised two sessions (one in Kurmanji and one in Turkish), with at least three days between the sessions. Participants were shown a short video of a car accident and were asked to imagine themselves having witnessed it. After that they were asked to recount the incident in four different imagined situations:

to a friend via a WhatsApp voice message (informal spoken), to a friend via a WhatsApp text message (informal written), to the police via a voice mail (formal spoken), to the police in a form of a written witness report (formal written). To exclude a possible effect of priming, the order of the communicative situations and the order of the languages were balanced.

At the end of the second session, each participant was asked to fill out an extensive questionnaire. The questionnaire comprised nine sections: participants' general information, educational and professional background of the participants, family information, information about the languages, self-assessment of their language skills (in Kurmanji, Turkish, as well as in foreign languages, on a five-point scale), participants' language use with family members and peers, a section concerning media use and free time (texting WhatsApp messages or writing emails in Kurmanji, Turkish, as well as in foreign languages, three scores of frequency), questions concerning personal character traits, and feedback on the participation in the study. The questionnaire was available in Turkish as it is the language for official documents in Turkey. It was always filled out by the participants themselves in the presence of an elicitor.

5.3 Annotation and statistical analysis

For the current study on the post-predicate elements in Iranian and neighboring languages, only the informal spoken productions of the speakers were analyzed.[1] The Kurmanji data comprised 745 communication units[2] , out of which 507 were analyzed because they contained pre- or post-predicate arguments (Iefremenko 2021a). The Turkish data consisted of 799 communication units, out of which 587 were analyzed (Iefremenko 2021b). Data coding was done in line with the WOWA coding method (Haig et al. 2024 [this volume]) and annotated for animacy, weight, semantic role, flagging, and position. What is important for the current analysis is that afterthoughts and self-repairings, which usually occur in the post-predicate position in OV languages, were not excluded from the analysis. Besides, there are several aspects of coding that are important for the analysis of Turkish and Kurmanji data. First, adpositions are not counted as part of the weight because they are considered to be part of the flagging. Thus, for instance, in the example (9) the token *rê da* (on the street) is considered to have weight 1 because the postposition *da* is not counted.

---

[1] (In-)formality and mode as factors that might influence the word order in Turkish and Kurmanji are analyzed and discussed in another study by Iefremenko (Submitted).

[2] A communication unit is an independent clause with its modifiers and dependent (subordinate) clauses.

(9) Northern Kurdish Ankara (Iefremenko 2021a: A, 3)
   iii   ez rê   da di-çû-m
   hmm 1SG street on PROG-go.PST-1SG
   'I was going on the street.'

Second, subjects were excluded from the analysis across all the analyzed languages in WOWA. This has to be taken into account when interpreting the results for Turkish because as indicated in Section 3, unlike Kurmanji, the word order in Turkish is determined by information structure and as a result, subjects can also be placed in the post-predicate position as long as they are not new information.

The statistical analyses were run in R (R Core Team 2023). In addition to the base package of R, I used tidyverse for data manipulation and visualization (Wickham et al. 2019). I ran binomial generalized linear regression models using the lme4 package (Bates et al. 2015).

6 Results

6.1 Kurmanji

First, I will start with the analysis of post-predicate constituents in Kurmanji.

Figure 1 shows the percentage of tokens in the pre- and post-predicate position in the Kurmanji data. The line inside each box indicates the median, while a dot represents the percentage of tokens used in pre- or post-predicate position by one speaker, calculated in relation to the total number of communication units for each speaker. Figure 1 demonstrates that the majority of the communication units have pre-predicate arguments, with around 20% taking post-predicate arguments. Also, we can see that individual variability in the group is high: some participants place tokens post-predicatively in up to 70% of their utterances, while in some speakers all utterances are pre-predicate.

As stated in Section 2, word order in Kurmanji is determined by verb semantics where goals are placed in post-predicate position. Therefore, high individual variability in the group might be caused by some speakers' frequent use of goals. In order to establish whether there is an effect of one of the variables from the WOWA coding scheme, namely animacy, weight, role or flagging (as described in Section 5.3), I ran four binomial generalized linear regression models with the dependent variable Position (preverbal, coded as 0 vs. postverbal, coded as 1) and the independent variables Animacy, Weight, Flagging, and Role.
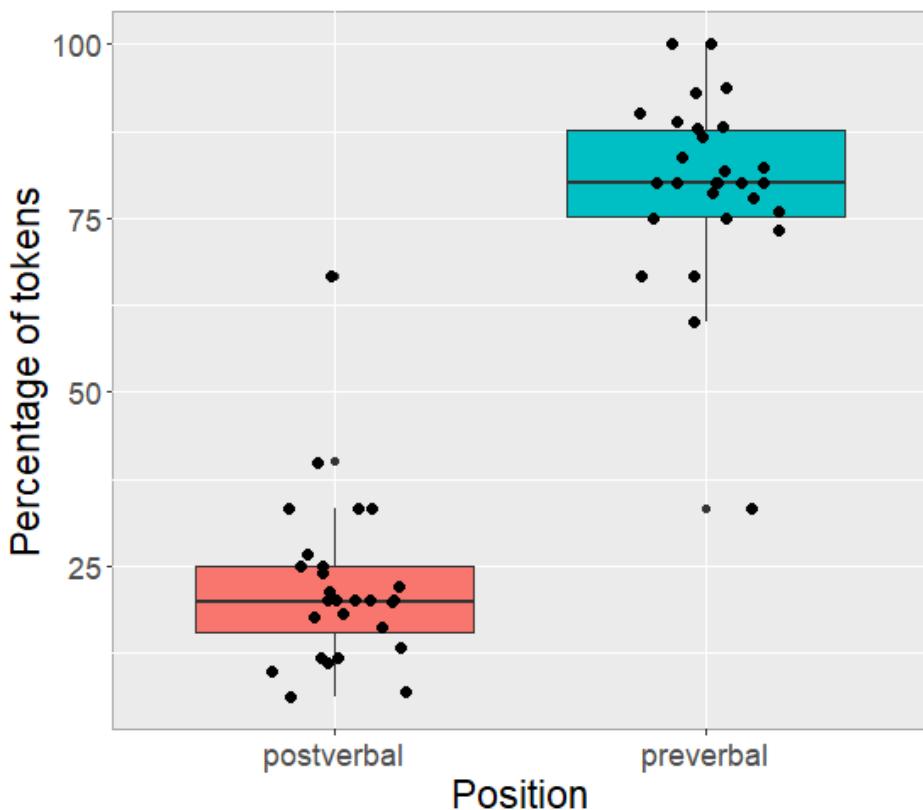
Figure 1: Percentage of tokens in pre- or post-predicate position in Kurmanji

6.1.1 Animacy

Animacy is a categorical independent variable with seven levels in the Kurmanji data. The results of the regression given in Table 1 show that there is no correlation between animacy and the position of the tokens in the Kurmanji data.

6.1.2 Weight

Weight is a discrete numeric variable, based on number of phonological words. Four values are distinguished: one to three phonological words, and more than three (which is coded as "four"). The results of the regression provided in Table 2 show that there is no correlation between the weight of the constituents and the position of the tokens in relation to the verb.

Table 1: Regression table for binomial GLM with the dependent variable Position and the independent variable Animacy in Kurmanji.

fixed effect	β	σ	p-value	
(intercept)	-1.65e+01	5.656e+02	0.97	(ns)
Anim-adv	1.28e+01	5.656e+02	0.98	(ns)
Anim-anim	1.36e+01	5.656e+02	0.98	(ns)
Anim-bp	1.35e+01	5.656e+02	0.98	(ns)
Anim-hum	1.45e+01	5.656e+02	0.97	(ns)
Anim-inan	1.56e+01	5.656e+02	0.97	(ns)
Anim-other	-9.53e-09	2.465e+03	1.00	(ns)

Table 2: Regression table for binomial GLM with the dependent variable Position and the independent variable Weight in Kurmanji

fixed effect	β	σ	p-value	
(intercept)	-1.26	0.30	< .0001	
Weight	-0.15	0.19	0.43	(ns)

6.1.3 Flagging

Flagging is a categorical independent variable with eight levels in the Kurmanji data. The results of the regression analysis presented in Table 3 show that there is an effect of Flagging on the placement of arguments in the pre- or post-predicate position in Kurmanji.

The model showed that arguments flagged with case (see example 10), locational nouns (see example 11), or a preposition together with a locational noun (see example 12) are more likely to be placed in the post-predicate position than in the pre-predicate one.

(10)   Northern Kurdish Ankara (Iefremenko 2021a: Y, 629)
    kûçik xe   *avêt-e*   *top-e*
    dog   oneself throw.PST.3SG-DRCT ball-OBL.F
    'The dog threw itself towards the ball.'

Table 3: Regression table for binomial GLM with the dependent variable Position and the independent variable Flagging in Kurmanji

fixed effect	β	σ	p-value	
(intercept)	-2.67	0.36	< .0001	
Flag-case	2.23	0.43	< .0001	***
Flag-circ	-0.09	0.69	0.88	(ns)
Flag-lvc-poss	-11.89	882.74	0.98	(ns)
Flag-postp	0.84	0.57	0.11	(ns)
Flag-prep	-0.06	0.51	0.89	(ns)
Flag-prep-relnoun	1.91	0.58	0.001	**
Flag-relnoun	3.84	0.52	< .0001	***

(11) Northern Kurdish Ankara (Iefremenko 2021a: Y, 626)
      çû   iii ser rê
      go.PST.3SG hm to street
      'He went to the street.'

(12) Northern Kurdish Ankara (Iefremenko 2021a: C, 57)
      iii di rê   da iii jin-ekî   kûçik-ê wî ew ê derbas
      hm in street CIRC hm woman-INDEF dog-EZ.M her they FUT passing
      bû-n-a   li_hemberê rê
      become.PST-3PL-DRCT across   street
      'Hm, on the street, (there was) a woman (and) her dog, they were about
      to cross the street to the opposite side.'

6.1.4 Role

Role is a categorical independent variable with 15 levels in the Kurmanji data. The results of the regression provided in Table 4 show that there is an effect of role on the placement of arguments in the post-predicate position.

Figure 2 shows the percentage of tokens with a particular role in the pre- and post-predicate position in the Kurmanji data, which was calculated in relation to the total number of tokens in the data. The line inside each box indicates the median, while dots represent the percentage of tokens used in pre- or post-predicate position by one speaker.

The model as well as Figure 2 show that goals of motion (see example 13) and caused motion (see example 14) are more likely to be placed in the post-predicate position than in the pre-predicate one.

Table 4: Regression table for binomial GLM with the dependent variable Position and the independent variable Role in Kurmanji

fixed effect	β	σ	p-value	
(intercept)	-2.85	0.59	< .0001	
Role-addr	0.28	1.19	0.8	(ns)
Role-ben	0.9	1.22	0.45	(ns)
Role-com	1.06	0.96	0.27	(ns)
Role-cop	-14.71	907.61	0.98	(ns)
Role-cop-loc	-14.71	3956.1	0.99	(ns)
Role-do	-1.61	1.16	0.16	(ns)
Role-do-def	-14.71	1978.09	0.99	(ns)
Role-goal	3.65	0.63	< .0001	***
Role-goal-c	2.39	0.66	< .0001	***
Role-instr	-14.71	1318.72	0.99	(ns)
Role-loc	-0.15	0.72	0.83	(ns)
Role-other	-0.51	1.17	0.66	(ns)
Role-rec	20.41	3956.10	0.99	(ns)
Role-stim	2.85	1.53	0.06	(ns)

(13) Northern Kurdish Ankara (Iefremenko 2021a: C, 69)
    çû   *ber_bi*   *ereba*
    go.PST.3SG towards car
    'He went towards the car.'

(14) Northern Kurdish Ankara (Iefremenko 2021a: E, 108)
    û   *kûçik jî*   *xe*   *awêt-e*   *top-ê*
    and dog   also oneself throw.PST.3SG-DRCT ball-OBL.F
    'And the dog threw itself to the ball.'

Thus, the analysis demonstrated that in Kurmanji goals of motion and caused motion, which are flagged with case, locational noun, or sometimes a preposition and a locational noun, are likely to be placed in post-predicate position. This is in fact what is known from previous literature on the post-predicate position in Kurmanji (Haig 2015; Haig 2019; Gündoğdu 2019).

Regarding the other types of goals, as shown in Figure 2, there are no examples of addressees placed in post-predicate position in my data, and all examples containing an addressee argument are placed in pre-predicate position. The reason
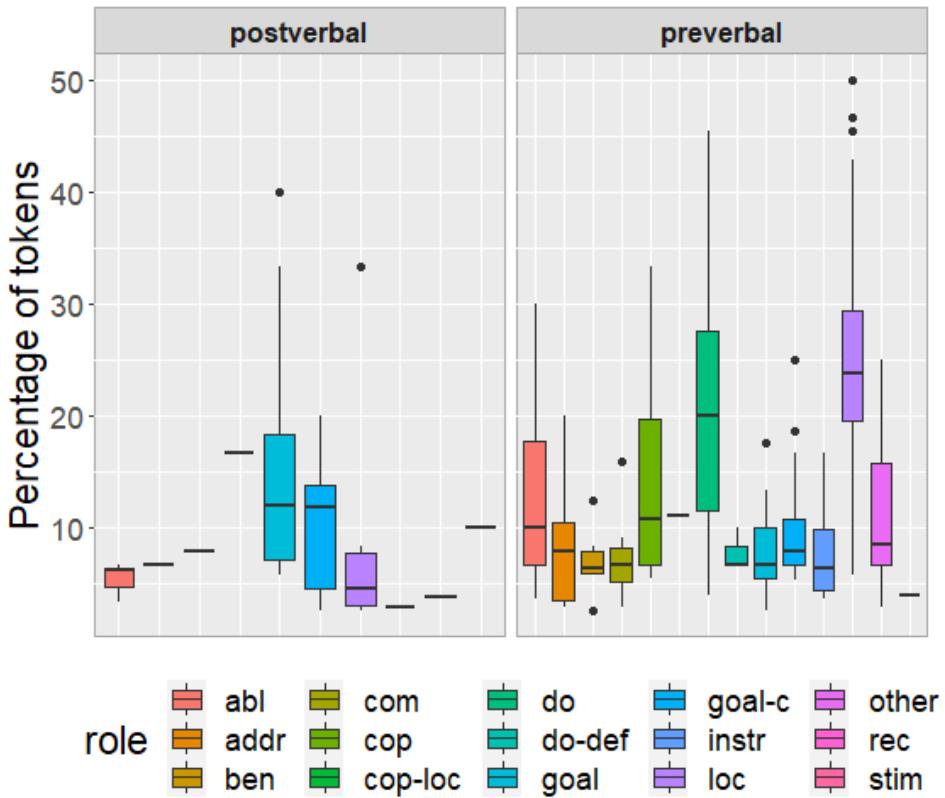
Figure 2: Percentage of tokens with a particular role in the pre- and post-predicate position in Kurmanji.

for this might be flagging: all pre-predicate examples are flagged with a circum-position (example 15) or a postposition (example 16).

(15) Northern Kurdish Ankara (Iefremenko 2021a: F, 134)
      mi   go   e ji te   ra bi-bêj-im
      1SG.OBL say.PST hm to 2SG.OBL CIRC SUBJ-say.PRS-1SG
      'I said, I am telling (this) to you.'

(16) Northern Kurdish Ankara (Iefremenko 2021a: F, 114)
      mi   go   ez te   re   bêj-im
      1SG.OBL say.PST 1SG 2SG.OBL POSTP SUBJ.say.PRS-1SG
      'I said, I am telling (this) to you.'

With regard to recipients, there are only two examples in the Kurmanji data, and they are placed post-predicatively (example 17). Presumably, the reason for a small number of recipients in the data is the method: the video shown to the participants did not trigger the use of verbs of transfer.

(17) Northern Kurdish Ankara (Iefremenko 2021a: T, 498)
xeber-ê   bi-di-m   te
news-OBL SUBJ-give.PRS-1SG 2SG.OBL
'Let me tell (lit. give) you the news.'

Apart from goals of (caused) motion, addressees, and recipients, Figure 2 shows that locations (abbreviated in Figure 2 as "loc" – see example 18) and sources of motion (abbreviated in Figure 2 as "abl" – see example 19) can also be placed in post-predicate position in Kurmanji.

(18) Northern Kurdish Ankara (Iefremenko 2021a: F, 126)
du erebe piştî hevdu   di-çû   him iii li_ser rê   da
two car   after each.other PROG-go.PST.3SG both hmm on   street CIRC
'Two cars were coming one after another on the street.'

(19) Northern Kurdish Ankara (Iefremenko 2021a: A, 28)
e   du erebe di-hat-in   ji   wî alî
hm two car   PROG-come.PST-3PL from his side
'Hm, two cars were coming from his side.'

Another interesting observation is placement of case-flagged goals of motion. As the previous literature (Haig 2015; Haig 2019; Gündoğdu 2019) shows, goals of motion flagged with case are always placed in immediate post-predicate position. Yet there are examples in my data where case-flagged goals of motion are placed in pre-predicate position (see example 20 and 21).

(20) Northern Kurdish Ankara (Iefremenko 2021a: N, 373)
nan û av erd-î   ket
bread and water ground-OBL.M fall.PST.3SG
'Bread and water fell on the ground.'

(21) Northern Kurdish Ankara (Iefremenko 2021a: O, 386)
pişte zilam iii gok-ê xwe imm (–) ji   dest-ê   xwa imm (–)
after man   hmm ball-EZ.F own hmm   from hand-EZ.M own hmm
erdêk/_erd-ê   ket
ground-OBL.M fall.PST.3SG
'Afterwards, the man, hm, his ball, hm, fell from his hands on the ground.'

In total in the data there are four examples with a case-flagged Goal of motion placed in pre-predicate position, and all four examples are actually the same construction[3] *erdê ketin* 'to fall on the ground'. In the contact language Turkish, *erdê ketin* is rendered as *yere düşmek*. In Turkish, *yere düşmek* is a fixed construction. This is also supported by the Turkish data from Kurmanji-Turkish bilinguals in Ankara: there are 10 instances of *yere düşmek*, and in all of them the argument *yere* 'on the ground' is placed in the immediate pre-predicate position. Besides, I have also checked other data collected with the help of the same method, namely Turkish monolingual speakers from Izmir and Eskişehir (Wiese et al. 2021) as well as Kurmanji-Turkish heritage speakers in Berlin, and found that in all the instances of *yere düşmek*, the argument *yere* was placed in the immediate pre-predicate position. Thus, I argue that the stable position of the argument *yere* in Turkish leads to transfer of the whole construction into Kurmanji and as a result, the argument *erdê* is placed pre-predicatively. Besides, in Kurmanji there are other constructions with the verb *ketin*, which are set phrases and where the Goal is always placed in pre-predicate position, e.g., *bi rê ketin* 'to set off (on a journey)'. This might explain the fact why I do not find examples of transfer of constructions with other verbs.

Thus, the analysis of the post-predicate position in Kurmanji of Kurmanji-Turkish bilingual speakers in Ankara showed that Kurmanji is an OV language where goals of motion and caused motion flagged with case, a locational noun, or a preposition used together with a locational noun, are systematically placed in post-predicate position.[4] However, the analysis showed that other elements, such as location and source of motion, can also be placed in post-predicate position in Kurmanji. Apart from this, the data demonstrated that some speakers employ the pre-predicate position for case-flagged goals of motion, which is a non-canonical position for such arguments.

6.2 Turkish

In this section, I will present the results of the analysis of the majority language of the society the speakers live in – which is Turkish.

Figure 3 shows frequency distribution of the coded tokens used in pre- or post-predicate position. The percentage was calculated in relation to the total number of utterances that contain tokens placed pre- or post-predicatively. Each dot represents the mean use of tokens by a speaker. In Figure 3, we can see that in

---

[3]For a definition of a construction see Croft & Cruse (2004: 257–262).

[4]As for addressees and recipients, there is not enough data in the data set to make conclusions about their position in a sentence in relation to the verb.

general the pre-predicate position of arguments is preferred: only 10% of utterances contain post-predicate elements. However, the figure also shows that there is individual variability: some speakers use up to 25% of post-predicate structures in their narrations.
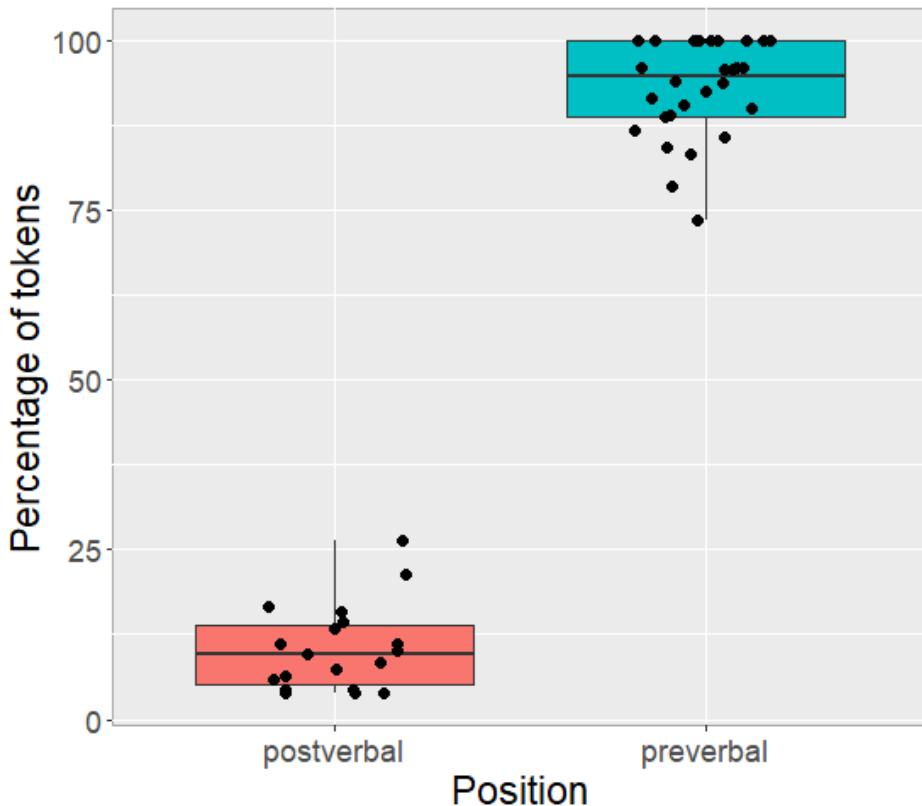
Figure 3: Percentage of tokens in pre- or post-predicate position in Turkish.

Similar to Kurmanji, in Turkish, each token was annotated for animacy, weight, role and flagging (see Haig et al. 2024 [this volume]). To find out whether there is an effect of one of these variables, I ran four binomial generalized linear regression models with the dependent variable Position (preverbal, coded as 0 vs. postverbal, coded as 1) and the independent variables Animacy, Weight, Flagging, and Role.

6.2.1 Animacy

Animacy is a categorical independent variable with six levels in the Turkish data. The results of the regression given in Table 5 show that there is no correlation between Animacy and the Position of the tokens in the Turkish data.

Table 5: Regression table for binomial GLM with the dependent variable Position and the independent variable Animacy in Turkish

fixed effect	β	σ	p-value	
(intercept)	-1.757e+01	2.284e+03	0.99	(ns)
Anim-adv	1.512e+01	2.284e+03	0.99	(ns)
Anim-anim	-5.966e-08	2.412e+03	1.00	(ns)
Anim-bp	1.498e+01	2.284e+03	0.99	(ns)
Anim-hum	1.612e+01	2.284e+03	0.99	(ns)
Anim-inan	1.446e+01	2.284e+03	0.99	(ns)

6.2.2 Weight

Weight is a discrete numeric variable, ranging from one to four and more phonological words. The results of the regression provided in Table 6 show that there is a positive effect of weight on the placement of the elements in post-predicate position, meaning that heavier elements (those that consist of three or four phonological words) are more likely to be placed in post-predicate position than those with the weight of one or two phonological words.

Table 6: Regression table for binomial GLM with the dependent variable Position and the independent variable Weight in Turkish

fixed effect	β	σ	p-value	
(intercept)	-3.32	0.35	< .0001	
Weight	0.5	0.19	0.008	**

Thus, for instance, in example (22) the structure *yolun diğer tarafından* 'from the other side of the street' consists of three phonological words and is placed in post-predicate position.

(22)   Turkish (Iefremenko 2021b: U, 622)
       bi   tane araba gel-iyor-du   *yol-un   diğer taraf-ın-dan*
       one piece car   come-PROG-PST.3SG road-GEN other side-POSS-ABL
       'There was a car coming from the other side of the street.'

6.2.3 Flagging

Flagging is a categorical independent variable with four levels in the Turkish data. The result of the regression analysis presented in Table 7 shows that there is no effect of flagging on the placement of arguments in pre- or post-predicate position in Turkish.

Table 7: Regression table for binomial GLM with the dependent variable Position and the independent variable Flagging in Turkish

fixed effect	β	σ	p-value	
(intercept)	-2.25	0.37	< .0001	
Flag-case	-0.38	0.41	0.35	(ns)
Flag-postp	0.05	0.71	0.94	(ns)
Flag-postp-relnoun	-0.5	0.81	0.47	(ns)

6.2.4 Role

Role is a categorical independent variable with 16 levels in the Turkish data. The results of the regression provided in Table 8 show that there is no correlation between the Role of a token and its Position in relation to the verb.

Thus, the analysis of word order in Turkish of Kurmanji-Turkish bilingual speakers showed that the only significant predictor is the weight of the elements: those constituents consisting of three and four phonological words are more likely to be placed in post-predicate position. Unlike Kurmanji, in Turkish such variables as Role or Flagging do not have an effect on the placement of arguments in post-predicate position in my data.

Before proceeding to the discussion of results, I would like to briefly discuss the findings from the Turkish data from Erzurum (and Erzurum Province) (Dogan 2021) that were analysed using the same scheme of the WOWA project. From the first look at the normalized numbers of different roles of the tokens, it is evident that the data sets from Ankara and Erzurum differ in the distribution of post-predicate arguments. For this reason, I have conducted similar analyses for the

Table 8: Regression table for binomial GLM with the dependent variable Position and the independent variable Role in Turkish

fixed effect	β	σ	p-value	
(intercept)	-2.68	0.46	< .0001	
Role-addr	1.38	0.79	0.08	(ns)
Role-ben	0.93	0.71	0.19	(ns)
Role-com	1.29	0.79	0.10	(ns)
Role-cop	-15.88	3261.3	0.99	(ns)
Role-cop-loc	-15.88	4612.2	0.99	(ns)
Role-do	-0.67	0.74	0.36	(ns)
Role-do-def	-15.88	1171.5	0.98	(ns)
Role-goal	0.16	0.56	0.77	(ns)
Role-goal-c	0.19	1.13	0.86	(ns)
Role-instr	-15.88	1966.64	0.99	(ns)
Role-loc	-0.3	0.65	0.63	(ns)
Role-other	0.65	0.61	0.28	(ns)
Role-poss	-15.88	6522.6	0.99	(ns)
Role-rec	-15.88	4621.2	0.99	(ns)
Role-stim	21.24	6522.6	0.99	(ns)

Erzurum data as I did for the data from Ankara (see Appendix for the results of the regression analyses).

First of all, the Turkish data from Erzurum is substantially pre-predicate (with around 12% of the utterances containing post-predicate positions), that is very similar to the Turkish data collected in Ankara. However, the regression analyses run on the Turkish data from Erzurum showed that there is no effect of weight, meaning that regardless of the number of phonological words that tokens consist of, they can be placed in the pre- or post-predicate position with no preference. But the two variables that have a positive effect on the employment of the post-predicate position are Role and Flagging. Namely, the model with Role as an independent variable showed that goals of motion are often placed in post-predicate position in Turkish. At the same time, the regression model with Flagging as an independent variable demonstrated that tokens flagged with case or a postpositional relational noun are likely to be placed in post-predicate position.

Looking back at the results of the Kurmanji data from Ankara, we see that these two variables, Role and Flagging, were also found to be significant. Namely,

the analysis of the Kurmanji data showed that goals of motion, and tokens flagged with case or a locational noun (sometimes in a combination with a preposition) are likely to be placed in post-predicate position in Kurmanji. Hence, similar factors influence the employment of the post-predicate position in the data sets from Kurmanji in Ankara and Turkish in Erzurum, while there are no common factors in the Turkish data sets collected in Ankara and Erzurum.

Nevertheless, it must be acknowledged that the two Turkish data sets (from Ankara and Erzurum) are not exactly comparable. First, the Turkish data from Erzurum come from three speakers, whereby one speaker is a young adult and two speakers are in their sixties, and their educational status is not known, unlike the speakers in Ankara who are all educated young adults. Second, the data were collected in the nineties, nearly 30 years prior to the data collection in Ankara. But what is most important is that it is not known whether the speakers from Erzurum are in fact bilingual in Turkish and Kurmanji, or if they are Turkish monolingual speakers. But what definitely differentiates the speakers in Erzurum from the speakers in Ankara is the number of Kurmanji speakers in the community in general and the societal status of Turkish. Even though Turkish certainly remains the language of formal contexts (such as education, business, etc.), Kurmanji is used more extensively in informal contexts, compared to Ankara (though we cannot compute the index of language use for the Erzurum data due to the absence of metadata for the speakers).

7 Discussion

In the study, I investigated word order, namely the post-predicate domain, in Turkish and Kurmanji that have been in a long-lasting contact with each other. Both languages are OV, but each of them employs the post-predicate position in a different way. While in Turkish the word order is determined by information structural requirements and the post-predicate position is reserved for background information, in Kurmanji the post-predicate position is the position for Goal arguments, particularly those flagged with case. Taking into account the different conditions encoded in WOWA, I investigated whether there is a structural convergence in word order in Turkish and in Kurmanji in Turkey.

First of all, the quantitative analysis showed that both languages predominantly place arguments in the pre-predicate position: with around 20% of the utterances in Kurmanji and 10% of the utterances in Turkish being post-predicate. Thus, both languages retain OV word order.

For Kurmanji, the regression analyses showed that the employment of the post-predicate position depends on the semantic role of the elements in the respective

clause and their flagging. Namely, goals of motion and caused motion are likely to be placed in post-predicate position, and this is in line with what has been described in previous research on Kurmanji word order (Haig 2015; Haig 2019; Gündoğdu 2019). As for addressees and recipients, the regression analysis did not show that there is a tendency to place them post-predicatively in my data. The main reason for this is different types of flagging: while the post-predicatively placed goals are flagged with case, all the pre-predicate examples of addressees and recipients are flagged with a circumposition or a postposition. Besides, in general I have not found many examples of addressees and recipients in my data possibly because the video shown to the participants did not trigger the use of verbs of speech and verbs of transfer. In fact, another variable that showed an effect on the placement of tokens in post-predicate position in Kurmanji is flagging. That is, tokens that are flagged with case, locational noun, or a preposition together with a locational noun are likely to be placed in post-predicate position. Thus, so far the analysis showed that the Kurmanji data is in line with most of the previous studies on word order in Kurmanji: Goal arguments flagged with case or a locational noun are placed post-predicatively.

At the same time, the qualitative analysis of the Kurmanji data showed that, albeit infrequently, such elements as location and source of motion are placed in the post-predicate position. Nevertheless, it is not clear whether the use of these elements in post-predicate position is language contact-induced, for two reasons: First, I did not find enough examples of this kind, and second, the analysed data is spoken and therefore is full with afterthoughts and self-repairings, and it is not fully clear whether certain post-predicative elements are in fact afterthoughts or self-corrections.

Another interesting observation from the Kurmanji data is that there are examples where case-flagged goals of motion are placed in pre-predicate position. Remember that in Section 2, I emphasized that case-flagged goals are licensed in post-predicate position. However, my data demonstrate that such constructions in Kurmanji can also be pre-predicate, but this concerns only one construction *erdê ketin* 'to fall on the ground', whereby *erdê* 'on the ground' as a case-flagged Goal argument is placed before the verb *ketin* 'to fall'. Interestingly, the same construction in the contact language Turkish is rendered as *yere düşmek*, whereby the Goal argument *yere* 'on the ground' is always placed in pre-predicate position. I argue that the pre-predicate position of the argument *erdê* occurs due to transfer of the whole construction from the dominant language Turkish. Besides, the reason why transfer occurs only with this particular construction might lie in the verb itself: in Kurmanji there are other set phrases with the verb *ketin* where

the Goal is always placed in pre-predicate position, e.g., *bi rê ketin* 'to set off (on a journey)'.

As for the post-predicate position in Turkish, the analysis showed that the only significant predictor of the employment of the post-predicate position, among those encoded in WOWA, is the weight of tokens, that is, tokens that consist of three and four phonological words are likely to be placed in post-predicate position. The placement of heavier constituents in post-predicate position is generally a phenomenon typical of spoken language. As Schroeder (1995) emphasizes, in spoken discourse information is conveyed in smaller chunks to make it more accessible to the hearer and the post-predicate position allows the hearer to keep track of the topical development and the deictic framework in which a predication takes place. On the other hand, the analysis of the Turkish data from Erzurum, where the contact with Kurmanji is more intense and Kurmanji is present in more spheres of life compared to Ankara, showed that Turkish speakers in Erzurum tend to place goals of motion, particularly those flagged with case and relational noun, in post-predicate position. At the same time, weight of constituents did not have an effect on the placement of constituents in the data from Erzurum.

In sum, the results of the analysis based on the WOWA encoding scheme show a minimal degree of convergence between the two languages of Kurmanji-Turkish speakers in Ankara. In Kurmanji semantic role of constituents and their flagging are determining factors in the placement of the constituents in relation to the verb, which is in line with the previous research (Haig 2015; Haig 2019; Gündoğdu 2019). In Turkish weight was proven to be a significant factor: longer constituents are more likely to be placed in the post-predicate position. The result that points to possible signs of an ongoing contact-induced language change is the changes in the word order of particular constructions. However, such changes are observed only in the minority language Kurmanji and not the majority language Turkish. Thus, such results point to the effect of societal status of the languages on the direction of the language change: a more prestigious language Turkish influences a less prestigious Kurmanji in Ankara.

At the same time, the analysis of the Erzurum Turkish data suggests that another social factor – the intensity of contact – has an impact on the occurrence of changes in word order of the languages in contact. Unlike Turkish in Ankara, in Erzurum Turkish semantic role and flagging have an effect on the employment of the post-predicate position. Hence, when Kurmanji is present in more spheres of life and the community size is bigger, changes also happen in the majority language Turkish. Even though it is not clear whether the speakers in the Erzurum data set are in fact bilinguals, the predictors for the placement of tokens post-predicatively are the same as in Kurmanji.

Finally, it is important to emphasize that information structure was not encoded in WOWA and, consequently, has not been analyzed in this chapter. Therefore, it remains unclear whether information structural constraints are loosened in the Turkish of bilinguals and whether information structure plays a role in the word order of Kurmanji as a result of contact with Turkish. This presents a limitation of the current study; however, the role of information structure on word order in Turkish and Kurmanji of the same speakers is discussed in another study by Iefremenko (Submitted).

Abbreviations

ABL	ablative	NEG	negation
ACC	accusative	OBL	oblique
CIRC	circumposition	PL	plural
DRCT	directional	POSS	possessive
EZ	ezafe	POSTP	postposition
F	feminine	PROG	progressive
FUT	future	PRS	present
GEN	genitive	PST	past
INDEF	indefinite	SG	singular
M	masculine	SUBJ	subjunctive

Acknowledgements

References

Asadpour, Hiwa. 2022. *Typologizing word order variation in northwestern Iran.* Frankfurt am Main: Goethe-Universität. (Doctoral dissertation).

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67. 1–48.

Chafe, Wallace. 1985. Linguistic differences produced by differences between speaking and writing. In David Richard Olson, Nancy Torrance & Angela Hildyard (eds.), *Literacy, language and learning: The nature and consequences of reading and writing*, 105–123. Cambridge: Cambridge University Press.

Croft, William & D. Alan Cruse. 2004. *Cognitive linguistics*. Cambridge: Cambridge University Press.

Csató, Éva Ágnes. 2002. Karaim: A high-copying language. In Mari C. Jones & Edith Esch (eds.), *Language change. The interplay of internal, external and extralinguistic factors*, 315–328. Berlin: De Gruyter Mouton.

Dogan, Mahîr. 2021. Oghuz (Erzurum). In Geoffrey Haig, Donald Stilo, Nils N. Schiborr & Mahîr Dogan (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation*. (accessed on October 20, 2022). Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa.

Doğruöz, A. Seza & Ad Backus. 2007. Postverbal elements in immigrant Turkish: Evidence of change? *International Journal of Bilingualism* 11(2). 185–220. DOI: 10.1177/13670069070110020301.

Erguvanlı, Eser. 1984. *The function of word order in Turkish grammar*. Berkeley: University of California Press.

Gündoğdu, Songül. 2019. Asymmetries in Kurmanji morphosyntax. In Songül Gündoğdu, Ergin Öpengin, Geoffrey Haig & Erik Anonby (eds.), *Current issues in Kurdish linguistics*, 93–115. Bamberg: University of Bamberg Press.

Haig, Geoffrey. 2015. Verb-goal (VG) word order in Kurdish and Neo-Aramaic: Typological and areal considerations. In Geoffrey Khan & Lidia Napiorkowska (eds.), *Neo-Aramaic and its linguistic context*, 407–425. Piscataway, NJ: Gorgias Press.

Haig, Geoffrey. 2019. Northern Kurdish (Kurmanjî). In Geoffrey Haig & Geoffrey Khan (eds.), *The languages and linguistics of Western Asia: An areal perspective*, 106–158. Berlin: De Gruyter Mouton.

Haig, Geoffrey. 2022. Post-predicate constituents in Kurdish. In Yaron Matras, Geoffrey Haig & Ergin Öpengin (eds.), *Structural and typological variation in the dialects of Kurdish*, 335–377. Cham: Springer International Publishing. DOI: 10.1007/978-3-030-78837-7_8.

Haig, Geoffrey & Yaron Matras. 2002. Kurdish linguistics: A brief overview. *STUF-Language Typology and Universals* 55(1). 3–14.

Haig, Geoffrey, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils Schiborr. 2024. Post-predicate elements in the Western Asian Transition Zone: Data, theory, and methods. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 3–54. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266331.

Haig, Geoffrey & Hannah Thiele. 2014. Post-predicate Goals in Northern Kurdish and neighbouring languages: A pilot study in quantitative areal linguistics languages. Paper presented at the *Second International Conference on Variation and Change in Kurdish, mardin artuklu university, turkey,* 8-9 October 2014.

Heine, Bernd. 2008. Contact-induced word order change without word order change. In Peter Siemund & Noemi Kintana (eds.), *Language contact and contact languages*, 33–60. Amsterdam: Benjamins. DOI: 10.1075/hsm.7.04hei.

Iefremenko, Kateryna. 2021a. Kurdish (Northern, Ankara). In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* (accessed on October 20, 2022). Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa.

Iefremenko, Kateryna. 2021b. Oghuz (Ankara). In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/#turkic.

Iefremenko, Kateryna. Submitted. *Word order in Turkish and Kurmanji Kurdish in language contacts.* Potsdam: University of Potsdam. (Doctoral dissertation).

Iefremenko, Kateryna, Cem Keskin & Christoph Schroeder. 2023. The post-predicate position in Turkish in language contact situations: How resistant to change? Paper presented at the *45th Annual Meeting of the German Linguistic Society* (DGfS) in Cologne, Germany 7-10 March 2023.

Johanson, Lars. 2002. *Structural factors in Turkic language contacts*. Richmond/Surrey: Curzon Press.

Keskin, Cem. 2023. On the directionality of the Balkan Turkic verb phrase: Variationist and theoretical perspectives. *Languages* 8(1). 2. DOI: 10.3390/languages8010002.

Kornfilt, Jaklin. 1997. *Turkish*. London/New York: Routledge.

Matras, Yaron et al. 2016. *The dialects of Kurdish.* University of Manchester. https://www.kratylos.org/~raphael/kurdish/database/.

Menz, Astrid. 1999. *Gagausische Syntax: Eine Studie zum kontaktinduzierten Sprachwandel*. Wiesbaden: Harrassowitz.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna. https://www.R-project.org/.

Schroeder, Christoph. 1995. Post-predicate arguments in Turkish. *Sprachtypologie und Universalienforschung* 48. 204–218.

Schroeder, Christoph, Kateryna Iefremenko & Mehmet Öncü. In press. The postverbal position in heritage Turkish. A comparative perspective with a focus on non-clausal elements. In Zeynep Kalkavan-Aydın & Yazgül Şimşek (eds.), *Zweisprachigkeit Deutsch-Türkisch*. Münster: Waxmann.

Sheyholislami, Jaffer. 2015. Language varieties of the Kurds. In Wolfgang Taucher, Mathias Vogl & Peter Webinger (eds.), *The Kurds: History–religion–language-politics*, 30–51. Vienna: Austrian Federal Ministry of the Interior.

Thomason, Sarah G. 2001. *Language contact*. Edinburgh: Edinburgh University Press.

Thomason, Sarah G. & Terrence Kaufman. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy Mc-Gowan D'Agostino McGowan, Roman François, Garrett Grolemund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Lin Pedersen, Evan Miller, Stephan Milton Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo & Hiroaki Yutani. 2019. Welcome to the Tidyverse. *Journal of Open Source Software* 4(43). 1686. DOI: 10.21105/joss.01686.

Wiese, Heike. 2020. Language situations: A method for capturing variation within speakers' repertoires. In Yoshiyuki Asahi (ed.), *Methods in dialectology XVI*, 105–117. Frankfurt a. M.: Lang.

Wiese, Heike, Artemis Alexiadou, Shanley Allen, Oliver Bunk, Natalia Gagarina, Kateryna Iefremenko, Esther Jahns, Martin Klotz, Thomas Krause, Annika Labrenz, et al. 2021. *RUEG Corpus*. Available online. DOI: 10.5281/ZENODO.3236068.

Yağmur, Kutlay. 2001. Turkish and other languages in Turkey. In Guus Extra & Durk Gorter (eds.), *The other languages of Europe. Demographic, sociolinguistic and educational perspectives*, 407–426. Clevedon: Multilingual Matters.

Appendix A Erzurum Turkish (Turkic, Dogan 2021)

Table 9: **Animacy:** Regression table for binomial GLM with the dependent variable Position and the independent variable Animacy in Turkish spoken in Erzurum

fixed effect	β	σ	p-value	
(intercept)	-1.557e+01	1.029e+03	0.98	(ns)
Anim-adv	1.439e+01	1.029e+03	0.98	(ns)
Anim-anim	1.279e+01	1.029e+03	0.99	(ns)
Anim-bp	1.339e+01	1.029e+03	0.99	(ns)
Anim-hum	1.331e+01	1.029e+03	0.99	(ns)
Anim-inan	1.383e+01	1.029e+03	0.98	(ns)
Anim-other	2.205e-08	1.188e+03	1.00	(ns)

Table 10: **Weight:** Regression table for binomial GLM with the dependent variable Position and the independent variable Weight in Turkish spoken in Erzurum

fixed effect	β	σ	p-value	
(intercept)	-2.10	0.29	< .0001	
Weight	0.16	0.19	0.39	(ns)

Table 11: **Flagging:** Regression table for binomial GLM with the dependent variable Position and the independent variable Flagging in Turkish spoken in Erzurum

fixed effect	β	σ	p-value	
(intercept)	-3.20	0.45	< .0001	***
Flag-case	1.46	0.47	0.02	**
Flag-postp-relnoun	2.25	0.63	0.0003	***

Table 12: **Role:** Regression table for binomial GLM with the dependent variable Position and the independent variable Role in Turkish spoken in Erzurum

fixed effect	β	σ	p-value	
(intercept)	-2.30	6.05	< .0001	***
Role-addr	5.129e-02	9.588e-01	0.95	(ns)
Role-becm	-9.531e-02	1.207e+00	0.93	
Role-ben	1.386e+00	8.466e-01	0.1	(ns)
Role-cop	-1.526e+01	9.890e+02	0.98	(ns)
Role-do	-7.793e-01	7.588e-01	0.3	(ns)
Role-do-def	-2.632e-02	6.632e-01	0.96	(ns)
Role-goal	2.042e+00	6.438e-01	0.001	**
Role-goal-c	1.022e+00	7.032e-01	0.14	(ns)
Role-instr	-1.526e+01	1.251e+03	0.99	(ns)
Role-loc	-6.592e-01	1.251e+03	0.43	(ns)
Role-other	1.372e-14	9.574e-01	1.00	(ns)
Role-poss	-1.526e+01	2.797e+03	0.99	(ns)
Role-rec	-3.365e-01	9.499e-01	0.72	(ns)
Role-stim	-1.526e+01	3.956e+03	0.99	(ns)

Chapter 3

Post-predicate prosody in OV languages

Stavros Skopeteas

University of Göttingen

This article examines the prosodic properties of post-predicate material in OV languages of the Western Asian Transition Zone. The core question is whether the right edge of the predicate is associated with the edge of a prosodic domain. The facts reported for these languages are summarized in two major phenomena that reveal an asymmetry between the material at the left and right side of the predicate. In some languages (e.g., in Standard Turkish), the nuclear stress must be realized within the domain preceding the right edge of the predicate, that is the pre-predicate material or the predicate itself. In some languages (e.g., in Georgian), the post-predicate material is separated from the predicate with a prosodic event that demarcates a prosodic constituent. Both phenomena provide evidence for a syntactic constituent that is mapped on prosody and separates the post-predicate elements from the core clause.

1 Introduction

In a study on the punctuation of the Georgian translation of the *Physiologus* (11th c. CE), Boeder (1991) reports that scribes often used punctuation within clauses indicating – or possibly prescribing – the prosodic phrasing of subclausal units. The use of punctuation in this manuscript revealed an interesting asymmetry between material at the left and the right side of the verb: setting clitics aside, arguments or adjuncts following the verb were often separated by punctuation, as indicated by the colon in (1), but this did not apply when the same elements appeared preverbally.

(1) Old Georgian (*Physiologus* 6.177.29, edited by Marr 1904; cited from
     Boeder 1991)
     da   sameupo-s i-p̣ov-eb-in   : kalandr-i   igi
     and realm-NOM PASS-[3SG]find-SM-PASS   charadrius-NOM DEM.REM
     'and it is found in the realm, that charadrius (bird).'

The asymmetry between the left and the right side of the predicate, as reported
in Boeder (1991), opens an interesting agenda: is there a difference between pre-
predicate and post-predicate elements regarding their prosodic properties? Un-
der which conditions does the post-predicate material appear outside the phono-
logical domain that contains the predicate? In particular for languages with OV
properties, a reasonable hypothesis is that the prosodic separation of the post-
predicate domain may reflect properties of a predicate-final constituent structure.
In many languages, prosodic structure is used for the demarcation information
structural domains. At the empirical side, the challenge is to disentangle effects
of constituent structure and effects on information structure. At the analytical
side, the interesting question is whether the information structural options of a
language can be partially predicted from the array of possible prosodic structures
in this language.

In order to tackle these questions, the present study examines selected OV
languages of the Western Asian Transition Zone, which comprises the area of
Northern Iran, Northern Iraq, Eastern Turkey and the Caucasus (Stilo 2015: 350);
see Chapter 1, this vol. The languages in this area are not syntactically uniform,
but most languages generally share several "OV" properties, such as preverbal
placement of bare objects and auxiliaries following the lexical verb, which at
least distinguish them from VO languages. The contextual conditions of OV and
VO differs very much between languages, as established by various studies (Stilo
2018, Haig et al. In press, Chapter 1, this vol.), which will be one of the main
issues in the following discussion (see Section 2).

The primary data discussed in the present study was elicited with native speak-
ers of Turkish, Georgian, Caucasian Urum, Eastern Armenian, and Persian. The
aim of this elicitation was to obtain comparable data that illustrate the phenom-
ena at issue; however, the main source of the discussed generalizations is the
available research on these languages.[1] Information about further languages of

---

[1]The data were collected in qualitative interviews with native speakers, who were raised in
the object language and currently use it in their everyday life. Beyond their native language,
these speakers were also competent in English or German. The Caucasian Urum speaker was
also native in Russian and Georgian; the further speakers did not speak other languages of
the area. The speakers prepared the sentences in collaboration with the instructor and were
instructed to imagine a conversation in which the scripted utterances are performed as answers
to context questions. They were free to repeat their performance until they were confident that
it corresponds to a natural way to produce the target utterance in the given context.

this area is drawn from the available research. The majority of the available sources is based on controlled data, either scripted speech or controlled speech-production tasks. This method comes with limitations in the generalizability of the findings, which should not be neglected.

Based on the available comparisons between spontaneous and controlled data in the phenomena under consideration, the major issue is that elicited data may present "idealizations" of what happens in real discourse. A comparison between scripted speech and semi-spontaneous narratives in Georgian reports that post-predicate elements are often separated by a high prosodic boundary from the predicate in either type of data: spontaneous narratives differ from controlled speech in that they display greater variability (Skopeteas et al. 2018: 44). This means that controlled data present an "idealization" of patterns that are learnt in real life (Stokhof & van Lambalgen 2011). Idealization of phonological events may have the form of hyper-articulation, e.g., expansion of the pitch range of tonal events that are less salient in spontaneous speech. In a study on Persian, spontaneous data were found to be often under-specified in comparison to controlled data: since prosodic demarcation is often not necessary for conveying the intended content, it is expected that real communication will be less rich in prosodic marking than the performance in careful speech (Sadat-Tehrani 2017: 28–31). Finally, the difference does not only lie in the clarity of signaling a certain information structure, but also in the intention of the speaker to convey the information structure of the utterance or not. Information structure is not an automatic reflex of the context, which often results in discrepancies between the intuitions of the speakers and what we find in corpora. For instance, Forker & Belyaev (2016: 243) report that some speakers of Tsakhur do not accept postverbal objects in narrow focus (with reference to Testelets 1999), but postverbal objects appear in Tsakhur texts as answers to questions that license a narrow focus on the object. This discrepancy may have two explanations: (a) the speakers report wrong generalizations about their native language; (b) since the information structure of the answer is made clear through the context, an answer that does not necessarily mark narrow focus is fully adequate in view of the communicative goals. Hence, these cases involve the challenge of identifying the intention of the speaker in producing an utterance in a certain context. This is not trivial of course, but it is relevant for the interpretation of the facts – if we do not assume that the relation between context and target utterance is deterministic. Finally, either type of speech production may be influenced by phenomena that are not related to the matter at issue: scripted speech may be performed with listing intonation (i.e., continuation rises at the end of declarative clauses), if the native speakers are not instructed to realize the utterances as complete discourse

units; spontaneous data often show reflexes of speech planning, e.g., hesitation pauses, which lead to shorter intonational units than in controlled settings (Sadat-Tehrani 2017: 31-33).

With this background, the goal of the present study is to identify prosodic properties of post-predicate elements in the available data and to set up hypotheses about their relation to syntax and/or information structure. The properties of post-predicate elements are not uniform across constructions: at least in many Iranian languages, the verb governs direct objects on its left and oblique objects (especially Goal arguments) on its right side, as it has been demonstrated with rich data in Haig & Rasekh-Mahand (2019), Haig et al. (In press), and Rasekh-Mahand et al. (2024 [this volume]). This difference is also reflected in the focus options and prosodic properties of the different classes of complements (Sadat-Tehrani 2007: 138–139), which means that post-predicate elements form different types of domains with the predicate, which are also reflected in prosody. In order to restrict the sources of variation, we will only examine preverbal and postverbal direct objects in the present study.

The following exposition starts with an overview of the variation between languages in the information structure of post-predicate objects; see Section 2. After a short summary of the necessary assumptions for prosodic description in Section 3, the core part of the study is organized around the information structure of post-predicate objects. The basic distinction is whether the post-predicate object is part of the partition of the utterance that is focused (i.e., the focus domain) or not. In the former case, it can be part of a larger focus domain that contains the post-predicate and the predicate: these are instances of "broad focus," that may contain a VP or an entire sentence (see Section 4). Alternatively, the focus domain of the utterance may be exactly the post-predicate object; these are cases of "narrow focus" (i.e., a focus domain only containing a simple lexical projection such as a noun phrase) (see Section 5). Finally, the post-predicate object may be outside the focus domain (or simply "out of focus"), in which case it is background information (see Section 6). Based on the properties introduced in these sections, Section 7 draws conclusions about the interaction between syntax and prosody in the realization of post-predicate elements.

2 Focus options

V-final languages of the Western Asian Transition Zone differ regarding the focus possibilities of post-predicate objects. They either cannot be focused (Turkish), or they can be part of a narrow/broad focus domain (Georgian) or of a broad focus domain only (Persian).

Post-predicate elements in Modern Standard Turkish can be either background information (discourse given information that is outside the focus domain of the utterance) or afterthoughts (discourse new information that completes the utterance) (see Taylan 1984: 50–56, İşsever 2003, Kilicaslan 2004: 727-728). Afterthoughts are a different type of phenomenon since they were not part of the sentence plan at the critical time point that the speaker selected the linearization of her utterance. Leaving afterthoughts aside, post-predicate elements in this type of language cannot be focused. This phenomenon is reported for Western Armenian (Donabedian-Demopoulos 2018: §2.7), Laz (Lacroix 2019: 852), and Balochi (Delforooz 2010: 66-68). Northwest Caucasian languages presumably belong to this group as well – at least those languages like Ubykh in which postverbal constituents are rare (Forker 2021: 977).

In a second type of languages, post-predicate objects are contextually unrestricted. In Georgian, postverbal objects can be either part of a broad focus (e.g., a focus domain encompassing the entire clause) or a narrow focus (e.g., focus on the postverbal object) (Skopeteas & Fanselow 2010, Gosby 2016: 170, Borise 2019: 106, 235, Chapter 10, this vol.). Similar facts are reported for Mingrelian and Svan (Forker 2021: 994), Caucasian Urum (Schröter 2017: 221), Ossetic (Erschler 2012: 686), and at least some Nakh-Daghestanian languages (see summary in Forker 2021: 977). In Eastern Armenian, the same flexibility of the post-predicate domain applies to synthetic verbs (in contrast to periphrastic verbs, in which case the auxiliary cliticizes to the element bearing the nuclear stress, such that the material following the auxiliary is de-accented (Comrie 1984, Kahnemuyipour & Megerdoomian 2011, Samvelian et al. 2023).

The difference between languages with and languages without postverbal foci is well established in the research on V-final languages. Interestingly, there is a third pattern that deserves more attention. Forker & Belyaev (2016: 243) (with reference to Testelets 1999) report that some speakers of Tsakhur accept postverbal objects in broad focus, but not so in narrow focus. A similar asymmetry is gathered through speaker judgments in Persian: *specific* objects can follow the predicate when they are part of a broad focus domain, but they are judged to

be impossible in narrow focus (see Karimi 2003: 115 and Sadat-Tehrani 2007: 68, 138–139 for the latter statement); see (2)[2]

(2)   Persian (Y. Sanei, p.c.)

  a. Broad focus (entire clause)
    A: 'What happened?'
    B: *Nâzanin livân-am-o   var-dâšt.*
     Nazanin glass-POSS.1SG-RA up-have:PST[3SG]
     'Nazanin picked up my glass.'
    B´: *Nâzanin var-dâšt   livân-am-o.*
      Nazanin up-have:PST[3SG] glass-POSS.1SG-RA
     'Nazanin picked up my glass.'

  b. Narrow focus (on the object)
    A: 'What did Nazanin pick up?'
    B: *Nâzanin livân-am-o var-dâšt.*
    B´: #*Nâzanin var-dâšt livân-am-o.*

OV languages of this region share some properties in this respect. First, postverbal arguments can be background information. Furthermore, in all these languages there is evidence for a preverbal position for narrow focus, which is immediately left adjacent to the verb: Turkish (Göksel & Özsoy 2000), Laz (Lacroix 2019: 852), Georgian (Borise 2019: 235), Chechen (Komen 2013: 322), Nakh-Daghestanian (in general) (Forker 2021: 977), Northwest Caucasian (in general) (Forker 2021: 986), Ossetic (Erschler 2012: 686, Borise & Erschler 2023), Eastern Armenian (Comrie 1984: 19, Samvelian et al. 2023: 464–467), Persian (Kahnemuyipour 2001). Beyond the preverbal option, there is a lot of variation regarding

---

[2]These intuitions imply that speakers will use a preverbal object whenever they intend to signal that this constituent is narrowly focused. The prediction for corpus observations is that preverbal objects are *more likely* in contexts licensing a narrow focus on the object than in contexts licensing a broader focus domain including the object (assuming that a subset of the data from speech production are under-specified for information structure). The only corpus study that offers data for this comparison is a small-size corpus study on written Persian, showing that SVO appears rarely in sentence/predicate focus and never in narrow focus (Majidi & Dabir Moghaddam 2012: 861) – but the size of the corpus does not allow for strong statements. The corpus study by Roberts (2009: 139–141) on written/oral Persian shows that postverbal specific objects may be part of the focus domain, but does not consider instances of narrow focus. Forker & Belyaev (2016: 243) present examples with postverbal objects in Tsakhur that are answers to questions licensing a narrow focus on the object, which contradict the speakers' intuitions, but without excluding that these answers are underspecified for information structure since this is clearly indicated by the context (see discussion in Section 1).

the exact focus options in the preverbal domain, e.g., whether the preverbal elements can be marked for focus in their canonical position, whether contrastive foci may have an effect on linear order, as in Persian (Karimi 2003: 92), which is out of the scope of the present study. In contrast to preverbal foci, postverbal foci are not restricted to a certain position that is designated to focus: any element within the post-predicate domain can be focused by prosodic means.

The language types presented so far are summarized in Figure 1. The first node contains the already established distinction between languages in which the post-predicate domain cannot be focused and languages allowing for focus on post-predicate elements. Within the latter type, there is a distinction between languages that allow for any type of focus and those that only allow for broad focus in the post-predicate domain.

Can a post-predicate object be part of a broad focus domain?

no   Can a post-predicate object be in narrow focus?

Turkish, etc.   no   yes

Persian, etc.   Georgian, etc.
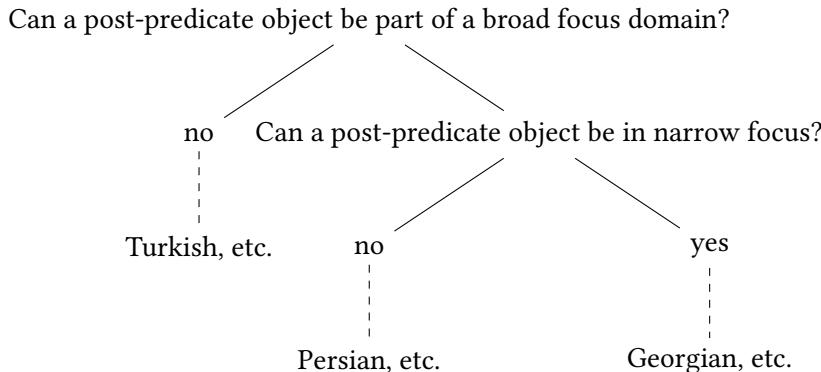
Figure 1: Focus options of post-predicate objects in OV languages of the Western Asian Transition Zone

Note that Figure 1 reveals an implicative relation between the focus options of post-predicate objects, since no language in this sample expresses narrow but not broad focus in the post-predicate domain (see 3).

(3)   Between focus options of post-predicate objects
       Broad focus ← Narrow focus

The origin of this asymmetry will be discussed in Section 7.2 in the light of the prosodic properties of the languages at issue, which are examined in the subsequent sections.

3 Prosodic assumptions

The description of the collected data in the following sections is built upon some basic assumptions about the relevant prosodic events and the syntax-prosody interface, that are introduced in the following. Prosodic constituents form a prosodic hierarchy such that constituents of higher layers optionally contain constituents of lower layers. The correspondence between prosodic constituents and syntactic constituents is roughly as follows: prosodic words (ω) correspond to morphological words including clitics; phonological phrases (φ) correspond to syntactic phrases; intonation phrases (ι) to clauses (Selkirk 2011: 437). These correspondences are modulated by purely phonological rules (e.g., constraints on the size of prosodic constituents), by other functions reflected in prosodic phrasing (e.g., focus) and by aspects relating to speech performance, such as speech tempo and hesitation breaks. For instance, the expression of focus may create prosodic constituents that do not correspond to syntactic phrases; see $(SV)_\varphi$ in example (7b) below. We assume that prosodic constituents allow for recursion, such that φ-phrases can be embedded in other φ-phrases as in $(\alpha\ (\beta)_\varphi)_\varphi$ (see discussion in Ladd 1986, Selkirk 2011: 455, Féry 2018: 78–85).

The pitch track annotations in the following examples follow the conventions of the Autosegmental-Metrical framework (Pierrehumbert 1980). Tonal events represent the assumed phonological entities that underlie the pitch contour – and not every peak or dip in the pitch excursion. The labels indicate the scaling of tonal targets with respect to their environment (H for 'high', L for 'low', L+H for 'rise', etc.), and for their association with two classes of entities: (a) edge tones are associated with the edges of prosodic constituents (H_φ: is a high tone associated with the right edge of a φ-phrase, $_\varphi H$ is a high tone associated with the left edge of a φ-phrase) and (b) pitch accents are associated with stressed syllables (H^* is a high tone associated with the stressed syllable; $L+H^*$ is a rise, whose H-target is associated with the stressed syllable). Languages often mark the constituent that hosts the nuclear stress of the intonation phrase (i.e., the constituent that is perceived with maximal prominence) with different tonal events from the prosodic constituents that precede or follow it ("prenuclear" or "postnuclear" prosodic constituents).

The phonetic cues of prosodic constituents are language specific. Prosodic constituents may be demarcated by tonal events or breaks at their edges as well as by effects on the duration of edge syllables (e.g., final lengthening). Further evidence for prosodic constituents comes from register lowering: the pitch scaling of tonal events at sister constituents incrementally decreases (Ladd 1988, Féry 2013).

4 Broad focus

A broad focus encompassing the entire utterance is elicited with a context question that does not introduce any presuppositions, such as 'What happened?'. Figure 2 illustrates the contrast between the SOV and the SVO order in this context in Georgian. The SOV utterance (Left Panel) is realized with a series of rising contours ($_{\varphi}$L...H$_{\varphi}$), stretching from a low target at the left edge to a high target at the right edge of the φ-phrase (see various analyses in Skopeteas et al. 2009, Vicenik & Jun 2014, Skopeteas et al. 2018, Borise 2021a). The second H$_{\varphi}$ is scaled lower than the first H$_{\varphi}$. The final verb is realized with low intensity and creaky voice, which results to a fragmentary pitch signal and ends with a final lowering, represented by the L target at the right edge of the ι-phrase (L$_{\iota}$). The SVO realization (Right Panel) differs (see detailed discussion in Skopeteas & Féry 2010, Skopeteas et al. 2018): the subject and the verb are integrated in a single phonological domain that is determined by a $_{\varphi}$H tone at the left edge and a H$_{\varphi}$ tone at the right edge. Crucially, the latter phrase tone is not lowered with reference to the former one. [3]
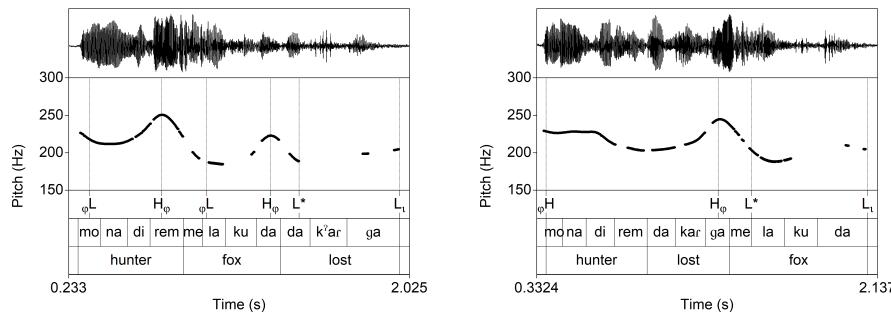
Figure 2: Broad focus in Georgian; answers to 'What happened?', Left Panel: SOV; Right Panel: SVO; morphemic transcription in (4)-(5)

The register lowering of the second H-target in the SOV order (Figure 2/Left Panel) is compatible with two alternative analyses, see (4): (a) the H-targets can be associated with the right edge of two sister φ-phrases that are mapped onto the preverbal arguments; (b) alternatively, the second φ-phrase can be associated with a φ-phrase mapped onto the object, that is nested within the φ-phrase of the

---

[3]The figures in this article contain an oscillogram (Top) and a track of the f0 (Middle) temporally aligned with the phonetic transcription of the utterance (Bottom); created in Praat (Boersma & Weenink 2023).

VP. Register lowering applies to this case too, since the φ-phrase of the VP is a sister constituent to the φ-phrase of the subject. Hence, both prosodic structures in (4) are possible and the realization of the example does not provide any further cues that would support the one or the other structure.

(4) Georgian (D. Kakashvili, p.c.)

$_φ$L   H$_φ$ $_φ$L   H$_φ$ L$_ι$
(($monadire$-$m_ω$)$_φ$ ($melakuda_ω$)$_φ$ da-$k'arg$-$a_ω$)$_ι$
(($monadire$-$m_ω$)$_φ$ (($melakuda_ω$)$_φ$ da-$k'arg$-$a_ω$)$_φ$)$_ι$
hunter-ERG   fox[NOM]   PFV-loose-AOR.S.3SG
'The hunter lost the fox.'

In the SVO utterance (Figure 2/RIGHT PANEL), the verb and the preverbal material are phrased as a single prosodic constituent. The left edge of this constituent is demarcated with a high phrase tone ($_φ$H), while the two phrase tones ($_φ$H and H$_φ$) are just interpolated. The relevant issue is that the predicate is phrased together with the preceding material and forms a separate prosodic constituent from the post-predicate. Quantitative studies on controlled and spontaneous data show that the high phrase tone at the right edge of the verb is a frequent (but not necessary) correlate of verb-medial orders (Skopeteas & Féry 2010, Skopeteas et al. 2018).

(5) Georgian (D. Kakashvili, p.c.)

$_φ$H   H$_φ$ L$_ι$
(($monadire$-$m_ω$ da-$k'arg$-$a_ω$)$_φ$ ($melakuda_ω$)$_φ$)$_ι$
hunter-ERG   PFV-loose-AOR.S.3SG   fox[NOM]
'The hunter lost the fox.'

The special role of the predicate in phrasing is also reported for other languages (see Borise 2021b: 771, 780 for a summary of earlier mentions in Adyghe, Circassian, and Chechen; see Hasan 2012: 275 on preverbal and postverbal quotations in Kurmanji), but more data is required in order to qualify these reports. However, this phenomenon does not apply to all OV languages of the area, as illustrated below for Persian.

The available accounts on Persian prosody do not predict a high phrase tone at the right edge of the verb for independent reasons: the nuclear φ-phrase cannot be after the predicate (except for Goal arguments) and this phrase normally ends with a low phrase tone (Sadat-Tehrani 2007: 9, 68).[4] Prenuclear φ-phrases are

---

[4]This does not mean that a high phrase tone at the right edge of the verb is impossible (see an example in Mahjani 2003: 54).

realized with rising contours that reach a H-target within the stressed syllable (i.e., a L+H* pitch accent). This pitch accent is followed by a phrase tone (H_φ) in prenuclear φ-phrases (Sadat-Tehrani 2007: 9, 42, 68) and is tonally compressed or totally erased in postnuclear φ-phrases (Abolhasanizadeh et al. 2012). The reflexes of this pitch accent are illustrated in Figure 3 with specific objects, which may precede or follow the V in Persian. The L+H* pitch accent is aligned with the stress, which is final with nouns (see *Nâza'nin* 'Nazanin') or falls on an earlier syllable in the presence of clitics (see *li'vân-am-o* 'glass-POSS.1SG-RA') or in the verb complex (see *ne'gâh kard* 'look do:PST[3SG]'[5]). Prenuclear and nuclear pitch excursions differ with respect to the presence of a H_φ phrase tone. In prenuclear φ-phrases, the pitch excursion rises up to the right edge of the phrase (see *li'vân-am-o* 'glass-POSS.1SG-RA' in the LEFT PANEL); in nuclear φ-phrases, the rising in the stressed syllable (L+H*) is followed by a falling pitch excursion (see *ne'gâh kard* 'look do:PST[3SG]' in the RIGHT PANEL), while the postnuclear domain is tonally compressed (Sadat-Tehrani 2007: 7–8, Ardali & Xu 2012). Nuclear rises are scaled lower than prenuclear rises and their H-target (H*) is aligned with the middle of the stressed syllable while the H-target of phrase tones (H_φ) is aligned with the right edge of the word (Sadat-Tehrani 2009, Hosseini 2014: 104–107).
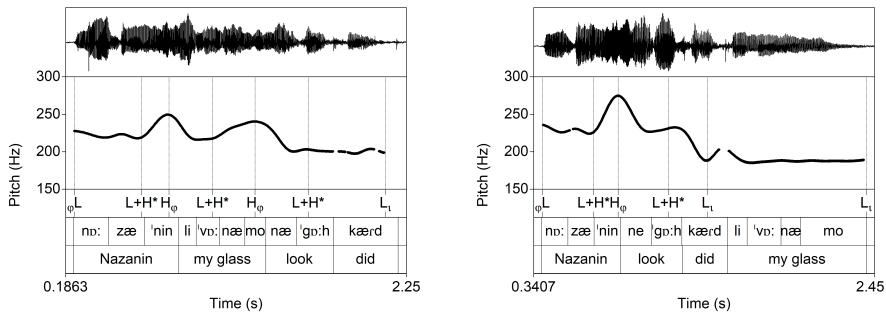
Figure 3: Broad focus in Persian; answers to 'What happened?', LEFT PANEL: SOV; RIGHT PANEL: SVO; morphemic transcription in (6)

In the SOV order with a specific object (LEFT PANEL), the nuclear stress falls on the verb and the φ-phrase of the object precedes the nucleus and is enclosed by a H_φ phrase tone (as expected for specific objects and in contrast to bare objects (Hosseini 2014: 81); see (6a)). In the SVO order (RIGHT PANEL), the nuclear stress

---

[5]Word stress is located on the final syllable of nouns/adjectives with the exception of unstressed enclitics; with complex verbs, the stress falls on the embedded non-verbal element that expresses the lexical content and not on the light verb (Kahnemuyipour 2003).

is again realized within the verb (as indicated by the absence of a H_φ), while the post-predicate domain is realized as a low plateau without any tonal events; see (6b). This realization follows from a general property of Persian prosody: (to the exception of Goal arguments) the nuclear stress cannot be realized in the domain following the predicate (Sadat-Tehrani 2007: 9, 68). Since the domain of the nuclear stress is the intonation phrase, we conclude that the low edge tone is associated with the right edge of an intonation phrase (L_ι).

(6) Persian (Y. Sanei, p.c.)

   a. L_φ L+H* H_φ L+H* H_φ L+H* L_ι
       $((nâza'nin_\omega\ )_\varphi$ $(li'vân\text{-}am\text{-}o_\omega$ $)_\varphi$ $(ne'gâh_\omega\ kard_\omega)_\varphi$ $)_\iota$
       Nazanin   glass-POSS.1SG-RA   look   do:PST[3SG]
       'Nazanin watched my glass.'

   b. L_φ L+H* H_φ L+H* L_ι L_ι
       $(((nâza'nin_\omega\ )_\varphi$ $(ne'gâh_\omega\ kard_\omega)_\varphi$ $)_\iota$ $(li'vân\text{-}am\text{-}o_\omega)_\varphi$ $)_\iota$
       Nazanin   look   do:PST[3SG]   glass-POSS.1SG-RA
       'Nazanin watched my glass.'

In sum, the examples of broad focus illustrated two distinct phenomena:

- In Georgian, post-predicate objects are prosodically separated from the predicate (by a φ-phrase boundary).

- In Persian, post-predicate objects are outside the phonological domain (presumably an ι-phrase) that may host the nuclear stress.

5 Narrow focus

In some languages, the post-predicate domain can host a narrow focus (see Section 2). Since all these languages also have a preverbal position hosting narrow foci (see 2), the relevant question is whether these options differ prosodically. The data presented in this section show that preverbal foci are prosodically integrated to the phonological constituent that contains the verb, while postverbal foci are prosodically separated from verb.

Caucasian Urum is an Anatolian dialect of Turkish spoken by a population that migrated to the Small Caucasus from Kars/Erzurum in the 19th century (Skopeteas 2016). Similar to the Anatolian vernaculars of Turkish – and in contrast to Standard Turkish – Urum allows for postverbal foci (Schröter 2017). Similarly to Turkish, prenuclear phrases are realized with rising contours delimited

by a low target at their left edge and a high target at their right edge ($_\varphi$L...H$_\varphi$). A falling pitch accent (H*+L) appears in the stressed syllable in nuclear φ-phrases or when the word stress precedes the final syllable (e.g., in case of compounds or unstressed enclitics); cf. Kamali (2011: 93), Güneş (2013), Féry (2018: 250–257). In Figure 4/Left Panel, a high phrase tone (H$_\varphi$) separates the preverbal focus from the preceding material; the focus is marked with a falling accent (H*+L), while the postnuclear domain (the verb) is de-accented. The postverbal focus (Figure 4/Right Panel) does not essentially differ: it is separated from the verb by a high phrase tone (H$_\varphi$), while the clause-final nuclear stress is marked with a low pitch accent (L*).
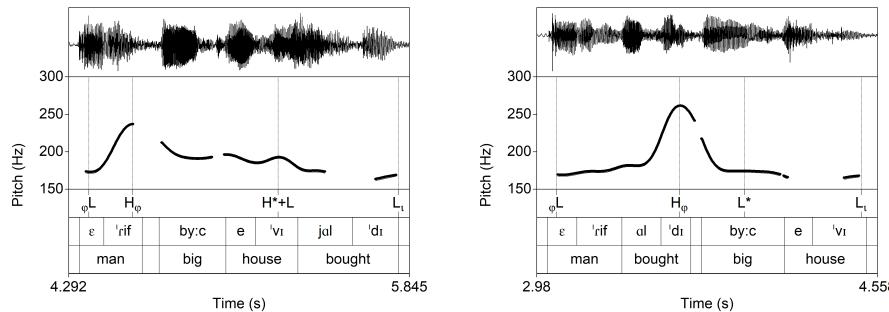
Figure 4: Narrow focus in Caucasian Urum; answers to 'What did the man buy?', Left Panel: SO$_F$V; Right Panel: SVO$_F$; morphemic transcription in (7)

There are several indicators of phrasing in the examples of Figure 4. The SOV order (Left Panel) displays a break between the S and O and the SVO order (Right Panel) between V and O. The epenthetic [j] between *ev-i* and *al-di* in the SOV order is inserted to avoid the hiatus, which indicates that there is no φ-phrase boundary between O and V. Furthermore, the H$_\varphi$ phrase tones are aligned with the right edge of the subject in SO$_F$V and of the verb in SVO$_F$. These facts suggest a phrasing (S)$_\varphi$(O$_F$V)$_\varphi$ for preverbal and (SV)$_\varphi$(O$_F$)$_\varphi$ for postverbal foci; see (7).

(7)   Caucasian Urum (V. Moisidi, p.c.)

a. $_\varphi$L   H$_\varphi$ H*+L   L$_\iota$
     ((*erif$_\omega$*   )$_\varphi$ (*büyük$_\omega$ ev-i$_\omega$ al-di$_\omega$*)$_\varphi$)$_\iota$
     man[NOM]   big   house-ACC buy:PST[3]
     'The/a man bought the/a big house.'

b. $_φ$L   H$_φ$ L*   L$_ι$
  ((*erif* $_ω$ *al-di* $_ω$)$_φ$ (*büyük* $_ω$ *ev-i* $_ω$)$_φ$)$_ι$
  man[NOM] buy:PST[3]   big   house-ACC
  'The/a man bought the/a big house.'

The prosody of Eastern Armenian has very similar properties: prenuclear φ-phrases are rising contours with $_φ$L and H$_φ$ tones at their edges – independent of word stress (see Toparlak 2019: 61–81 for a pitch-accent based account). Prenuclear φ-phrases show exactly this pattern in Figure 5: see the contour of the subject before a preverbal focus (LEFT PANEL) and the contours of the subject and the verb preceding a postverbal focus (RIGHT PANEL). Nuclear accents have the form of falling contours in declarative clauses (Dum-Tragut 2009: 53), differing from prenuclear accents in that (a) the peak of the H*+L accent is aligned earlier in the stressed syllable than the peak of the H$_φ$ phrase tone and (b) the nuclear accents are aligned with the stressed syllable and not with the final syllable in words with non-final stress, e.g., with unstressed enclitics; see Dolatian (2020: 864) on stress. In Figure 5, the nuclear accents (H*+L) are aligned with the stressed syllable of the narrow focus, either preverbal (LEFT PANEL) or postverbal (RIGHT PANEL).
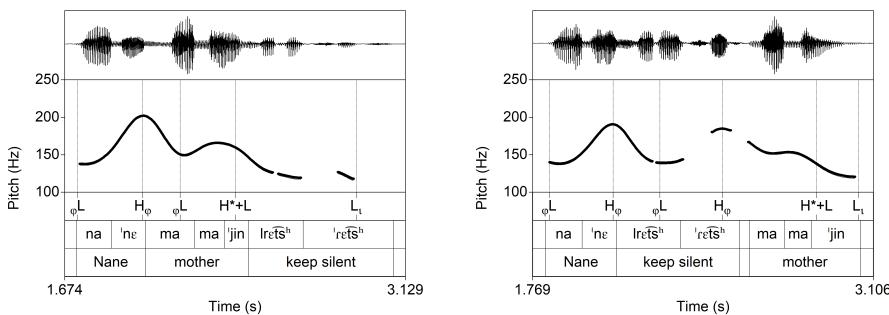
Figure 5: Narrow focus in Eastern Armenian; answers to 'Whom did Nane keep silent?', LEFT PANEL: SO$_F$V; RIGHT PANEL: SVO$_F$; morphemic transcription in (8)

The prosodic excursions of Figure 5 also differ with respect to the relative pitch scaling of the H-targets: register lowering applies to the second H-target of the LEFT PANEL but not to the second H-target of the RIGHT PANEL. The absence of register-lowering in the latter case is evidence that the right edge of the φ-phrase mapped on the verb is not a sister constituent to the φ-phrase of the subject, but it is a higher prosodic constituent; see ((S)$_φ$(V)$_φ$)$_φ$ in (8b).

(8) Eastern Armenian (H. Hovhannisyan, p.c.)

a.
$_φ$L $\quad\quad$ H$_φ$ $_φ$L $\quad$ H*+L $\quad\quad\quad\quad\quad\quad\quad\quad$ L$_ι$
((*nane*$_ω$ $\quad$)$_φ$ (*mama-i-n*$_ω$ $\quad\quad$ *lrre-c'r-ec'*$_ω$)$_φ$ $\quad$)$_ι$
Nane[NOM] $\quad$ mother-DAT-DEF silent-CAUS-AOR.3SG
'Nane kept mother silent.'

b.
$_φ$L $\quad\quad$ H$_φ$ $_φ$L $\quad\quad\quad\quad\quad\quad$ H$_φ$ $\quad\quad$ H*+L $\quad$ L$_ι$
((*nane*$_ω$ $\quad$)$_φ$ (*lrre-c'r-ec'*$_ω$)$_φ$ $\quad\quad$)$_φ$ (*mama-i-n*$_ω$)$_φ$ $\quad$)$_ι$
Nane[NOM] $\quad$ silent-CAUS-AOR.3SG $\quad$ mother-DAT-DEF
'Nane kept mother silent.'

The same difference between preverbal and postverbal foci is found in Georgian; see Figure 6. In the preverbal focus (LEFT PANEL), the prenuclear φ-phrase of the subject is separated from the focus with a high phrase tone – similar to the pattern observed in Caucasian Urum and Eastern Armenian. The initial syllable of the focus is realized with a steep fall towards a low target that is reached within the first syllable (L*). The material after this syllable is realized with low intensity (see oscillogram), very low pitch and creaky voice, which results into inaccurate pitch measurements (the real pitch level is much lower than the displayed pitch level in the pitch track). In contrast to the broad focus in Figure 2/LEFT PANEL, the material after the stressed syllable of the focus is de-phrased such that there is no H$_φ$ at the right edge of the focused object. In the postverbal focus in Figure 6/RIGHT PANEL, the prenuclear domain is mapped on a single φ-phrase, delimited by two high targets and a pitch excursion interpolating between the high edge events. After the high phrase tone at the right edge of the verb, the pitch is falling towards a low target within the first syllable of the focus, while the remainder is realized with a radical intensity drop and creaky voice (see earlier descriptions in Vicenik & Jun 2014: 177, Skopeteas & Féry 2010, Borise 2019: 291).

In either case (preverbal and postverbal focus), the intonational nucleus (i.e., the φ-phrase in narrow focus) is separated by a high phrase tone (H$_φ$) from the prenuclear material on its left side. This is a robust property of focus expressions in Georgian (Skopeteas & Féry 2010, Féry 2013: 711–713, Borise 2021a, Vicenik & Jun 2014: 181). It results in a difference in phrasing between preverbal and postverbal foci: while the former are phrased together with the verb, the latter are prosodically separated from the verb; see (9).
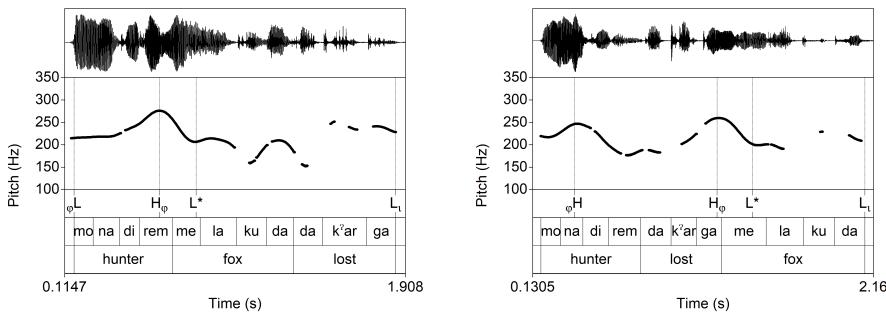
Figure 6: Narrow focus in Georgian; answers to 'What did the hunter loose?', Left Panel: SO_FV; Right Panel: SVO_F; morphemic transcription in (9)

(9)   Georgian (D. Kakashvili, p.c.)

    a. $_\varphi$L   H$_\varphi$ L*   L$_\iota$

        ((*monadire-m*$_\omega$)$_\varphi$ (*melakuda*$_\omega$ *da-k'arg-a*$_\omega$)$_\varphi$)$_\iota$

        hunter-ERG   fox[NOM]   PFV-loose-AOR.S.3SG

        'The hunter lost the fox.'

    b. $_\varphi$H   H$_\varphi$ L*   L$_\iota$

        ((*monadire-m*$_\omega$ *da-k'arg-a*$_\omega$)$_\varphi$ (*melakuda*$_\omega$)$_\varphi$)$_\iota$

        hunter-ERG   PFV-loose-AOR.S.3SG   fox[NOM]

        'The hunter lost the fox.'

The examples of this section demonstrate that preverbal foci are integrated into the φ-phrase of the predicate, while postverbal foci are separated by H$_\varphi$ phrase tones from the predicate. Crucially, the high φ-phrase tone does not depend on the syntactic category of the material preceding the focus: in the examples discussed so far, it appears on the right edge of S in SO_FV and the right edge of V in SvO_F. When a postpredicate focus is not adjacent to the verb, the phrase tone just appears at the right side of the preceding φ-phrase, as in Figure 7, with an adverb intervening between the V and the narrow focus. In the broad focus realization (Left Panel), the first φ-phrase is determined with H phrase tones at its left and right edge; the next H$_\varphi$ at the right edge of the adverb is lowered with reference to the earlier H$_\varphi$. The final object is realized with low pitch and a drop in intensity that renders the pitch indefinable after the stressed syllable. Crucially, when the final object is focused (Right Panel), the final object is preceded by a high boundary at the right edge of the prenuclear prosodic constituent that contains all material preceding the focus.
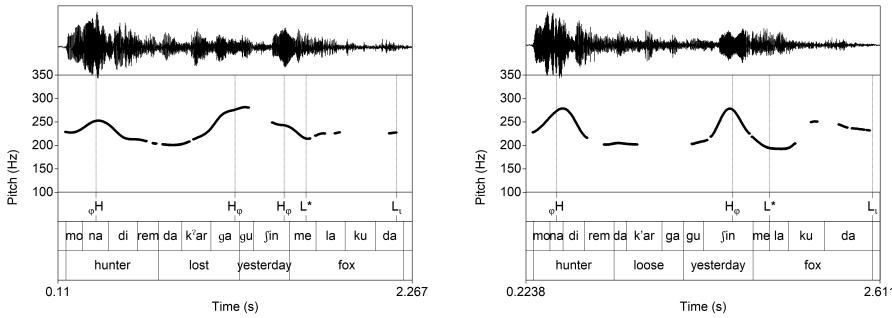
Figure 7: SVAdvO in Georgian; Left Panel: broad focus, answer to 'What happened?'; Right Panel: narrow focus, answer 'What did the hunter loose yesterday?'; morphemic transcription in (10)

In the broad focus realization (Left Panel), the predicate is phrased together with the preceding material, as in (10a), which confirms the observations in broad focus; compare phrasing in (5). Narrow focus on the clause-final object (Right Panel) induces a major φ-phrase boundary preceding the focus, as in (10b).

(10)   Georgian (D. Kakashvili, p.c.)

   a. $_\varphi$H   H$_\varphi$ H$_\varphi$ L*   L$_\iota$
        ((*monadire-m$_\omega$* *da-k'arg-a$_\omega$*   )$_\varphi$ (*gušin$_\omega$*   )$_\varphi$ (*melakuda$_\omega$*)$_\varphi$)$_\iota$
        hunter-ERG   PFV-loose-AOR.S.3SG   yesterday   fox[NOM]
        'The hunter lost the fox yesterday.'

   b. $_\varphi$H   H$_\varphi$ L*   L$_\iota$
        ((*monadire-m$_\omega$* *da-k'arg-a$_\omega$*   *gušin$_\omega$*   )$_\varphi$ (*melakuda$_\omega$*)$_\varphi$)$_\iota$
        hunter-ERG   PFV-loose-AOR.S.3SG yesterday   fox[NOM]
        'The hunter lost the fox yesterday.'

In sum, preverbal foci are phrased together with the predicate, while postverbal foci are prosodically separated. It is crucial that postverbal foci are separated from the prefocal material of any category, e.g., from a prefocal adverb in (10b). Hence, we are dealing with a prosodic property of focus that does not refer to the constituent structure. In all these languages, the left edge of the focus is aligned with a φ-phrase boundary at its left edge (see typology in Féry 2013), whose exponent is a high phrase tone at the right edge of the preceding φ-phrase. This phenomenon is independent of the syntactic category preceding the focus.

6 Out of focus

In all languages that allow for post-predicate elements, the post-predicate material can be background information, which is outside the focus domain of the utterance. Studies on the languages of the area conclude that the material following the nuclear stress (including post-predicate elements) is de-accented, which means that the tonal events are eliminated; see Persian in Ardali & Xu (2012), Sadat-Tehrani (2007: 7), and Rahmani et al. (2018),[6] and Turkish in Ipek (2011), Özge & Bozsahin (2010: 144), and Kamali (2011: 34), while de-accenting is reported to be optional in Georgian (Skopeteas et al. 2009: 115, Vicenik & Jun 2014: 177).

The question is whether the predicate has particular effects on prosodic phrasing when the post-predicate material is outside the focus domain. For instance, it is reported for Western Armenian – which belongs to the languages that use post-predicate material only for background information – that constituents following the verb are separated from it by a prosodic break (Donabedian-Demopoulos 2018: §2.7).

Similar phenomena apply to Standard Turkish, which is a further language that only allows background information to follow the predicate, as illustrated in the following. The nuclear stress cannot be realized after the predicate (Taylan 1984: 51, Göksel 1998: 103, Özge & Bozsahin 2010: 148–152). This limitation restricts the focus options in this language, as demonstrated with focus-sensitive particles in (11), whose scope depends on the focus (Kural 1992: 24). The focus particle *yalnızca* 'only' can be placed in the postverbal domain. With neutral intonation, the nuclear stress falls on the preverbal object, which is then interpreted as being in the scope of 'only'. Changing the position of the nuclear stress changes the scope of the focus, e.g., on the subject or on the verb. However, one reading is excluded: the nuclear stress – and correspondingly the scope of 'only' – cannot fall on a postverbal element.

(11) Turkish (Kural 1992: 24)
Ahmet   o   kitab-ı   göster-mış   yalnızca Berna'-y-a.
Ahmet[NOM] this book-ACC show-PST.EVID[3] only   Berna-∅-DAT
'Ahmet must have shown only THIS BOOK to Berna.' (with nuclear stress on *kitab-ı*)
'Only AHMET must have shown this book to Berna.' (with nuclear stress on *Ahmet*)

---

[6]A different result is presented in Abolhasanizadeh et al. (2012) on Persian: tonal events associated with the stress (pitch accents) are tonally compressed but still recognizable after the nuclear stress. However, the conclusions of this study are refuted by Rahmani et al. (2018).

'Ahmet only must have SHOWN this book to Berna.' (with nuclear stress
on *göster-mış*)
'Ahmet must have shown this book only to BERNA.' (not possible)

These examples show that the nuclear stress of the intonation phrase must
be realized within the phonological domain that ends at the right edge of the
verb. We conclude from this fact that the verb determines the right edge of an
intonation phrase. The Turkish examples in Figure 8 illustrate the difference be-
tween V-final utterances in broad focus and non-V-final utterances with postver-
bal background information. In the broad focus realization (Top Panel), the
utterance is realized as a series of rising contours mapped onto the preverbal
constituents. Constituents with final stress (*maymu-'-nu* 'monkey-ACC') are re-
alized with a rising pitch excursion stretching from their left to their right edge
($_\varphi$L...H$_\varphi$); non-final stress (*A'lanya-lı-lar* 'Alanya-NMLZ-PL,' *cu'ma-rtesi* 'Friday-
after,' *ar-ı'yor-muş* 'look-PROG-PST.EVID[3]') is associated with a falling accent
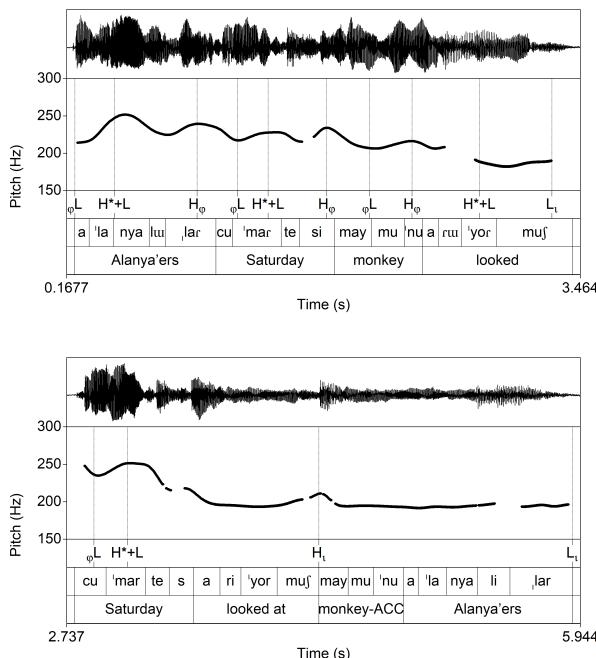
Figure 8: Broad focus vs. background information in Standard Turkish:
Top Panel [SAdvOV]$_F$, answer to 'What happened?'; Bottom Panel:
Adv$_F$VOS, answer to 'When did the Alanya'ers seem to have watched
a monkey?'; morphemic transcription in (12)

(H*+L) (see Kamali 2011: 93, Güneş 2013, Féry 2018: 250–257). When the narrow focus is preverbal, it is realized with a falling pitch accent (H*+L), late aligned with the stressed syllable (see *cu'ma-rtesi* 'Friday-after'in the Bᴏᴛᴛᴏᴍ Pᴀɴᴇʟ). There is no φ-phrase boundary between the focus and the verb, which is shown by the fact that the final /i/ is elided (apocope) to avoid the hiatus. The material following the focus is de-accented, that is, the falling pitch accents in stressed syllables are erased: compare the realization of *A'lanya-lı-lar* 'Alanya-ɴᴍʟᴢ-ᴘʟ' in the Tᴏᴘ and Bᴏᴛᴛᴏᴍ Pᴀɴᴇʟ.

Crucially, the postfocal domain is de-accented, but not fully de-phrased: a (compressed) rise is aligned with the right edge of the verb. The tonal events that are associated with stress (pitch accents) are eliminated, but the events that relate to phrasing are tonally compressed. Since the phonological domain that contains the nuclear stress is an intonation phrase, we assume that the right edge of the predicate is aligned with an ɪ-phrase boundary.

(12) Turkish (D. Balıkçıoğlu and S. Seyis, p.c.)

a. ₀L H*+L   H_φ ₀L H*+L   H_φ ₀L   H_φ
   ((*alanya-lı-lar*_ω   )_φ (*cuma-rtesi*_ω)_φ (*maymun-u*_ω)_φ
   Alanya-ɴᴍʟᴢ-ᴘʟ[ɴᴏᴍ]   Friday-after   monkey-ᴀᴄᴄ

                                L_ɪ
   *ar-ıyor-muş*_ω   )_ɪ
   look-ᴘʀᴏɢ-ᴘsᴛ.ᴇᴠɪᴅ[3]

   'The people from Alanya seem to have watched a monkey on
   Saturday.'

b. ₀L H*+L   H_ɪ
   (((*cuma-rtesi*_ω)_φ *ar-ıyor-muş*_ω   )_ɪ (*maymun-u*_ω
   Friday-after   look-ᴘʀᴏɢ-ᴘsᴛ.ᴇᴠɪᴅ[3]   monkey-ᴀᴄᴄ

                                L_ɪ
   *alanya-lı-lar*_ω)_φ   )_ɪ
   Alanya-ɴᴍʟᴢ-ᴘʟ[ɴᴏᴍ]

   'It was on SATURDAY that the people from Alanya seem to have
   watched a monkey.'

In this section, we demonstrated two prosodic properties of post-predicate elements. For some languages that do not have postverbal foci, it is reported that post-predicate elements are separated by a prosodic boundary from the predicate (Western Armenian). Anecdotal data from Turkish show that the right edge of the predicate may be aligned with a tonally compressed high tone boundary, even if

this tonal event is within the de-accented domain that follows the nuclear stress. Second, in some languages, the post-predicate domain cannot bear the nuclear stress and can only contain information outside the focus domain of the utterance – as demonstrated for Standard Turkish.

7 Outlook

7.1 Left-right asymmetries

The prosodic facts presented in sections 4-6 reveal two major asymmetries between the left and the right side of the predicate in languages with OV properties. The typology must refer to constructions (instead of languages), because these observations do not necessarily generalize for all constructions of a language (see discussion about postverbal objects and obliques in Persian in Section 1).

- Left-Right Asymmetry in Prominence
  In some constructions, the nuclear stress cannot be hosted by elements after the predicate.

- Left-Right Asymmetry in Phrasing
  In some constructions, post-predicate material is prosodically separated from the predicate, while pre-predicate material is not.

The Left-Right Asymmetry in Prominence is consistently reported in earlier research on Standard Turkish (Taylan 1984, İşsever 2003, etc.) and Persian (Karimi 2003, Sadat-Tehrani 2007, to the exception of postverbal goals). The fact that the nuclear stress cannot follow the predicate excludes any element that needs to be accented from this domain: bare objects, focused constituents, *wh*-pronouns in ordinary questions, etc. If these elements do not bear the nuclear stress, they may follow the predicate: bare objects can follow the predicate if the focus is on the verb in Turkish (İşsever 2003: 1047); polar questions with fronted verbs (i.e., nuclear stress on the verb) may have postverbal bare objects in Persian (Karimi 2003: 123).

The limitation on nuclear stress placement indicates that post-predicate elements belong to a phonological domain whose potential nucleus cannot take scope over the core clause. This view does not exclude that an accent follows the predicate (this can be the case in afterthoughts for instance), but a post-predicate accent cannot be the nuclear stress of the intonation phrase encompassing the predicate and the material preceding it. In terms of prosody-syntax mapping, this

phenomenon implies that post-predicate elements are outside the core clause in these constructions, that is, they are right dislocated, as schematically represented in (13).

(13)   ((S V)$_\iota$ O)$_\iota$
       [[S V]$_{core}$ O]$_{clause}$

The Left-Right Asymmetry in Phrasing between the left and the right side of the predicate has been observed in contexts in which the predicate and the post-predicate elements are both part of the same information structural domain (broad focus or out of focus):

- In Old Georgian manuscripts, the verb is often separated by a colon from the post-predicate material; see Section 1 (Boeder 1991);

- In Modern Georgian, the right edge of the verb is often aligned with the boundary of a prosodic constituent, reflected in a high phrase tone that is not within the domain of register lowering of earlier high tones; see Section 4.

- Anecdotal evidence from Turkish shows that when the postnuclear material is not completely de-phrased, the right edge of the predicate is a natural way to split a background domain in prosodic constituents; see Section 6.

- For various languages of the area, it is reported that non-final verbs appear with phonetic cues (tonal events or breaks) indicating a prosodic boundary at their right edge: Western Armenian (Donabedian-Demopoulos 2018) as well as Adyghe, Circassian, and Chechen (summaries in Borise 2021b); see sections 4 and 6.

The prosodic separation of the predicate from post-predicate elements cannot be accounted for by information structure, since in all these cases the predicate and the post-predicate material are part of the same information-structural domain (e.g. a broad focus domain or a domain of background information). There are two possible ways to account for this data.

It may be that this phrasing pattern stems from purely phonological rules that are accidentally correlated with the right edge of the predicate. For instance, it may be that the phrasing of the final element in Georgian SVO clauses is just an indicator of a nuclear stress hosted by the final object. This hypothesis would apply to a part of the data (only broad focus) but is fatally defeated by the prosodic realization in Figure 7/Left Panel, since this example does not involve a major

phrasal boundary between the V and the material following it, while the nuclear φ-phrase does not immediately follow the predicate.

An alternative hypothesis is that the phrasing of the verb reflects the V-final properties of these languages. The post-predicate material is part of the domain that can host the nuclear stress in languages such as Georgian, which means that there is no reason to assume a lower domain for ι-phrases in this case; compare (13) and (14). Consequently, the φ-phrase containing the material preceding the verb can only be accounted for by assuming a syntactic phrase containing the subject and the predicate – to the exclusion of the object. We label this constituent as a projection expanding the verb (V). This conclusion follows from the assumptions introduced so far, but the reasoning behind this constituent structure needs to be supported by independent evidence.

(14)   ((S V)$_\varphi$ (O)$_\varphi$)$_\iota$
       [[S V]$_{V\ expansion}$ O   ]$_{core\ clause}$

7.2 Prosody and focus options of post-predicate elements

The prosodic properties of the post-predicate domain offer a partial explanation of the cross-linguistic differences in focus options in Figure 1. In prosodic view, if post-predicate elements can host the nuclear stress, they can be used for (narrow/broad) focus, as in Georgian, Armenian, Caucasian Urum and probably many Nakh-Daghestanian languages; see Figure 9 and discussion about the languages in Section 2. If post-predicate elements cannot host the nuclear stress,

Can the nuclear stress follow the predicate?

yes   Can unstressed material be part of a broad focus?

{NF, BF, −F}
Georgian, etc.   yes   no

                       {BF, −F}   {−F}
                       Persian, etc.   Turkish, etc.
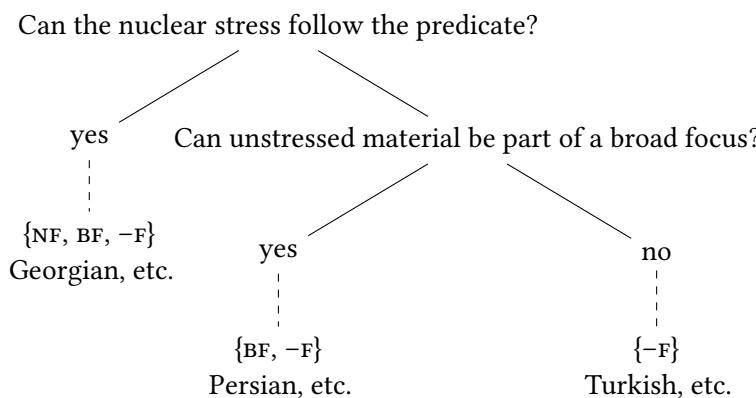
Figure 9: Focus options of post-predicate objects by prosodic properties
(NF: narrow focus; BF: broad focus, −F: out of focus)

then they cannot be narrowly focused, as in Persian, Turkish, Western Armenian. Within this group of languages, there is a further differentiation that cannot be accounted for by prosodic differences. In some languages (and at least at some level of meta-linguistic reflection), it is possible to use in broad focus certain constructions with post-predicate elements that do not bear the nuclear stress (see discussion about Tsakhur and Persian in Section 2). In Persian, this possibility appears with nuclear stress on the verb, which applies to constructions with (preverbal/postverbal) specific objects (see Modarresi 2014: 133–134 on Persian). The nuclear stress must be hosted by the specific object if it is in narrow focus and it is exactly this focus option that is banned from the post-predicate domain. This reasoning accounts for the facts from Persian, but does not offer an explanation why the same flexibility does not hold true in Turkish. The crucial issue is that there is nothing in the prosodic structure that would hinder a specific object from being postverbal under broad focus in these languages. Whether a language uses this construction (as Persian) or not (as Turkish) is an independent distinction that cannot be accounted for by prosody alone.

7.3 Prosody and postverbal narrow foci

Narrow foci come with an asymmetry in phrasing, such that preverbal foci are phrased together with the predicate, while postverbal foci are separated from the predicate with a prosodic break (see Section 5). This asymmetry follows from an independent phonological property of focus in these languages, namely that the left edge of the focus domain is aligned with a prosodic boundary, separating the focus domain from the prefocal material (see Féry 2013 for a typology in these lines). There is no evidence for a difference in the interpretation of preverbal and postverbal foci in languages that allow for both options: e.g., there is no difference in the interpretation of either option in Georgian (Skopeteas & Fanselow 2010).

An interesting question is whether the preference to align the left edge of the focus with the left edge of a prosodic constituent (which applies to all examined languages) correlates with the evolution of a preverbal focus position in these languages. If the underlying principle is to phrase the focus together with the predicate, the prediction is straightforward: languages with an immediately-preverbal focus are expected to insert a prosodic boundary at the left side of the focus (see Georgian, Eastern Armenian and Caucasian Urum in Section 5), while languages with an immediately-postverbal focus are expected to demarcate the focus by means of a prosodic boundary on its right. Languages of the

latter type are less frequent, but at least Zulu – that has been the object of extensive investigation in this respect – confirms this prediction: narrow focus is right adjacent to the predicate and the focus is aligned with a boundary on its right (Cheng & Downing 2012). However, there are various counterexamples to this putative generalization: various languages of India are reported to have preverbal foci and to align the right edge of the focus with a prosodic boundary; see Bengali (Selkirk 2008), Konkani (Féry 2013: 709). This left/right asymmetry suggests a bias towards aligning the left edge of the focus with a prosodic boundary, which is not surprising: postfocal material is de-accented in most languages, which means that prosodic events following the focus are less likely to appear in general.

Abbreviations

3	3rd person	PASS	passive
ACC	accusative	PFV	perfective
AOR	aorist	PL	plural
CAUS	causative	POSS	possessor
DAT	dative	PROG	progressive
DEF	definite	PST	past
DEM	demonstrative	RA	*ra* (Persian)
ERG	ergative	REM	remote
EVID	evidential	S	subject
IO	indirect object	SG	singular
NMLZ	nominalizer	SM	series marker
NOM	nominative		

Acknowledgements

References

Abolhasanizadeh, Vahideh, Mahmood Bijankhan & Carlos Gussenhoven. 2012. The Persian pitch accent and its retention after the focus. *Lingua* 122(13). 1380–1394. DOI: 10.1016/j.lingua.2012.06.002.

Ardali, Mortaza Taheri & Yi Xu. 2012. Phonetic realization of prosodic focus in Persian. *Proceedings of Speech Prosody 2012*. 326–329. DOI: 10.21437/SpeechProsody.2012-83.

Boeder, Winfried. 1991. The phrasing of Old Georgian according to scribal punctuation. In *7th Conference on the Non-Slavic languages of the Soviet Union, Chicago*, 22–24.

Boersma, Paul & David Weenink. 2023. *Praat: Doing phonetics by computer*. computer program. http://www.praat.org.

Borise, Lena. 2019. *Phrasing is key: The syntax and prosody of focus in Georgian*. Harvard University. (Doctoral dissertation).

Borise, Lena. 2021a. Focus projection need not be based on pitch accents: Evidence from Georgian. *Glossa: A journal of general linguistics* 6(1). DOI: 10.16995/glossa.5733.

Borise, Lena. 2021b. Tone and intonation in languages of the Caucasus. In Maria Polinsky (ed.), *The Oxford handbook of languages of the Caucasus*, 756–782. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780190690694.013.31.

Borise, Lena & David Erschler. 2023. Flexible syntax-prosody mapping of intonational phrases in the context of varying verb height. *Phonology* First View. 1–42. DOI: 10.1017/S0952675723000015.

Cheng, Lisa Lai-Shen & Laura J. Downing. 2012. Against FocusP: Arguments from Zulu. In Ivona Kučerová & Ad Neeleman (eds.), *Contrasts and positions in information structure*, 247–266. Cambridge: Cambridge University Press. DOI: 10.1017/cbo9780511740084.012.

Comrie, Bernard. 1984. Some formal properties of focus in Modern Eastern Armenian. *Annual of Armenian Linguistics* 5. 1–21.

Delforooz, Behrooz Barjasteh. 2010. *Discourse features in Balochi of Sistan: Oral narratives*. Uppsala: Uppsala Universitet.

Dolatian, Hossep. 2020. Cyclicity and prosodic misalignment in Armenian stems. *Natural Language and Linguistic Theory* 39(3). 843–886. DOI: 10.1007/s11049-020-09487-7.

Donabedian-Demopoulos, Anaïd. 2018. Middle East and beyond - Western Armenian at the crossroads: A sociolinguistic and typological sketch. In Christiane Bulut (ed.), *Linguistic minorities in Turkey and Turkic-speaking minorities of the periphery*, 89–148. Wiesbaden: Harrassowitz.

Dum-Tragut, Jasmine. 2009. *Armenian: Modern Eastern Armenian*. Amsterdam: Benjamins. DOI: 10.1075/loall.14.26.

Erschler, David. 2012. From preverbal focus to preverbal "left periphery": The Ossetic clause architecture in areal and diachronic perspective. *Lingua* 122(6). 673–699. DOI: 10.1016/j.lingua.2012.01.009.

Féry, Caroline. 2013. Focus as prosodic alignment. *Natural Language and Linguistic Theory* 31(3). 683–734. DOI: 10.1007/s11049-013-9195-7.

Féry, Caroline. 2018. *Intonation and prosodic structure*. Cambridge: Cambridge University Press.

Forker, Diana. 2021. Information structure in languages of the Caucasus. In Maria Polinsky (ed.), *The Oxford handbook of languages of the Caucasus*, 973–1002. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780190690694.013. 28.

Forker, Diana & Oleg Belyaev. 2016. Word order and focus particles in Nakh-Daghestanian languages. In M. M. Jocelyne Fernandez-Vest & Robert D. Van Valin (eds.), *Information structuring of spoken language from a cross-linguistic perspective*, 239–262. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110368758-012.

Göksel, Aslı. 1998. Linearity, focus and the postverbal position in Turkish. In Lars Johanson, Eva A. Csato & Vanessa Locke (eds.), *The Mainz meeting proceedings of the 7th International Conference on Turkish linguistics*, 85–106. Wiesbaden: Harrassowitz.

Göksel, Aslı & A. Sumru Özsoy. 2000. Is there a focus position in Turkish? In Aslı Göksel & Celia Kerslake (eds.), *Studies on Turkish and Turkic languages: Proceedings of the 9th International conference on Turkish linguistics*, 219–228. Wiesbaden: Harrassowitz.

Gosby, Geoffrey. 2016. *Information structure in Georgian*. University of Oxford. (Doctoral dissertation).

Güneş, Güliz. 2013. On the role of prosodic constituency in Turkish. In Umut Özge (ed.), *Proceedings of the 8th Workshop on Altaic formal linguistics*, 115–128. Cambridge: MITWPL.

Haig, Geoffrey, Paul M. Noorlander & Nils N. Schiborr. In press. Which word order features are stable in a contact setting? Corpus-based evidence from the Western Asian transition zone. In Jeroen Darquennes, Joe Salmons & Wim Vandenbussche (eds.), *Language contact: An international handbook*, vol. 2. Berlin: De Gruyter Mouton.

Haig, Geoffrey & Mohammad Rasekh-Mahand. 2019. *Post-predicate elements in Iranian and neigbouring languages: Inheritance, contact, and information struc-*

ture. *Position paper, for the project on post-predicate constituents in Iranian and neighbouring languages.* (10 July, 2023).

Hasan, Aveen Mohammed. 2012. *Kurdish intonation with reference to English.* Coleraine: University of Ulster. (Doctoral dissertation).

Hosseini, Seyed Ayat. 2014. *The phonology and phonetics of prosodic prominence in Persian.* Tokyo: University of Tokyo. (Doctoral dissertation).

Ipek, Canan. 2011. Phonetic realization of focus with no on-focus pitch range expansion in Turkish. In *Proceedings of the 17th International Congress of phonetic sciences (ICPhS XVII)*, 140–143.

İşsever, Selçuk. 2003. Information structure in Turkish: The word order–prosody interface. *Lingua* 113(11). 1025–1053. DOI: 10.1016/s0024-3841(03)00012-3.

Kahnemuyipour, Arsalan. 2001. On Wh-questions in Persian. *Canadian Journal of Linguistics/Revue canadienne de linguistique* 46(1-2). 41–61. DOI: 10.1017/s000841310001793x.

Kahnemuyipour, Arsalan. 2003. Syntactic categories and Persian stress. *Natural Language and Linguistic Theory* 21(2). 333–379.

Kahnemuyipour, Arsalan & Karine Megerdoomian. 2011. Second position clitics and the vP phase: The case of the Armenian auxiliary. *Linguistic Inquiry* 42(1). 152–162. DOI: 10.1162/ling_a_00033.

Kamali, Beste. 2011. *Topics at the PF interface of Turkish.* Cambridge, MA: Harvard University. (Doctoral dissertation).

Karimi, Simin. 2003. On object positions, specificity, and scrambling in Persian. In Simin Karimi (ed.), *Word order and scrambling*, 91–124. Malden: Blackwell. DOI: 10.1002/9780470758403.ch5.

Kilicaslan, Yilmaz. 2004. Syntax of information structure in Turkish. *Linguistics* 42(4). 717–765. DOI: 10.1515/ling.2004.024.

Komen, Erwin R. 2013. *Finding focus: A study of the historical development of focus in English.* Utrecht: LOT.

Kural, Murat. 1992. Properties of scrambling in Turkish. Manuskript, UCLA.

Lacroix, René. 2019. Laz. In Geoffrey Haig & Geoffrey Khan (eds.), *The languages and linguistics of Western Asia: An areal perspective*, 830–860. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110421682-021.

Ladd, D. Robert. 1986. Intonational phrasing: The case for recursive prosodic structure. *Phonology Yearbook* 3. 311–340.

Ladd, D. Robert. 1988. Declination 'reset' and the hierarchical organization of utterances. *Journal of the Acoustical Society of America* 84(2). 530–544.

Mahjani, Behzad. 2003. *An instrumental study of prosodic features and intonation in Modern Farsi (Persian).* Edinburgh: University of Edinburgh. (MA thesis).

Majidi, Setareh & Mohammad Dabir Moghaddam. 2012. Information structure in Persian: A comparison of systemic functional grammar and role and reference grammar. *Theory and Practice in Language Studies* 2(4). 856–863. DOI: 10.4304/tpls.2.4.856-863.

Marr, Nikolaj. 1904. *Fisiolog*. Saint Petersburg: Imperial University of Saint Petersburg.

Modarresi, Fereshteh. 2014. *Bare nouns in Persian: Interpretation, grammar, and prosody*. Ottawa: University of Ottawa. (Doctoral dissertation). DOI: 10.20381/RUOR-3768.

Özge, Umut & Cem Bozsahin. 2010. Intonation in the grammar of Turkish. *Lingua* 120. 132–175.

Pierrehumbert, Janet B. 1980. *The phonology and phonetics of English intonation*. New York: Garland Press. (Doctoral dissertation).

Rahmani, Hamed, Toni Rietveld & Carlos Gussenhoven. 2018. Post-focal and factive deaccentuation in Persian. *Glossa: A journal of general linguistics* 3(1). DOI: 10.5334/gjgl.328.

Rasekh-Mahand, Mohammad, Elham Izadi, Mehdi Parizadeh, Geoffrey Haig & Nils Schiborr. 2024. Post-predicate elements in modern colloquial Persian: A multifactorial analysis. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 197–230. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266343.

Roberts, John R. 2009. *A study of Persian discourse structure*. Uppsala: Upsala Universitet.

Sadat-Tehrani, Nima. 2007. *The intonational grammar of Persian*. Winnipeg: University of Manitoba. (Doctoral dissertation).

Sadat-Tehrani, Nima. 2009. The alignment of L+H* pitch accents in Persian intonation. *Journal of the International Phonetic Association* 39(2). 205–230. DOI: 10.1017/s0025100309003892.

Sadat-Tehrani, Nima. 2017. Intonation of Persian declaratives: Read vs. spontaneous speech. *Questions and Answers in Linguistics* 4(1). 21–43. DOI: 10.1515/qal-2017-0002.

Samvelian, Pollet, Pegah Faghiri & Victoria Khurshudyan. 2023. On the persistence of SVO: The case of Modern Eastern Armenian. *Linguistics* 61(3). 459–497. DOI: 10.1515/ling-2021-0013.

Schröter, Stefanie. 2017. *Information structure in Caucasian Urum: An empirical investigation of the interaction of focus and topic with word order*. Bielefeld: Bielefeld University. (Doctoral dissertation).

Selkirk, Elisabeth. 2008. Bengali intonation revisited: An optimality theoretic analysis in which FOCUS stress prominence drives FOCUS phrasing. In Chung-min Lee, Matthew Gordon & Daniel Büring (eds.), *Topic and focus*, 215–244. Dordrecht: Springer. DOI: 10.1007/978-1-4020-4796-1_12.

Selkirk, Elisabeth. 2011. The syntax-phonology interface. In John Goldsmith, Jason Riggle & Alan C. L. Yu (eds.), *The handbook of phonological theory, second edition*, 435–484. Malden, MA: Wiley. DOI: 10.1002/9781444343069.ch14.

Skopeteas, Stavros. 2016. The Caucasian Urums and the Urum language. In Süer Eker & Ülkü Çelik Şavk (eds.), *Handbook of endangered Turkic languages, vol. ii*, 527–552. Ankara/Astana: Uluslararası Türk Akademisi.

Skopeteas, Stavros & Gisbert Fanselow. 2010. Focus in Georgian and the expression of contrast. *Lingua* 120(6). 1370–1391. DOI: 10.1016/j.lingua.2008.10.012.

Skopeteas, Stavros & Caroline Féry. 2010. Focus and intonation in Georgian: Constituent structure and prosodic realization. Bielefeld University and University of Frankfurt.

Skopeteas, Stavros, Caroline Féry & Rusudan Asatiani. 2009. Word order and intonation in Georgian. *Lingua* 119(1). 102–127. DOI: 10.1016/j.lingua.2008.09.001.

Skopeteas, Stavros, Caroline Féry & Rusudan Asatiani. 2018. Prosodic separation of postverbal material in Georgian. In Evangelia Adamou, Katharina Haude & Martine Vanhove (eds.), *Information structure in lesser-described languages: Studies in prosody and syntax*, 17–50. Benjamins. DOI: 10.1075/slcs.199.02sko.

Stilo, Donald. 2015. An introduction to the Atlas of the Araxes-Iran linguistic area. In Bernard Comrie & Lucía Golluscio (eds.), *Language contact and documentation/Contact lingüistico y documentación*, 343–356. Berlin: De Gruyter Mouton.

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

Stokhof, Martin & Michiel van Lambalgen. 2011. Abstractions and idealisations: The construction of modern linguistics. *Theoretical Linguistics* 37(1-2). 1–26. DOI: 10.1515/thli.2011.001.

Taylan, Eser Erguvanlı. 1984. *The function of word order in Turkish grammar*. Berkeley: University of California Press.

Testelets, Yakov G. 1999. Porjadok slov [Word order]. In Aleksandr E. Kibrik (ed.), *Èlementy caxurskogo jazyka v tipologičeskom osveščenii [Elements of Tsakhur in typological perspective]*, 293–313. Moscow: Nasledie.

Toparlak, Tabita. 2019. *Etudes phonétiques en arménien*. Paris: Université Sorbonne Nouvelle, Paris 3. (MA thesis).

Vicenik, Chad & Sun-Ah Jun. 2014. An autosegmental-metrical analysis of Georgian intonation. In Sun-Ah Jun (ed.), *Prosodic typology II*, 154–186. Oxford: Oxford University. DOI: 10.1093/acprof:oso/9780199567300.003.0006.

Part II

Case studies from Iranian and Indo-Aryan languages

Chapter 4

Balochi: A cross-dialect investigation of post-verbal elements

Maryam Nourzaei[a] & ⓘ Geoffrey Haig[b]

[a]Uppsala University [b]University of Bamberg

> This chapter investigates word order in three Balochi varieties: Coastal Balochi (Coastal), Koroshi Balochi (Koroshi) and Turkmenistan Balochi (Turkmen). Although all three are closely related, they are areally widely dispersed, making Balochi an interesting test case for investigating the effects of areality on closely related varieties. All three varieties are predominantly OV. However, pronominal direct objects show a stronger tendency to post-verbal placement, especially in Coastal, echoing similar findings for other Iranian languages of the region. All three varieties exhibit predominantly post-verbal Goals (VG), with the highest values found in Koroshi, confirming the expected correlation between higher frequency of VG and geographic proximity to the southwestern Mesopotamian region of the Western Asian Transition Zone.

1 Introduction: Language background and data sources

Balochi is a northwestern Iranian language, which belongs to the Indo-Iranian branch of Indo-European. Syntactically, Balochi is OV, but shows mixed adpositional typology (see Section 3.2) as well as dialectally differentiated alignment systems (see below). Word order in the NP is generally head-final: adjectives precede nouns, and take an attributive linker *-ēn/ẽ̄*. Possessors also precede the possessed, and the possessor takes a so-called genitive case, though this may vary in the westernmost dialects; see Section 3 for details.

Balochi provides an excellent window on the interplay of areal and genetic influence in shaping word order. The unity of 'Balochi' as the descendants of a

historically reconstructable sub-group of Iranian is justified in Korn (2005: 21), and there is no doubt that the three doculects under consideration here can be assigned to Balochi. Today, however, due to successive population movements, varieties of Balochi are spoken across a vast area, including Southeastern and Southwestern Iran, Southwestern Pakistan, and also in Afghanistan, India, Africa, Turkmenistan, Oman and the UAE. Almost all Balochi speakers are multilingual, with contact languages belonging to four different language families, and different genera within them: Indo-European (Iranian, Indo-Aryan and Slavic), Dravidian, Turkic and Semitic. Disentangling what is inherited Proto-Balochi from the multiple contact effects is methodologically challenging, for syntax just as it is for phonology and morphology. We address the implications of our study for broader questions of Iranian diachronic syntax in Section 5.

Research on Balochi recognizes three main dialects: southern, eastern and western Balochi. Each of these dialects demonstrates its own sub-divisions (see Jahani & Korn 2009: 636–637). In addition, a group of dialects to the southwest is distinguished, collectively referred to as Koroshi. The total number of Balochi speakers is uncertain, though, Jahani (2013) reports an estimate of at least 10 million speakers. For the comparisons undertaken in this chapter, we have selected data from three geographically dispersed locations, each of which lies within some larger dialect region of Balochi. We refer to the data from these locations as doculects (the variety documented in a specific data set), rather than dialects, because we cannot assume that each doculect is necessarily representative of its larger dialect region. The three doculects are labelled Turkmen (Turkmenistan Balochi), Coastal Balochi (Coastal Balochi), and Koroshi (Koroshi Balochi) respectively. The data for Coastal Balochi come from two villages, Korsar and Sedighzahi, Dashtiyari, Iran. For Koroshi, data comes from from Deh Piyaleh, Shiraz, and Marvdasht, Iran, while for Turkmen the data come from the Mari region, Turkmenistan. The location of the three doculects, and the main dialect divisions of Balochi, are provided in Fig. 1.

Turkmen belongs to the larger Western Balochi dialect group, Coastal Balochi dialect belongs to the Southern dialect group, and Koroshi is part of Koroshi, but would probably also be considered part of Southern Balochi in a broad dialect division (see also Nourzaei et al. 2015: 22).

It is important to bear in mind that our doculects are not necessarily representative for the entire dialect group to which they belong, and in fact our results suggest that there is considerable internal variation (at least on the word order parameters that we have investigated) within the larger dialect groups that have traditionally been recognized. For each doculect, a WOWA data set was compiled as the basis for quantitative comparison. Text types are traditional narrative texts for Coastal Balochi and Koroshi, while the texts for Turkmen include
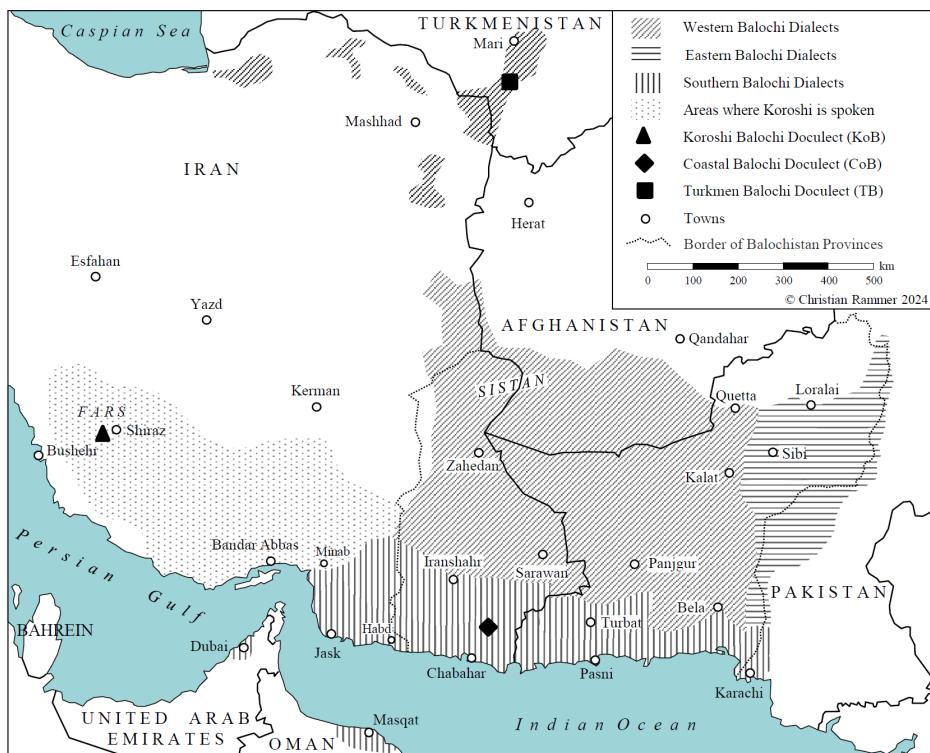
Figure 1: Location of the main Balochi dialects and doculects analysed

traditional narratives, a procedural text, and a life story; see Nourzaei (2021a,b) and Haig (2022a) for details and access to the data. In citing examples we follow the original transcriptions for Coastal Balochi and Koroshi but adapt the original transcriptions of Turkmen to bring it closer to the other two.

The doculects vary according to their alignment systems: Turkmen and Koroshi display the same alignment system in both past and present domain, while Coastal Balochi demonstrates ergative only for 3rd person in transitive clauses based on past-tense verb forms. Koroshi, due to reduction of its morphological case system, uses more person clitics than Turkmen and Coastal Balochi to index verb arguments. The most recent studies on Coastal Balochi are Nourzaei (2017, In press) and Korn & Nourzaei (2019). Earlier works on Koroshi include a grammatical sketch in Persian (Emādi 2005), Nourzaei et al. (2015), Nourzaei (2017, In press). It is assumed that speakers of Koroshi originally migrated from Balochistan to the Southwest of Iran. The most significant study on Turkmen dialect is Axenov (2006). Turkmen Baloch is the result of migration from either

Iran or Afghanistan and north-eastern parts of Iran. For the history of the Baloch migration to Turkmenistan, see Axenov (2000: 71), among others. In addition to the quantitative analysis based on the WOWA data sets, we also draw on additional published corpora and analysis (Barjasteh Delforooz 2010, Nourzaei et al. 2015, Nourzaei 2017, In press, Axenov 2006).

This chapter is organized as follows. Section 2 summarizes previous literature, while Section 3 provides a sketch of some relevant aspects of Balochi grammar, including NP-internal word order. In Section 4, we present the core of the quantitative analysis, based on the major clausal constituent types recognized in the WOWA framework, while Section 5 discusses the implications of the Balochi findings from a comparative (West) Iranian perspective.

2 Previous studies on word order in Balochi

With regard to clause-level constituents, empirical comparative research only commenced quite recently, in particular Jahani (2018) and Korn (2022). Jahani (2018) covers four dialects, viz., Koroshi, Sistani, Turkmen and Southern Balochi (Iran and Pakistan), and considers four argument types: goals, recipients, addressees, and final states of change-of-state predicates. She notes that across all varieties, goals of motion tend to be postverbal, while with regard to the other roles, there are differences: While Koroshi generally has all four argument types post-verbally, Sistani and Turkmenistan Balochi has goals, recipients, and addressees, but not final-states in post-verbal position. For Southern Balochi, also Jahani (2018) includes a more detailed study of spoken versus written genres, observing a very significant tendency towards strict pre-verbal placement of all argument types in written Southern Balochi, while spoken Southern Balochi places 90% of goals after the verb.[1] The differences between oral and written language are very much in line with the observations for Persian, where colloquial spoken Persian has a high frequency of post-verbal Goals (around 80%, see Rasekh-Mahand et al. 2024 [this volume]), while the frequency is much lower for formal written Persian. These pioneering observations suggest that the overall situation in Balochi shows some similarities with the situation that has been identified for languages further to the west (e.g. Kurdish, Haig 2022c), in particular the special status of goals (see Haig et al. 2024 [this volume] for summary), but there are also significant differences. We turn to explanations in Section 5 below.

---

[1]The data from Jahani (2018) are taken from the slides available from the conference talk. For Southern Balochi, the data for recipients on the overview slide do not match the findings reported from the oral vs. written case study, so we make no claim regarding recipients in Southern Balochi.

Korn (2022: 114) draws on data from Afro-Baloch varieties, taken from Nourzaei and Korn's unpublished Afro-Baloch data. Specifically, the data come from three locations: Shirgowaz (close to Bahukalat), Konark and Karewan. As such, they are only indirectly comparable to our Coastal Balochi doculect, although both Korn's data and our own Coastal Balochi doculect can be included under the broad umbrella term of Southern Balochi. Korn's data confirm the strong tendency to place goals post-verbally, though she suggests that goals of caused motion are overall less likely to be post-verbal, a finding which is confirmed in our own Coastal Balochi data in Section 4 below. Korn suggests that many of the caused goals in her data are human, making them overall more similar to recipients (Korn 2022: 104). Recipients are commonly post-verbal, especially when the recipient or benefactive is "mentioned for the first time "(Korn 2022: 105), but pre-verbal position is also well-attested, so that no clear conclusion can be reached with regard to a canonical or unmarked position; rather, the interaction of various factors such as animacy, verb semantics, overall presence of co-arguments, and information structure co-determine the placement. For addressees, only few overt examples were present in the data, but all occur preverbally (Korn 2022: 107). This finding confirms Jahani's (2018) observation that addressees are overall more likely to be preverbal than either goals or recipients. Again, this confirms observations from other languages in WATZ, according to which addressees are less likely to be post-verbal than recipients or goals (see Haig et al. 2024 [this volume], Section 4).

Korn also investigates direct objects, noting that post-verbal placement "is surprisingly frequent" (Korn 2022: 113), though no figures are provided; we return to this below. Korn also discusses the possible impact of flagging, and animacy, noting a tendency for [+human] arguments to be pre-verbal. She concludes with a suggestion for the pathway towards "generalization of the post-verbal position" in Balochi (Korn 2022: 118), according to which non-human goals would have been the first kinds of constituent in this position, extending then to include metaphorical uses of direction (purpose) and then other kinds of adverbial, while a second line of extension would proceed via [+human] goals to recipients and beneficiaries. Ultimately, an extension to direct objects is considered as a final possibility. Korn (2022) also discusses the impact of flagging, verb serialization, and interactions with subordination. In sum, Jahani (2018) and Korn (2022) provide a very informative overview of the relevant characteristics of Balochi, which already identifies some of the areas of cross-dialect variation. The present study builds on these observations but extends the the range of argument types considered, and is based on more accessible data.

3 Basic features of Balochi: Morphosyntax and NP-internal word order

3.1 Alignment, person marking, and nominal inflection

Among the three doculects considered here, only Coastal Balochi displays stem-sensitive ergativity, i.e. verbal forms with the present stem pattern accusatively (A in direct case and agreeing with the verb, P in oblique or object case) while verbal forms with the past stem pattern ergatively (P in direct case and agreeing with the verb (only for third person), A in oblique case) (see for details Nourzaei 2017). Both Koroshi and Turkmen Balochi exhibit only accusative alignment.

Doculects differ regarding the usage of person-marking clitics. Both Coastal Balochi and Turkmen only use the person marking clitics for the 3rd person, while Koroshi uses the person-marking clitics for all persons. The existence of person-marking clitics in Balochi has a strong correlation with the case system. In varieties such as Coastal Balochi and Turkmen (see above) which have a rich morphological case system, they are less commonly used, while in varieties such as Koroshi and Sarawani (see Baranzehi 2003: 86), which display a reduced case system, they are more common.

Balochi has a morphological case system containing at least a direct case, an oblique case, and a genitive case, which is the system found in Koroshi. In Sistani, Afghan and Turkmen Balochi, a locative case has developed from the genitive (cf. Buddruss 1988: 48, Axenov 2006: 80–82, Korn 2008, and the data in Barjasteh Delforooz 2010). In Coastal Balochi, it is sporadically attested as well, typically with human names (Nourzaei 2017: 61, Korn & Nourzaei 2019). In addition, another form appears in Sistani Balochi that is derived from the oblique case and is only used to mark direct objects, whence its name, "object case" (cf. Korn 2008: 61–63, Nourzaei 2017: 62). In Coastal Balochi and Koroshi doculects, object case is only available for pronouns (cf. Nourzaei 2017: 44).

With respect to plural marking, Koroshi differs from other varieties in that it has an innovated plural marker *-obār* in all cases (direct, genitive and oblique, Nourzaei et al. 2015: 29) while in Coastal Balochi and Sistani, the marker *-ān* does not mark the plural on nouns in the direct case, which are thus not directly marked for number. Instead, plural number can be indicated by plural agreement markers on the verb. Nouns in the oblique case, on the other hand, show a number opposition. No variety of Balochi has retained grammatical gender.

3.2 NP-internal word order

3.2.1 Adjective and noun

The most common pattern regarding adjective-noun ordering across the dialects is that adjectives precede nouns, and that attributive adjectives take a suffix *-ēn/ẽ̂* (adjective attribute suffix, ATTR).

(1)   a.   Coastal Balochi (Nourzaei 2021a, C, 1171)
           zar'd-ẽ̂   *negē'na*
           yellow-ATTR stone
           'the yellow stone'
      b.   Turkmen Balochi (Haig 2022a, A, 0049)
           ǰwān-ēn   *zāg=ē*
           good-ATTR son=INDV
           'a good son'
      c.   Koroshi Balochi (Nourzaei et al. 2015:42)
           bōr-ēn   *pašm-ā*
           beige-ATTR wool-OBL
           'the beige wool'

However, Koroshi exhibits borrowed ezafe constructions (N-ADJ order) from Persian shown in (2).

(2)   a.   Koroshi Balochi (Nourzaei 2021b, A, 0006)
           ādam=e   ***xūb=ī***
           person=EZ good=INDV
           'a good person'
      b.   Koroshi Balochi (UP)
           ǰāhel=e *nūrānī=ye*
           boy=EZ handsome=INDV
           'a handsome boy'

3.2.2 Possessor and noun

Across the dialects, possessors normally precede the possessed noun (POSS-N order), and the possessor takes a so-called genitive case (for details regarding different forms of genitive case (see Nourzaei 2017, Nourzaei et al. 2015, Korn & Nourzaei 2019).

(3)   a. Coastal Balochi (Nourzaei 2021a, A, 0315)
> ***sī'morg-e***   *dap-ā*
> fabulous_bird-GEN mouth-OBL
> '(into) the fabulous bird's mouth'

      b. Turkmen Balochi (Haig 2022a, A, 0083)
> ***xānbādorr-ī***   *ǰenēn-ā*
> Khanbadur-GEN wife-OBL
> 'the wife of Khanbadur'

      c. Koroshi Balochi (Nourzaei 2021b A, 0007)
> ***šāh-ay*** *awal-īn*   *bač*
> king-GEN first-ATTR son
> 'the king's first son'

In Koroshi, possessors may follow nouns (N-POSS) in a kind of ezafe construction, most likely borrowed from Persian. Most examples in Nourzaei et al. (2015) suggest not so much morphological borrowing as wholesale borrowing of phrases from Persian, as in (4). The actual productivity of N-EZ POSS constructions in Koroshi thus remains to be established.

(4)   Koroshi Balochi (Nourzaei et al. 2015:212)
> *ya* *banne=ye*   *xodā=ī*
> one servant=EZ god=INDV
> 'a fellow (lit. a servant) of God'

The ordering of pronominal possessors may differ from nominal possessors. In Coastal Balochi, pronominal possessors generally follow the possessed:

(5)   a. Coastal Balochi (Nourzaei 2021a, B, 505)
> *gohār-ā*   ***otīg-a***
> sister-OBL REFL-OBL
> 'your sister'

      b. Coastal Balochi (UP)
> *māt*   *manīg*
> mother mine
> 'my mother'

3.2.3 Demonstrative and noun

Across all dialects, demonstratives precede nouns, with some marginal exceptions which are ignored here (Korn & Nourzaei 2019).

(6)   a.   Coastal Balochi (Nourzaei 2021a, B, 0711)
          ē   pet
          PROX father
          'this father'

      b.   Turkmen Balochi (Haig 2022a, A, 0030)
          ē   bādešā
          PROX king
          'this king'

      c.   Koroshi Balochi (Nourzaei 2021b, B, 0570)
          ē   ǰo'ġlā
          PROX boy.OBL
          'this boy'

3.2.4 Numeral and noun

Plural marking retains to a large extent the archaic pattern, also found in Kurdish, whereby only nouns in oblique cases are overtly plural marked with a suffix, but plural subject nouns are not overtly marked. All dialects share the commonality of ordering the numerals before head nouns.

3.2.5 Adpositions

The dialects have prepositions, postpositions, and circumpositions. Postpositions are generally relational nouns in the oblique case, with the NP complement in the genitive case. Similar to nouns in the oblique case, the adpositions can also be used alone as adverbs. Prepositions usually trigger the oblique case of the noun. Note that there is a tendency for losing the oblique case after prepositions in Koroshi. The respective frequencies of prepositional and postpositional use shows interesting cross-dialectal variation, which we sum up in Table 1 below. In general, Koroshi has the strongest tendency to use prepositions (see for more details Nourzaei et al. 2015: 43–46). Examples of prepositions and postpositions follow:

(7)   a.   Coastal Balochi (Nourzaei 2021a, B, 0678)
          čāt-e   tōkā
          well-GEN inside
          'inside the well'

b. Turkmen Balochi (Haig 2022a, D, 0576)
  gis-ay **tā**
  house-GEN inside
  'inside the house'

c. Koroshi Balochi (Nourzaei et al. 2015:36)
  dawr=e *ī mēdag-ā*
  around=EZ PROX encampment-OBL
  'around this encampment'

d. Koroshi Balochi (Nourzaei 2021b, B, 0553)
  čāh-ay **tōxā** *por=e šamšīr=o nayza a=kan-t*
  well-GEN in full=EZ sword=and spear VCL=do.PRS-3SG
  'She fills the well with swords and spears.'

In contrast to Koroshi, Coastal Balochi and Turkmen possess circumpositions:

(8) Turkmen Balochi (Axenov 2006:150, glosses follow the source)
  bi *diga gis-ē tā*
  to other house-IND inside
  'to another house'

For our quantitative analysis of adpositions in actual usage, based on the WOWA data, we will consider the respective frequencies of prepositional and postpositional flagging, across those functions that we would generally expect to favour adpositional over other types of flagging (e.g. case marking, or lack of any overt flagging).[2] Table 1 shows the respective frequencies of postpositional and prepositional flagging across the three dialects for NPs in these functions.

Table 1 suggests a two-way split across Balochi, between the predominantly prepositional Koroshi and Turkmen on the one hand, versus predominantly postpositional Coastal Balochi on the other. The high frequency of postpositional flagging in Coastal Balochi is intriguing. At this point we have no convincing explanation. It may be a retention of earlier Balochi structures, which has been lost in other varieties through greater contact with other west Iranian languages,

---

[2]The procedure was as follows: Taking the three Balochi WOWA data sets, we selected the total number of tokens in the following functions: ABL(ative); ADDR(essee); BEN(efactive); COM(itative); GOAL; GOAL-C(aused); INSTR(umental); LOC(ative); REC(ipient); REC-BEN. We then extracted those that were flagged with preposition or pre-nominal relational nouns (lumped together as "prepositional"), and those that were flagged with postpositions, or post-nominal relational nouns (lumped together as "postpositional"). Together this yielded 343 tokens.

Table 1: Balochi prepositional and postpositional flagging frequency

Adpositional type	Coastal		Koroshi		Turkmen		Totals
	N	%	N	%	N	%	
Postpositional	83	86	2	2	12	10	97
Prepositional	13	14	129	98	104	90	246
Totals	96		131		116		343

notably Persian. It may be connected to multilingualism with Urdu and other postpositional Indo-Aryan languages spoken in the region, but this is speculative. Interestingly, in our data Turkmen is dominant prepositional, although we might have expected higher rates of postpositions due to contact with Turkic. It is possible that Turkic has had less influence because the migrations of Turkmen speakers to Turkmenistan was relatively recent, and they retain contact with Sistani Balochi speakers in Iran. This would be in line with the findings of Haig et al. (In press), according to which adpositional type is a relatively conservative word order parameter that only shifts under intense and long contact.

3.3 Auxiliary and main verb, complement clause and matrix clause

TAM categories are expressed by the presence or absence of verbal prefixes (Jahani & Korn 2009, Axenov 2006, Nourzaei et al. 2015 among others). The perfect system uses the participle followed by the inflected copula, which cliticizes to the verb, while the progressive (e.g., Koroshi and Coastal Balochi) is built from infinitive plus clitic copula. Historically, then, we can assume at least some examples of V-Aux order, which have since univerbated through cliticization of the original auxiliary. In contemporary Balochi, however, prosodically independent (non-clitic) auxiliary verbs are infrequent, so establishing a regular order of auxiliary and verb is not straightforward. Examples with cliticized auxiliaries are the following:

(9)   a.   Turkmen Balochi (Haig 2022a, D, 0548)
      ammā pa wat-ī   māl-ān   yakk yakk=ī   nām
      1PL   for REFL-GEN animal-PL one   one=INDV name
      išt=at-an
      put.PST=COP.PST-1PL
      'We had given names to everyone of our sheep.'

b. Koroshi Balochi (Nourzaei 2021b, A, 0296)
*hasan kačal faġat **nay-āk-ag=en***
Hasan bald   only NEG-come.PST-PP=COP.3SG
'Only Hasan the Bald has not come along.'

The Sistani variety of Western Balochi has an auxiliary verb *dāšten* 'have', a recent borrowed element from Persian, to build progressive construction in past and present domain (see for details Nourzaei 2024). As in Persian, this auxiliary precedes the main verb as in (10):

(10)   a.   Sistani Balochi (Nourzaei 2024)
dār-īn   *wān-īn*   *sāket be*
have.PRS-1SG read.PRS-1SG quiet IMP.become.PRS.2SG
'Be quiet, I am studying!'
b.   Sistani Balochi (Nourzaei 2024)
dāšt-on   *šot-on*   *ke čākar āt*
have.PST-1SG go.PST-1SG CLM Chakar come.PST.3SG
'I was going when Chakar came.'

The subordinator *ke* may introduce various kinds of subordinate clause, i.e, relative, complement and adverbial, as well as quoted speech. Across the dialects, the complement clauses normally follow the main clause and are either linked to it by asyndetic subordination (juxtaposition) without any overt marker of subordination other than rising intonation, or with an overt complementizer. A number of compound conjunctions, composed of nouns or other elements plus *ke*, such as *mawġeī ke/waġteke* 'when', and *be šartī ke* 'on the condition that' are also used. Additional subordinating conjunctions include *tā/ta* 'until, so that' and *aga/aya* 'if'. In all dialects subordination closely follows the basic pattern of Persian and copies its compound conjunctions (see also Jahani & Korn 2009: 678). Examples of complement clauses with verbs of speech and perception, with and without complementizers follow:

(11)   a.   Coastal Balochi (Nourzaei 2021a, C, 1061)
pet-ā   ***go***   ***ke*** *man-ī čo   nī 'mã 'ta-rā 'sīr*
father-OBL   say.PST CLM 1SG-GEN child now 1SG 2SG-OBJ wedding
da'y-ã̃
give.PRS-1SG
'The father said /that/, "My son, now I will marry you off." '

b. Turkmen Balochi (Axenov 2006)

gis-ay   **_wāond gušt=ī_**   **_ke_** _mnī_   _piss_   _iškārī=e_
house-GEN owner   say.PST=PC.3SG CLM 1SG.GEN father hunter=INDV

_at-ī_
COP.PST-3SG

'The owner of the house said that my father was a hunter.'

c. Koroshi Balochi (Nourzaei et al. 2015:143)

a=genn-an   **_bale_** _aždahā=am_ _pīk-ay_
VCL=see.PRS-3PL yes   dragon=ADD twist.PST.PP-COP.PRS.3SG

dawr=e   _šāh-ay_   _ǰanek-ay_   _garden-ā_
around=EZ king-GEN daughter-GEN neck-OBL

'They see [that] indeed the dragon was wrapped around the neck of the king's daughter.'

4 Order of clause-level constituents: A quantitative analysis

In this section we present the results of the quantitative analysis, drawing on the set of constituent types defined in the WOWA framework; not all categories are considered, as some have too few tokens for meaningful quantitative analysis.

4.1 Direct object and verb

Across the dialects, nominal direct objects are overwhelmingly in the preverbal position (>90% OV in all three WOWA Balochi doculects). This confirms previous research on Balochi (Jahani & Korn 2009 and Korn 2022), and also reflects the overall tendency for Iranian languages in WOWA to be consistently OV in discourse, with the exception of Kumzari (see Haig 2022b, Haig et al. 2024 [this volume]). Our analysis does, however, identify some cross-dialectal differences. Table 2 provides the figures for direct objects, distinguishing between pronominal direct objects (12a) and nominal direct objects (12b). WH-words, clitic pronouns, and adverbials in object function have been excluded from the pronoun category, but we include reflexives. In Table 2, and in the following Tables, N refers to the absolute number of tokens included in each category, "Po" refers to the number of those tokens that were post-verbal, and "%" provides the percentage of post-verbal tokens in each category.

The absolute number of pronominal objects in the data is quite low, particularly in Koroshi, and those that are present are overwhelmingly human (82%, 142

Table 2: Nominal vs. pronominal post-verbal direct object frequencies

	Coastal			Koroshi			Turkmen			Totals
	N	Po	%	N	Po	%	N	Po	%	
Nominal	339	23	6.8	182	4	2.2	193	3	1.6	714
Pronominal	98	27	27.6	20	0	0	55	2	3.6	173
Totals	437			202			248			887

out of 173); we turn to the interplay of humanness and pronominality in Table 5 below. Examples of pronominal and nominal direct objects respectively are the following:

(12)   a.   Turkmen Balochi (Haig 2022a, A, 0007)
       *annūn b-raw-an   ke   **šmā-rā** gis=a   da-īn*
       now   SBJV-go.PRS-1PL CLM 2PL-OBJ house=VCL give.PRS-1SG
       'Now let us go, I will marry you off.'

   b.   Koroshi Balochi (Nourzaei 2021b, A, 0022)
       *man 'wad=om   **as'p-ok-ā**   'gott a=kan-ān*
       1SG   REFL=PC.1SG horse-DEF-OBL raise VCL=do.PRS-1SG
       'I myself will raise the horse.'

With regard to nominal objects, the three dialects exhibit <10% levels of post-verbal placement. However, in Coastal Balochi, rates of post-verbal nominal objects are more than double the other two dialects. A pair-wise Fisher's Exact Test indicates that the difference between Coastal Balochi and the other two reach significance (p<0.05). For pronominal objects, the differences are much more pronounced, and again, it is Coastal Balochi that diverges from the other two.

Korn (2022: 112) had already noted that postverbal direct objects "are surprisingly frequent" in the geographically close variety of Southern Balochi that she investigates. Our data suggest that the spoken varieties of Southern Balochi, such as our "Coastal Balochi," may indeed differ from other varieties of Balochi, in particular with regard to pronominal objects. Other research (Stilo 2018, Haig et al. In press) has suggested that pronominal objects are the most mobile, in the sense that they are more likely to shift across the predicate from the canonical object position, and our findings provide further support for this assumption.

As Korn (2022: 113) notes, post-posing direct objects is probably related to information structure, but the exact nature of the triggering factors is "not always

obvious." The WOWA data base does provide a rough classification of direct objects into definite and indefinite, which we have analysed in Table 3.

Table 3: Definite vs. indefinite post-verbal nominal direct object frequencies

	Coastal			Koroshi			Turkmen			
	N	Po	%	N	Po	%	N	Po	%	Totals
Definite	110	9	8	126	3	2	89	2	2	325
Indefinite	229	14	6	56	1	2	104	1	1	389
Totals	339			182			193			714

Based on the admittedly blunt instrument of the definiteness classification in WOWA, the distinction between definite and indefinite does not contribute much to the explanation. Either the absolute figures are too low (Koroshi and Turkmen), or do not reach significance (Coastal Balochi).

A second factor that is often claimed to be relevant in placement of direct objects is weight. The WOWA data set distinguishes four levels of syntactic weight, based on orthographic words excluding clitics and adpositions: 1, 2, 3, and >3.[3] An example of a heavy direct object (three words) is 13, a light direct object is (12b) above.

(13) Coastal Balochi (Nourzaei 2021a, A, 0028)
      'mã ō'tī   'mačč-e   'hōš   de'gar-ā
      1SG REFL.GEN date_palm-GEN cluster other-OBL
      'na-dāt-ag=ã
      NEG-give.PST-PP=COP.PRS.3PL
      'I didn't give **my date palm clusters** to anyone.'

The figures comparing postverbal placement of the lightest (1 word) with the heaviest (>2 words) nominal direct objects are shown in Table 4 (all dialects combined).

Table 4 echoes findings from the WOWA spoken language corpora, which suggest that weight is not a significant factor in triggering object postposing (see Haig et al. 2024 [this volume]). In fact, the opposite tendency is suggested by our data.

---

[3]An additional measure of weight (number of characters) is also available, but was not applied here; see Haig et al. 2024 [this volume].

Table 4: Frequencies of post-verbal nominal direct objects, light vs. heavy

	All dialects		
	N	Po	%
Light (1 word)	329	16	4.9
Heavy (>2 words)	86	0	0
Totals	415		

Turning now to the factor of animacy, here reduced to human versus non-human, there is an interaction between humanness, pronominality, and post-verbal placement. Table 5 provides the relevant figures:

Table 5: Post-verbal direct objects (all doculects), according to human-ness and pronominality

	+Hum			-Hum			
	N	Po	%Po	N	Po	%Po	Totals
Nominal	171	10	5.8	543	20	3.7	714
Pronominal	142	27	19	31	2	6.5	173
Totals	313			574			887

First, these figures suggest that there is a strong correlation between humanness and pronominality: more than 80% of all pronominal objects are human (142 out of 173). Thus non-human pronominal objects are a rarity in Balochi, confirming a cross-linguistic tendency to avoid non-human object pronouns in discourse (Haig et al. 2022). Second, pronominality generally increases the likelihood of post-verbal placement, irrespective of humanness. Third, humanness alone is only marginally relevant: a human, nominal direct object is not significantly more likely to be postverbal than a non-human, nominal object. The difference between human and non-human only becomes relevant when the direct object is pronominal.

Closer inspection of the data reveal that the general tendency to avoid non-human object pronouns is most pronounced in Coastal Balochi, where almost 90% of object pronouns are human. Furthermore, in Coastal Balochi the tendency

to place these [+human] object pronouns after the verb reaches around 30%, and it is the Coastal Balochi data which actually account for most of the effects shown in Table 5. In fact, Coastal Balochi exhibits a nascent tendency towards the system found in Kumzari (Haig 2022b, based on Anonby 2015), where the majority of pronominal objects are post-verbal, and nominal objects are still predominantly pre-verbal.

In sum, we have seen that all doculects exhibit dominant OV order, but closer inspection reveals a difference between Coastal Balochi and the other two: although all three dialects are dominant OV overall, Coastal Balochi shows a significantly higher number of postverbal objects than the other two, with the strongest effect found for human pronominal objects. This finding is difficult to reconcile with our earlier observation that Coastal Balochi is more postpositional than either of the other doculects (Section 3). If it is more strictly postpositional, we might have expected it to be more rigidly OV, but this is not the case. We see no obvious explanation for the relatively high degree of object mobility in Coastal Balochi. But we have seen elsewhere that dominant use of postpositions does not necessarily entail strict OV; both Georgian and Armenian are post-positional, yet object positioning is actually more flexible than it is for the mostly prepositional Iranian languages (Hodgson et al. 2024 [this volume]). Conversely, strict OV does not necessarily entail postpositions; witness some of the Iranian languages closer to the Mesopotamian core of WOWA, which are strictly OV but dominant prepositional, or the varieties of Neo-Aramaic which have shifted to OV, but remain prepositional (see Noorlander 2024 [this volume]). While we have no explanation for the divergent characteristics of Coastal Balochi, they further underscore the potential disconnect between adpositional order and verb/object ordering in discourse.

Our data further suggest that weight is not a significant factor in predicting post-verbal placement of objects. We have identified the difference between nominal and pronominal form as relevant, though only in conjunction with humanness. Basically, we find that neither humanness, nor pronominal form alone predicts post-verbal placement. However, with pronouns, a strong interaction with humanness becomes apparent; this tendency is strongest for Coastal Balochi, but we can only speculate on the reasons for this at present.

4.1.1 Copula complements

A similar picture emerges here as with direct objects: all dialects overwhelmingly prefer preverbal placement, as in the following examples:

(14)   a.   Coastal Balochi (Nourzaei 2021a, A, 0299)
       te'yōk-ā *bī*
       alone-ᴏʙʟ COP.PST.3SG
       'He was alone.'

   b.   Turkmen Balochi (Haig 2022a, A, 0154)
       xānbādorr=o   *man=on*
       Khanbadur=ꜰᴏᴄ 1SG=COP.PRS.1SG
       'I am Khanbadur.'

   c.   Koroshi Balochi (Nourzaei 2021b, A, 0006)
       ē   *'šāh 'ǰan=e da'wom=ī*   ***'xeylī ā'dam=e 'xūb=ī***
       ᴘʀᴏx king wife=ᴇz second=ᴘᴄ.3SG very   person=ᴇz good=ɪɴᴅv
       'na-bod-ag=en
       ɴᴇɢ-become.PST-PP=COP.PRS.3SG
       'This king's second wife was not a very good person.'

The relevant figures are provided in Table 6.

Table 6: Frequencies of post-verbal copula complements

	Coastal			Koroshi			Turkmen			
	N	Po	%	N	Po	%	N	Po	%	Totals
Copula complements	310	13	4.2	53	0	0	66	2	3	429

Examples of a post-verbal copula from Coastal Balochi and Turkmen are the following:

(15)   Coastal Balochi (Nourzaei 2021a, C, 0997)
   ke   *ā-'ī*   *'nām=a*   *rahīm'baxš*
   ᴄʟᴍ ᴅɪsᴛ-ɢᴇɴ name=COP.PST.3SG Rahimbakhsh
   'who is called Rahimbakhsh'

(16)   Turkmen Balochi (Haig 2022a, A, 0022)
   man=om ast-om   ***pādišā=īē***
   1SG=ᴀᴅᴅ exist.PRS-1SG king=ɪɴᴅv
   'I am a king.'

These findings confirm a broader tendency observable across the entire WOWA data base, namely that in dominant OV languages, the position of copula

complements is generally conservative in the sense that they are even less prone to post-posing than direct objects are. In our sample, we find near-categorical pre-verbal placement of copula complements.

4.2 Verb and goal

Korn (2022) and Jahani (2018) suggest that among endpoint constituents, goals of verbs of movement (e.g., 'go', 'come') and goals of verbs of caused motion (e.g., 'put', 'throw', 'bring') have the highest rates of post-predicate position among all argument types, and across all data. Our data confirm this result. Examples of goals and goals of caused motion are the following:

(17)   a.   Goal
      Coastal Balochi (Nourzaei 2021a, A, 0043)
      ’šo   ***mē’tag-ā***
      go.PST.3SG home-OBL
      'He went home.'

   b.   Goal
      Koroshi Balochi (Nourzaei 2021b, A, 0047)
      ma’rō ’raft-ay   ***lō’g-ā***
      today go.PST-2SG home-OBL
      'Today (when) you go home.'

   c.   Goal
      Turkmen Balochi (Haig 2022a, A, 0160)
      šot   *bi* ***yak ǰā=ē***
      go.PST.3SG to one place=INDV
      'He went to a certain place.'

   d.   Caused Goal
      Coastal Balochi (Nourzaei 2021a, A, 0066)
      ǰat=e   *’zahm=e*   ***’dast=e***
      beat.PST=PC.3SG sword=INDV hand=PC.3SG
      'He struck the sword at its hand.'

   e.   Caused Goal
      Koroshi Balochi (Nourzaei 2021b, B, 0510)
      ma-ba’r-ā   *bod-a*   *ma-prē’n-ā*
      IPFV-take.PRS-BACKG.3SG become.PST-PP IPFV-throw.PRS-BACKG.3SG
      bod-a   *mā* ***dar’yā-hā***
      become.PST-PP into sea-OBL
      'She used to take it [and] throw it into the sea.'

f. Caused Goal
Turkmen Balochi (Haig 2022a, D, 0609)

mn-ā   ǰat   be ḍigār-ā
1SG-OBL beat.PST.3SG on ground-OBL

'(The donkey) threw me onto the ground.'

However, the dialects in our sample differ in the degree to which goals are postposed. Table 7 compares nominal goals and caused-motion goals, excluding adverbs:.

Table 7: Frequencies of post-verbal nominal goals

	Coastal			Koroshi			Turkmen			Totals
	N	Po	%	N	Po	%	N	Po	%	
Goal	85	60	70.6	68	63	92.6	23	8	34.8	176
Caused Goal	21	6	28.6	17	13	76.5	13	7	53.8	51
Totals	106			85			36			227

For Koroshi, the results are typical for an Iranian language close to the Mesopotamian core of the Western Asian Transition Zone, and also for colloquial spoken Persian (Haig et al. 2024 [this volume]): more than 80% of all goals are post-verbal. For the other two dialects, the results are somewhat puzzling. For Turkmen, significantly lower rates of post-verbal Goals can plausibly be related to effects of Central Asian varieties of Turkic, in line with the predictions of Haig et al. (In press) and Haig et al. 2024 [this volume], which suggest that the placement of Goals is highly sensitive to language contact. However, the absolute number of goals in the Turkmen data set is quite low, so this is a provisional conclusion. For Coastal Balochi, the most puzzling aspect is the overall low levels of postverbal caused goals, an observation that echoes Korn's (2022) finding for another Southern Balochi variety. For most data sets in the WOWA sample, the difference between caused and simple goals is not significant, and the two roles can be unified for most analyses. But for Coastal Balochi, the difference is striking; we have no explanation for this; it definitely merits further investigation (verb-specific effects, for example).

4.3 Recipients and addressees

A number of previous studies have shown that goals, recipients, and addressees do not necessarily pattern alike in the OV languages of Western Asia (Haig et al. 2024 [this volume], Korn 2022, Stilo 2018). Particularly addressees may exhibit quite different word order properties, even between closely related varieties (Jahani 2018 for Balochi, Haig 2022c for Kurdish). Small corpora of naturalistic spoken language are problematic for testing word order of addressees and recipients, because relevant tokens are not particularly frequent, and often pronominal. In varieties with clitic pronouns, the position of the pronoun is determined by language-specific clitic-placement principles, which may be quite distinct from the principles governing word order of prosodically independent constituents. Our findings here are correspondingly tentative. Examples of recipients and addressees are provided below, while Table 8 gives the frequencies per doculect.

(18)   a.   Recipient
          Coastal Balochi (Nourzaei 2021a, A, 0 527)
          k-ā'r-ã   *de'-yã*   ***'pet-a***
          K.IPFV-bring.PRS-3PL give.PRS-3PL father-OBL
          'They bring [and] give [them] to [their] father.'

      b.   Addressee
          Coastal Balochi (Nourzaei 2021a, A, 0023)
          o'š-ī   ***čo'k-ān***
          say.PRS-3SG child-OBL.PL
          'He says to [his] children.'

      c.   Recipient
          Koroshi Balochi (Nourzaei 2021b, B, 580)
          pu'l-ā   *a='dā*   ***āle'm-ok-ā***
          money-OBL VCL=give.PRS-3SG wise_man-DEF-OBL
          'She gives the money to the doctor (lit. wise man).'

      d.   Addressee
          Koroshi Balochi (Nourzaei 2021b, A, 0320)
          *'ya 'rō 'šāh **ba dūmād-o'bār=ay**   *a='š-īt*
          one day king to son_in_law-PL=PC.3SG VCL=say.PRS-3SG
          'One day, the king says to his sons-in-law.'

The findings do not entirely match those of Jahani (2018) and Korn (2022: 109), who suggest regular pre-verbal addressees in "Southern Balochi", to which our Coastal Balochi would belong. The findings for Coastal Balochi are also puzzling

Table 8: Frequencies of post-verbal nominal addressees and recipients (includes recipient/benefactives)

	Coastal			Koroshi			Turkmen			
	N	Po	%	N	Po	%	N	Po	%	Totals
Addressees	40	16	40	12	4	33	5	0	0	57
Recipients	19	4	20	3	2	66	5	1	20	27
Totals	59			15			10			84

in view of a general tendency noted in Haig et al. 2024 [this volume], according to which addressees are generally less likely to be post-verbal, which would align with the findings of Jahani (2018) and Korn (2022). However, in our Coastal Balochi sample, this is not the case; we have no explanation for this mismatch; this requires further research.

Turning to pronominal addressees and recipients, Table 9 provides the relevant figures.

Table 9: Frequencies of post-verbal pronominal addressees and recipients (includes recipient/benefactives)

	Coastal			Koroshi			Turkmen			
	N	Po	%	N	Po	%	N	Po	%	Totals
Addressees	13	2	15	8	1	1	6	0	0	27
Recipients	40	9	25	15	8	50	6	2	33	48
Totals	43			23			12			75

With pronominal arguments, the expected trend for higher frequency of postposed recipients is confirmed in all doculects. Examples of nominal addressees are provided in (19a-19c) and a pronominal example is shown in (19d).

(19)   a. Coastal Balochi (Nourzaei 2021a, B, 0504)

       ’nī   lōṭā-’ēn-ī   **brā’t-ā̃**

       now call-CAUS.PRS-3SG brother-OBL.PL

       'Then he called the brothers.'

   b. Coastal Balochi (Nourzaei 2021a, A, 0188)
      'nī ǰa'nek-ā go'š-ī   'to 'dar ā
      now girl-OBL say.PRS-3SG 2SG PREV SBJV.come.PRS.2SG
      'Then he [the boy in the well] said to the girl, "You get out."'

   c. Koroshi Balochi (Nourzaei 2021b, A, 0320)
      *'ya 'rō 'šāh **ba dūmād-o'bār=ay** a='š-īt*
      one day king to son_in_law-PL=PC.3SG VCL=say.PRS-3SG
      'One day, the king says to his sons-in-law.'

   d. Turkmen Balochi (Haig 2022a, D, 0540)
      *mnī   piss **pamman** sendbad-ī nakl-ā   kort*
      1SG.GEN father for.1SG   Sindbad-GEN story-OBL do.PST.3SG
      'My father told me the story of Sindbad.'

In summary, the findings for addressees and recipients largely confirm a trend observed across other Iranian languages of the region, according to which recipients are generally more likely to be post-verbal than addressees. However, in Coastal Balochi for nominal arguments (Table 8) only, the trend is reversed. We currently lack an explanation for this, which definitely requires more research.

4.4 Complements of 'become', Place, Source, Instrument, Benefactive, Comitative

The absolute numbers of tokens for the remaining roles are quite low in several doculects, so quantitative analysis is not always meaningful. We have combined the results in an overview Table 10. Due to low absolute figures, we include all possible POS types, including pronouns, adverbs etc. Figure 2 provides an overview visualization of all roles considered.

With regard to complements of change-of-state verbs ('become'), Coastal Balochi and Turkmen are overwhelmingly pre-verbal, while Koroshi shows around one third post-verbal placement. This would be expected given that the attested Iranian languages which have dominant post-verbal 'become'-complements are all from the southwestern Mesopotamian periphery of WATZ (Haig et al. 2022, Haig et al. 2024 [this volume]). An example of post-verbal 'become'-complement from Koroshi is provided in (20).

(20)   Koroshi Balochi (Nourzaei 2021b, A, 0491)
     ba ha'm=ī   kasa-ō-'ēn   ga'hār=eš   a='b-ant
     to EMPH=PROX small-DIM-ATTR sister=PC.3PL VCL=become.PRS-3PL

Table 10: Other roles: frequencies of post-verbal placement

	Coastal			Koroshi			Turkmen			
	N	Po	%	N	Po	%	N	Po	%	Totals
'become'-compl.	14	1	7	12	4	33	10	0	0	36
Place	61	12	19.7	21	7	33.3	25	0	0	107
Source	12	1	8.3	7	3	43.9	21	2	9.5	40
Instrument	18	5	28	5	0	0	12	1	8	35
Comitative	18	3	17	13	4	30	18	2	11	49
Benefactive	13	2	15	12	1	8	2	0	0	27

> ***ka'nīz=o   naw'kar***
> maidservant=and male_servant
> '[You know, these six sons-in-law and their wives came and] became servants to this their youngest sister.'

Turning to local roles source and place, it has been suggested that oblique arguments generally tend to prefer post-predicate position (Jing et al. 2021, based predominantly on written-language corpora). Above, we have shown that this certainly applies to goals, but the data for source and place provide only weak support for assuming a general tendency applying to all obliques. Nevertheless, it is noteworthy that the other obliques in Table 10 do show notably higher rates of post-verbal placement than direct objects. Examples for place and source are the following:

(21) a. Coastal Balochi (Nourzaei 2021a, A, 0327)
      *ha'm=ē   'gōšt pa'dā kašt=ī   **da'p-ā   'če***
      EMPH=PROX meat again pull.PST.3SG=PC.3SG mouth-OBL from
      'It took out this meat from its mouth again.'

      b. Coastal Balochi (Nourzaei 2021a, A, 0063)
      *'hanga 'nešt   ha'm=ē   **'mačč-e   'čērā***
      still   sit.PST.3SG EMPH=PROX date_palm-GEN under
      'Still he [the boy] was sitting under this date-palm.'

      c. Koroshi Balochi (Nourzaei 2021b, A, 0217)
      *'kār a=kan-ān   **'mā 'ī   bā'ġ-ā***
      work VCL=do.PRS-1SG in   PROX garden-OBL
      'I will work in this garden.'

d. Koroshi Balochi (Nourzaei 2021b, B, 0612)
   ba mad'rasā=om   *ġada'ġan=eš*   *kod-a*
   to school.OBL=ADD forbidden=PC.3PL do.PST-PP
   'At school they have actually forbidden (your) coming.'

The data for instruments, comitatives, and benefactives are quite thin, but suggest a tendency for all three roles to be predominantly pre-verbal. This is particularly noteworthy for benefactives, which are >80% pre-verbal in all doculects. Benefactives are sometimes included under the umbrella term of "target" (Asadpour 2022a,b, Korn 2022) together with goals, but the Balochi data suggest that benefactives are far less likely to be postverbal than goals, and in fact no more likely to be postverbal than sources and locations. However, it should be noted that the majority of benefactives (about 90%) in our data are pronominal, so the bias towards preverbal position may be linked to the pronominal status, but given the paucity of nominal benefactives in the data, this remains to be clarified.

5 Summary: Post-verbal constituents in Balochi

Figure 2 summarizes the quantitative data from post-verbal positioning of various non-subject constituents considered in the preceding section. The values for the roles Recipient, Addressee, Benefactive, and Comitative are combined to "Oblique" in Figure 2; see the preceding section for individual roles.
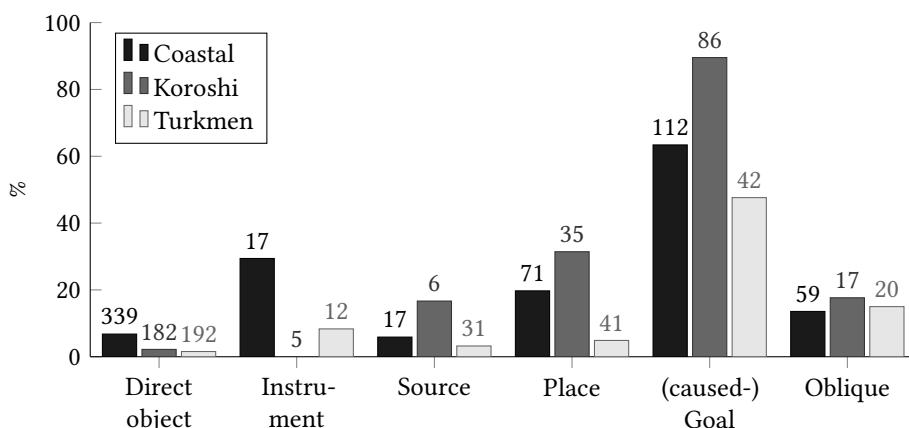
Figure 2: Frequency of post-verbal arguments in three Balochi doculects. *Note*: Ticks above bars indicate absolute count.

As expected, the highest rates of post-verbal placement for all doculects are found for goals, confirming the by-now familiar finding across the entirety of WOWA. Note that it is only in Koroshi that both recipients and goals are dominantly (>60%) post-verbal, a pattern that has parallels with varieties of Kurdish (Haig 2022c). In our data, addressees do not pattern with goals in any doculect, and overall do not differ from other obliques, showing less than 50% post-verbal placement. With regard to direct objects, all doculects are overwhelmingly OV, with post-verbal objects accounting for less than 10% of all objects. However, Coastal Balochi shows a significantly higher number of post-verbal objects than either of the other two doculects. In our sample, weight and definiteness do not seem to be relevant in accounting for post-verbal direct objects, in any doculect. We do, however, find an effect of humanness, particularly in combination with pronominal direct objects (Table 5 above). Overall, the most likely direct objects to be postposed are human, pronominal objects, with the difference most obvious for Coastal Balochi.

6 Conclusions and implications for comparative Iranian syntax

Table 11 summarizes the overall degree of head-initial syntax in various phrase types in the three doculects, based on the discussion in Section 3.

Table 11: Distribution of head-initial syntax in three Balochi doculects

Parameter	Koroshi	Turkmen	Coastal
N/Adjective	N-Adj	Adj-N	
N/Possessor	N-Poss	Poss-N	
Adp/N	Prep-N	Prep-N	N-Postp
Main clause/Compl. clause	Main-Compl. Clause		
Aux(Modal)/Main verb	Aux(Modal)-Main verb		
Complementizer/Clause	Compl-Clause		

Across these typological parameters, Koroshi is the only variety that is consistently head-initial, presumably linked to a greater degree of contact with Persian and other western Iranian languages exhibiting head-initial NPs. All doculects linearize main and complement clauses, complementizers, and auxiliaries alike, sharing these values with probably all other Iranian languages of WOWA. The most surprising aspect of Table 11 is the dominance of postpositions in Coastal

Balochi, for which we have no ready explanation; this is certainly an area for future research.

Finally, we consider the relevance of the Balochi data for addressing the larger question of how post-verbal syntax in Iranian may have emerged: What kind of historical scenario is compatible with the areal distribution that is now becoming evident? First, let us consider the Balochi data in the light of a feature that characterizes all other contemporary spoken western Iranian language in the WOWA sample: in connected discourse, nominal goals (including caused-motion goals) are 60–100% post-verbal. The sole exception turns out to be Turkmen Balochi (Turkmen), where the figure drops to around 40%. Turkmen Balochi is thus different from its relatives in this respect, and from any of the Iranian languages spoken in Mesopotamia, where post-verbal Goals are >85%. The most obvious explanation for the outlier status of Turkmenistan Balochi is its outlier geographic position, beyond the northeastern periphery of WATZ, and farthest from the Mesopotamian core that has been identified as harbouring the highest levels of post-verbal Goals. We take this as tentative support for the hypothesis that the placement of goals is particularly sensitive to contact influence.

Nevertheless, 40% post-verbal Goals is higher than many other OV languages in our sample (vernacular standard Turkish of Ankara, Iefremenko 2024 [this volume] for example), so geography cannot be the whole story. Generally, it appears that Iranian languages are just more prone to post-verbal Goals than Turkic languages, and this is a matter of shared inheritance, rather than geography alone. What we might therefore assume is that the shared ancestor of all western Iranian languages had some initial baseline level of post-verbal spatial goals, which may already have exceeded 50% in some of the (unattested) languages of that period. This feature was thus present prior to the dispersion of the various branches of what is traditionally termed 'West Iranian' (see also Korn 2022: 122). The result would be traces of post-verbal Goals (in the narrow sense) in all West Iranian languages, with actual levels dependent on their respective contact biographies over the last 2000 years. In those Iranian languages that shared territory with with Semitic languages (in particular Aramaic) for at least a millennium in the Mesopotamian and western Zagros regions, levels of post-verbal Goals converge to the near 100% that characterize the Semitic contact languages. Where Semitic influence is lacking, and contact with other OV languages (e.g. Indo-Aryan) was present, levels would have dropped, as in Balochi of Turkmenistan, or at least not risen further. Alternatively, the unrelated OV contact languages might converge to Iranian in this respect, as is the case for Qashqai Turkic, which exhibits 70% post-verbal nominal goals (Schreiber 2021), a figure close to its Iranian neighbours. The actual outcome would thus depend

on the local patterns of multilingualism and power relations among the speech communities, but a broad geographical tendency is nonetheless discernible.

This is essentially a refinement of the account outlined in Haig (2015), who noted that the wide distribution of post-verbal "Goals" (see below on terminology) across the Iranian languages of western Asia is suggestive of "an old trait of West Iranian that was inherited by the daughter languages," rather than independent parallel development (Haig 2015: 421). post-verbal Goals are attested in the Old Iranian period, as in the following Old Persian examples, cited from Haig (2015: 421); see also Jahani (2018) for further examples:

(22)   a. Old Persian (DB V, 9–10, Kent 1953: 133)
          pasāva Gaubaruva hadā kārā ašiyava **ūvjam**
          then   Gobryas   with army marched to_Elam
          'Then Gobryas marched to Elam with an army.'
       b. Old Persian (DB V, 21, Kent 1953: 133)
          pasāva hadā kārā ašiyavam abiy **sakām**
          then   with army went.1sg to   Scythia
          'Then I went with an army to Scythia.'

However, the formulation in Haig (2015) was still framed in terms of an assumed macro-role sense of "Goal", which included addressees, and recipients, among other things. More recent research (Iefremenko 2024 [this volume]), including this chapter, demonstrates that this approach is not tenable. Instead, a finer-grained approach is required that consistently distinguishes Goal in the narrower sense of motion and caused motion from other roles. As shown above in Figure 2, outside of spatial goals, no other role surpasses 50% rates of post-verbal placement in Balochi, with the exception of recipient, and that only in Koroshi, and this finding is replicated for several other Iranian languages.

Our revised suggestion would be that in proto-West Iranian, post-verbal Goals (in the narrow sense) were already a frequent option (perhaps majority), while postposing other arguments was possible, but determined by pragmatic and semantic principles, with no clear role-driven preferences; the multi-variate analysis of spoken Persian in Rasekh-Mahand et al. 2024 [this volume], demonstrates such a system empirically. In some Iranian languages, most notably Kurdish of the Mesopotamian region, post-verbal placement spread from goals to other argument types, ultimately encompassing a bundle of semantically-related roles. The initial focus on these languages led to the (misleading) assumption of a macro-role ("Goal" in Haig & Thiele 2014 and Haig 2015; "Target" in Asadpour 2022b and Jügel 2022; see Haig et al. 2024 [this volume], for discussion. The role-specific

analysis adhered to here suggests an areally-mediated shift in the frequency of post-verbal Goals in Balochi (lowest in Turkmen), and in the range of arguments that behave similarly to goals (the Koroshi clustering of recipient and Goal).

The developments just sketched are speculative to the extent that they reconstruct possible diachronic developments, based on synchronic feature distributions. The little historical data that is available is unfortunately difficult to interpret. There is a general problem with projecting from written texts to spoken language; an investigation of contemporary formal written Persian, for example, would fail to register the fact that spoken contemporary Persian has a fairly stable rate of around 80% post-verbal Goals (Haig et al. 2024 [this volume]). Early Classical Persian texts (11–13 centuries AD) have virtually no trace of post-verbal Goals (Parizadeh & Rasekh-Mahand 2024 [this volume]), but we doubt that these figures reflect the spoken language of the time, given the comparative evidence from other Iranian languages. Similarly, Jahani (2018) compares oral and written Southern Balochi, noting that in written Balochi, goals are never post-posed, while in spoken Southern Balochi the relevant figure is 90%. These differences raise a methodological problem with regard to reconstructing the history of post-verbal syntax; differences between yesterday's written and today's spoken language may just reflect differences between written and spoken registers that have always existed, rather than being evidence of any change across time.

This chapter provides a tentative cross-dialect survey of word order in Balochi, based on three doculects representing geographically dispersed varieties, with differing contact profiles. The picture that emerges confirms some of the previous observations in Jahani (2018) and Korn (2022), but extends the purview of these studies. We investigated a larger number of roles, and considered additional factors such as weight and humanness. On the other hand, the pan-dialectal approach means that we are unable to zoom in on finer semantic, pragmatic, or stylistic factors, such as the effects of individual verb classes, register, or information structure that figure in Korn (2022). Generally, our findings confirm the overall prediction that both overall frequencies of post-verbal Goals, and the number (and nature) of other role types that are prone to post-verbal placement, decrease with increasing distance from the Mesopotamian core of WATZ. We have also identified Coastal Balochi as exhibiting a significantly higher frequency of post-postions in discourse, and interestingly, a higher number of post-verbal direct objects, with a strong interaction with pronominal form and humanness. These findings raise interesting questions regarding the possible role of contact with Indo-Aryan languages in the region, which have yet to be explored.

Abbreviations

ADD	additive particle	NEG	negation
ATTR	attributive	OBJ	object case
BACKG	backgrounding	OBL	oblique case
CAUS	causative	PC	person-marking
CLM	complementizer		enclitic
Coastal	Coastal Balochi doculect	PL	plural
		PP	past participle
COP	copula	PROH	prohibitive prefix
DEF	definite	PROX	proximal
DIST	distal	PRS	present
EMPH	emphasis	PST	past stem
EZ	ezafe particle	REFL	reflexive pronoun
FOC	focus particle	SBJV	subjunctive
GEN	genitive	SG	singular
IND	indefinite	Turkmen	Turkmenistan
INDV	individuation clitic		Balochi doculect
IPFV	imperfective	UP	Unpublished text
K.IPFV	*k*-variant of	VCL	verbal clitic
	imperfective	WATZ	Western Asian
Koroshi	Koroshi Balochi doculect		Transition Zone

References

Anonby, Christina van der Wal. 2015. *A grammar of Kumzari: A mixed Perso-Arabian language of Oman.* Leiden: Leiden University. (Doctoral dissertation).

Asadpour, Hiwa. 2022a. *Typologizing word order variation in northwestern Iran.* Frankfurt am Main: Goethe-Universität. (Doctoral dissertation).

Asadpour, Hiwa. 2022b. Word order in Mukri Kurdish - The case of incorporated targets. In Hiwa Asadpour & Thomas Jügel (eds.), *Word order variation: Semitic, Turkic and Indo-European languages in contact*, 63–87. Berlin: De Gruyter Mouton.

Axenov, Serge. 2000. Balochi orthography in Turkmenistan. In Carina Jahani (ed.), *Language in society – Eight sociolinguistic essays on Balochi* (Studia Iranica Upsaliensia 3), 71–78. Uppsala: Acta Universitatis Upsaliensis.

Axenov, Serge. 2006. *The Balochi language of Turkmenistan: A corpus-based grammatical description (StudiaIranica Upsaliensia, 10)*. Uppsala: Acta Universitatis Upsaliensis.

Baranzehi, Adam Nader. 2003. The Sarawani dialect of Balochi and Persian influence on it. In Carina Jahani & Agnes Korn (eds.), *The Baloch and their neighbours: Ethnic and linguistic contact in Balochistan in historical and modern times*, 75–111. Wiesbaden: Reichert.

Barjasteh Delforooz, Behrooz. 2010. *Discourse features in Balochi of Sistan (oral narratives)*. Uppsala: Uppsala Universitet. http://uu.divaportal.org/smash/record.jsf?pid=diva2:345413.

Buddruss, Georg. 1988. *Aus dem Leben eines jungen Balutschen, von ihm selbst erzählt*. Stuttgart: Steiner.

Emādi, Nezam. 2005. *Guyeš-e Korowš (The vernacular of the Korosh)*. Shiraz: Entešārāt-e Širāz Āvand Andišeh.

Haig, Geoffrey. 2015. Verb-goal (VG) word order in Kurdish and Neo-Aramaic: Typological and areal considerations. In Geoffrey Khan & Lidia Napiorkowska (eds.), *Neo-Aramaic and its linguistic context*, 407–425. Piscataway, NJ: Gorgias Press.

Haig, Geoffrey. 2022a. Balochi (Turkmenistan). In Geoffrey Haig, Donald Stilo, Nils N. Schiborr & Mahîr Dogan (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation*. Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

Haig, Geoffrey. 2022b. Kumzari (Musandam). In Geoffrey Haig, Donald Stilo, Nils N. Schiborr & Mahîr Dogan (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation*. Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

Haig, Geoffrey. 2022c. Kurdish (Northern, Muş). In Geoffrey Haig, Donald Stilo, Nils N. Schiborr & Mahîr Dogan (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation*. Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

Haig, Geoffrey, Paul M. Noorlander & Nils N. Schiborr. In press. Which word order features are stable in a contact setting? Corpus-based evidence from the Western Asian transition zone. In Jeroen Darquennes, Joe Salmons & Wim Vandenbussche (eds.), *Language contact: An international handbook*, vol. 2. Berlin: De Gruyter Mouton.

Haig, Geoffrey, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils Schiborr. 2024. Post-predicate elements in the Western Asian Transition Zone: Data, theory, and methods. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 3–54. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266331.

Haig, Geoffrey, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr. 2022. *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

Haig, Geoffrey & Hannah Thiele. 2014. Post-predicate Goals in Northern Kurdish and neighbouring languages: A pilot study in quantitative areal linguistics languages. Paper presented at the *Second International Conference on Variation and Change in Kurdish, mardin artuklu university, turkey,* 8-9 October 2014.

Hodgson, Katherine, Victoria Khurshudyan & Pollet Samvelian. 2024. Post-predicate arguments in Modern Eastern Armenian. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 375–410. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266355.

Iefremenko, Kateryna. 2024. Word order in the speech of Kurmanji-Turkish bilinguals. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 55–85. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266333.

Jahani, Carina. 2013. The Balochi language and languages in Iranian Balochistan. *Journal of the Middle East and Africa* 4(2). 153–167.

Jahani, Carina. 2018. Post-verbal arguments in Balochi. Paper presented at *the International Conference on Anatolia-Caucasus-Iran: Ethnic and linguistic contacts (ACIC),* Yerevan University, 10–12 May.

Jahani, Carina & Agnes Korn. 2009. Balochi. In Gernot Windfuhr (ed.), *The Iranian languages*, 634–692. London & New York: Routledge.

Jing, Yingqi, Paul Widmer & Balthasar Bickel. 2021. Word order variation is partially constrained by syntactic complexity. *Cognitive Science* 45(11). 1–25. DOI: 10.1111/cogs.13056.

Jügel, Thomas. 2022. Word order variation in Middle Iranic: Persian, Parthian, Bactrian, and Sogdian. In Hiwa Asadpour & Thomas Jügel (eds.), *Word or-*

der variation – Semitic, Turkic and Indo-European languages in contact, 39–62. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110790368-003.

Kent, Roland. 1953. *Old Persian*. 2nd revised edition. New Haven, CT: American Oriental Society.

Korn, Agnes. 2005. *Towards a historical grammar of Balochi*. Wiesbaden: Reichert.

Korn, Agnes. 2008. A new locative case in Turkmenistan Balochi. *Iran and the Caucasus* 12(1). 83–99.

Korn, Agnes. 2022. Targets and other post-verbal arguments in Southern Balochi: A multidimensional cline. In Hiwa Asadpour & Thomas Jügel (eds.), *Word order variation – Semitic, Turkic and Indo-European languages in contact*, 89–128. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110790368-005.

Korn, Agnes & Maryam Nourzaei. 2019. Notes on the speech of the Afro-Baloch of the southern coast of Iran. *Journal of the Royal Asiatic Society* 29. 623–657. DOI: 10.1017/S1356186319000300.

Noorlander, Paul M. 2024. Neo-Aramaic in Iran and northeastern Iraq. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 431–469. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266359.

Nourzaei, Maryam. 2017. *Participant reference in three Balochi dialects: Male and female narrations of folktales and biographical tales*. Uppsala: Acta Universitatis Upsaliensis. https://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-314090.

Nourzaei, Maryam. 2021a. Balochi (Coastal). In Geoffrey Haig, Donald Stilo, Nils N. Schiborr & Mahîr Dogan (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation*. Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

Nourzaei, Maryam. 2021b. Balochi (Koroshi). In Geoffrey Haig, Donald Stilo, Nils N. Schiborr & Mahîr Dogan (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation*. Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

Nourzaei, Maryam. 2024. Progressive constructions in different Balochi dialects from a diachronic and an areal linguistic point of view. *Journal of the Bulletin of the School of Oriental and African Studies*. 1–26. DOI: 10.1017/S0041977X24000144.

Nourzaei, Maryam. In press. *Documenting orality: Collecting folk narratives in Iran*. Uppsala: Acta Universitatis Upsaliensis.

Nourzaei, Maryam, Carina Jahani, Erik Anonby & Abbas Ali Ahangar. 2015. *Koroshi: A corpus-based grammatical description*. Uppsala: Acta Universitatis Upsaliensis.

Parizadeh, Mehdi & Mohammad Rasekh-Mahand. 2024. Post-predicate elements in Early New Persian (10–13th Century CE). In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 231–244. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266345.

Rasekh-Mahand, Mohammad, Elham Izadi, Mehdi Parizadeh, Geoffrey Haig & Nils Schiborr. 2024. Post-predicate elements in modern colloquial Persian: A multifactorial analysis. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 197–230. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266343.

Schreiber, Laurentia. 2021. Oghuz (Qashqai). In Geoffrey Haig, Donald Stilo, Nils N. Schiborr & Mahîr Dogan (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation*. Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

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

Chapter 5

Notes on word order in Bashkardi

 Agnes Korn

CNRS – UMR 8041 Centre de recherche sur le Monde iranien (CeRMI)

This contribution discusses some features of word order in Bashkardi, a group of varieties of the Iranian branch of Indo-European spoken inland of the Strait of Hormuz in Southern Iran. The data are from recordings made by Ilya Gershevitch in 1956, when Persian influence was less strong than today. The findings include an average of 30% non-subject elements being in postverbal position. Goals of motion and of caused-motion show a strong preference for this position, except for Goals of 'put'-expressions, which are close to the overall average. More than 20% of nominal direct objects are postverbal, while pronouns are very rare in this position.

1 Introduction

1.1 Affiliation and location

For the purposes of this contribution, the term "Bashkardi" (ISO 639-3: bsg) refers to the varieties spoken inland of the Strait of Hormuz in Southern Iran in a region called Bašākerd (see the map in Figure 1).[1] By its being situated far away from the core of the Western Asian Transition Zone (see Haig et al. 2024 [this volume]), Bashkardi offers a convenient point of comparison with Iranian languages within that zone.

Bashkardi has been said to belong to the South-Western sub-branch of Iranian (e.g. Skjærvø 1989a: 846). However, the differences among the Bashkardi varieties – particularly between North (NBsh.) and South Bashkardi (SBsh.) – and the features they share with Balochi (which is classified as North-Western Iranian) suggest that "Bashkardi" could be the result of a linguistic area where Iranian

---

[1]See Korn (2017: 79–80) for more details.

varieties of different genetic affiliations have converged and developed shared features (cf. Korn 2021b: 302f.). In this sense, "Bashkardi" is a collective term "merely on the basis of territorial coverage" (Voskanian & Boyajian-Sureniants 2007: 122).

1.2 Data and previous works

The data for this article are from recordings made by Ilya Gershevitch during his visit to Iran in 1956. He did not publish any text, but did use his data for his articles on historical linguistics.[2] The data he cites there are summarised in Skjærvø (1989a,b).

Two of the texts (H and G in what follows) are published in Korn (2021a,b), and the corpus used for the present study is available at Korn & Gershevitch (2023a,b).[3]

Elements of Bashkardi grammar are summarised in Korn (2017, description of the nominal system with historical interpretation), and a brief sketch of North Bashkardi grammar is presented in Korn (2021b).

Some data are available from varieties spoken in the region today: Seddiqi Nejad (2010), describing the South Bashkardi of Dahwast; Barbera (2023, 2024) on the South Bashkardi of Garu; Barbera (2005, 2020) on the variety of the town of Minab; and Pelevin (2010) on the speech of Bandar-Abbas. A comparison with Gershevitch's recordings shows that the morphosyntax of Bashkardi has since become much more similar to that of Persian (see Korn 2017: 93–95), which highlights the importance of the data of 1956.

1.3 Some elements of Bashkardi grammar

For the discussion to follow, some grammatical features need to be kept in mind (see also Section 2.4.1). In Bashkardi as defined in Sections 1.1-2, there is no case distinction. The only elements which are (historically) case-marked are the pronominal clitics (enclitic pronouns, PC). They are used for the possessor and the indirect object as well as for the direct object in the present domain (i.e. in clauses whose verb form is based on the present stem), and for the transitive subject in the past domain (clauses with verb forms based on the past stem; this

---

[2]See Gershevitch (1959) for an account of his journey, Korn (2015) for a description of Gershevitch's materials and Korn (2017: 81, 2021b: 301–302) for work on the data.

[3]All examples in what follows are from Korn & Gershevitch (2023a,b), specifying North / South Bashkardi as well as the text and sentence number in https://multicast.aspra.uni-bamberg.de/resources/wowa/#corpora. Further examples can be found in the works mentioned in this section.
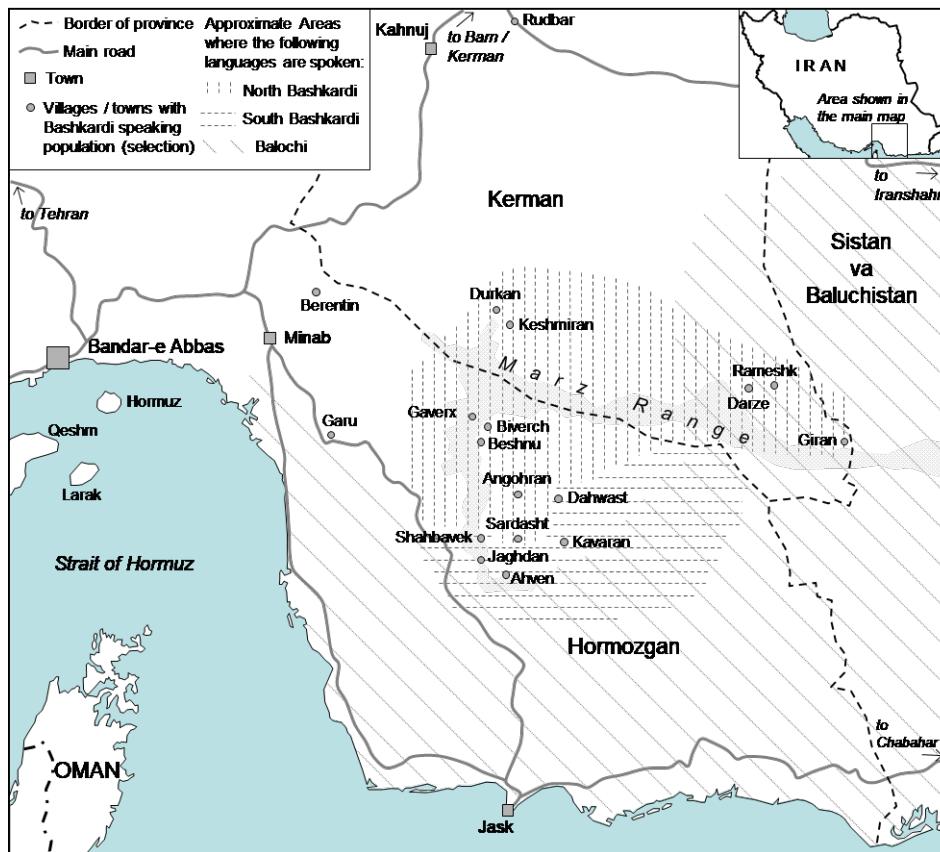
Figure 1: Location of Bashkardi varieties with locations mentioned in Gershevitch's materials

includes the perfect forms). In the latter domain, verbal agreement (if any) is with the direct object, and in some cases with the indirect object (recipient).

In both domains, nominal and pronominal indirect objects (see Section 2.4.3) and definite direct objects (see Section 2.4.2) can take a directional preposition ('to'), but are mostly unmarked, i.e. all elements marking syntactic relations mentioned in Section 2.1 are employed only sporadically.

Bashkardi is a heavily pro-drop language, which reduces potential examples of pronominal arguments as well as of other arguments that speakers can retrieve from the context.

In addition to the pronominal clitics, Bashkardi shows a number of other items which are consistently realized as enclitic to the preceding word, viz. the directional clitic (Section 2.1.3.2), the marker of specificity (Section 2.1.2), the possessor

clitic (Section 2.1.1.2), the forms of the copula, the subordinator *ke* (see Section 2.3) and the coordinating connective *o* 'and'.[4]

There are important differences between North and South Bashkardi (cf. Section 1.1), which include the use of TAM markers as well as the forms of the pronominal clitics and the verbal endings. Furthermore, there is a certain amount of variation also within each dialect group, and among speakers individually.

2 Word order profile

2.1 Noun phrases and adpositional phrases

Bashkardi noun phrases and adpositional phrases are head-initial (right-branching) as they are in Persian, but there are also left-branching patterns (which is the dominant pattern in Balochi). Demonstratives and numerals precede the noun.[5]

2.1.1 Adjectives and possessors

In general, adjectives and possessors follow the head noun and are linked to it by means of a clitic called "ezâfe" (=*i* and variants, EZ), which is also used for arguments of complex predicates (see Section 2.2).

2.1.1.1 Adjectives

Examples for the general pattern of an adjective being linked to its head noun by the ezâfe include (4) and (13a); an example without ezâfe is (11).

Some adjectives are preposed, cf. *aǰab* in (1) and *heil=o* in (3a) and (3b) – surely copied from Persian, where *aǰab* and *xeili* are preposed as well.

(1) North Bashkardi (I:59)
    aǰab   nöuk   ei mõ hast=ī
    amazing grandson to I   exists=PC.3SG
    'What an amazing grandson I have!'

---

[4]*ke* and *o* are noted as independent words and without '=' in the examples below; the same applies to the forms of the copula unless they are part of a verb form.

[5]For the Bsh. noun phrase, see Korn (2017), specifically pp. 91–92 for adjectives, pp. 88–90 for possessive NPs, p. 85 for the marking of definiteness and p. 83–86 for number, and 86–87 for adpositional elements.

2.1.1.2 Possessors

The general pattern of a possessor being linked to the possessum by ezâfe is shown in *mahala=i* PN 'X's place' in (16d) and *čürak-e šēr* in (14). No ezâfe is required when the possessum is a body part, or when the possessor is a pronoun of the 1st or 2nd person (4, 8b) or the reflexive pronoun, although the ezâfe may be used here as well (13a, 17b).

There is also a (rare) head-final pattern (looking like a Balochi noun phrase), with or without the possessor clitic (=*ī*, POSS), (2).

(2)   North Bashkardi (H:88, 103, 118, 134)
    tūlag-a=ī   *sīr=å*
    jackal-PL=POSS wedding=DIR
    '[I will go] to the jackals' wedding.'

Pronominal clitics are frequently used for the possessor (6a, 7b, 10c, 19).

They are also used for the possessor in the *mihi est* construction (the Bashkardi pattern expressing 'have', see Section 2.4.4), also in addition to a noun phrase (10d) or prepositional phrase (1) expressing the possessor. In most cases, 'exists' occupies the clause-final position (3a, 10b, 12a), but it can also be fronted (3b, 10b).

(3)   South Bashkardi (C:31f.)

    a. [Kadxodā:] *heil=o*   *būv=et*   *heš*
                     many=SPC date_palm=PC.2SG exists
      'Do you have many palm trees?' –

    b. *heš=om*   *haštåd ben*   *bū*
        exists=PC.1SG eighty trunk[6] date_palm
      'I have eighty palm trees.'

2.1.2 Demonstratives and numerals

Demonstratives precede the head noun without any linker. This is quite frequent, so that demonstratives approach the functions of a definite article (4, 5, 6a, 10c, 10d, 11, 14, 19).

There also is a (rare) suffix *-ak* that seems to express definiteness (4).

---

[6]*ben* can quite well be considered a numeral classifier here; 'trunk' is meant to render the literal meaning.

(4) North Bashkardi (H:18)

　　tūla=i　　xwara=ī　　hamī måst-ak-ūn　　mon a-xwar-ed
　　jackal=ᴇᴢ voracious=ѕᴘᴄ ᴅᴇᴍ1 yoghurt-ᴅᴇꜰ-ᴘʟ[7] I　　ɪᴘꜰᴠ-eat.ᴘʀѕ-3ѕɢ
　　'A voracious jackal keeps eating this yoghurt of mine.'

Numerals are followed by the head noun without (16b) or with an intervening numeral classifier (*ben* in (3b), *tå* in (10d)). In addition to the numeral 'one' (10d), singularity may be expressed by the specificity clitic (ѕᴘᴄ, etymologically 'one')[8] *=ī, =ē*, SBsh. also *=ō*, which is placed at the end of the NP (4). When introducing a new entity, the NP frequently shows both (9, 10a, 10b, 14) so that 'one' is "circumposed"; (5) even shows two instances of the specificity marker.

(5) North Bashkardi (K:83)

　　yak=ē　　gozer=ē　　hast=e　　　hamē mahal
　　one=ѕᴘᴄ important=ѕᴘᴄ exists=ᴄᴏᴘ.3ѕɢ[9] ᴅᴇᴍ1 place
　　'There is an important man in this place.'

2.1.3 Adpositions

2.1.3.1 Prepositions

Syntactic relations can optionally be clarified by prepositions, some of which are clearly borrowed from Persian (*barå_i* 'for' in (16c), *γeir=e* 'except for' in (17b), perhaps also *bå* 'with' in (11, 13a, 19)). Note the (sporadic) marking of objects with prepositions (see also Section 1.3).

For more precise meanings, nouns may be used as prepositions (relational nouns), e.g. *sar* 'head; on, above' or *dah* 'entrance' in (16e). Prepositions may also be combined with each other (6b) or with a relational noun.

(6) South Bashkardi (E:10f.)

　　a.　*ī　　dā=yom　　　kāšt=om　　　kūš*
　　　　ᴅᴇᴍ1 hand=ᴘᴄ.1ѕɢ[10] draw.ᴘѕᴛ=ᴘᴄ.1ѕɢ knife
　　　　'[With] this (= the other) my hand I drew the knife.'

---

[7]Certain liquids, dairy products etc. are treated as plural in Bashkardi.

[8]This term follows Hᴇɪɴᴇ, see Korn (2017: 85).

[9]Cf. the parallel constructions with ᴄᴏᴘ.ᴘѕᴛ in (10a, 10b, 10d). For copula forms added to a finite verb in Balochi, see Korn & Nourzaei (2019: 651–652).

b. *zat=om   be-rū der=e*
hit.PST=PC.1SG to-on heart=PC.3SG
I struck [it] into his [the leopard's] heart.'

2.1.3.2 Postposed elements

As for postposed elements, the (rare) directional clitic[11] *=å / =a* is appended to the noun, or to the NP (note its position following the pronominal clitic in (7b)). It may occur in combination with a preposition (16e).

(7)   North Bashkardi (K:126)

a. *xo, hålå på ba-Ø   tå   be-rr-īn*
well now foot become.PRS-IMP.2SG so_that MOD-go.PRS-1PL
'Now get up; let's go,

b. *tå   dehngōn[12]=et=a   seråk   be-dah-om.*
so_that landlord=PC.2SG=DIR showing MOD-give.PRS-1SG
[and] I will show [you] your landlord
(and you will see how miserable he is now).'

Note also the possessive clitic mentioned in Section 2.1.1.2.

2.2 Verbal expressions

Auxiliaries are rare in Bashkardi; *TAM categories* are expressed by the presence or absence of verbal prefixes, as seen throughout the examples above and below. The perfect system uses the perfect participle followed by the inflected copula, as does the progressive, which is built from the past stem or infinitive.[13]

The uninflected element *båål* 'it is necessary' (certainly copied from Persian *bāyad*, but not a transparent verb form in Bashkardi) is occasionally found; it is placed towards the beginning of the clause; the main verb is in the subjunctive (8b).

---

[10] *y* is the hiatus-bridging element attaching the pronominal clitic *=om*.

[11] This clitic should be distinguished from the "directional clitic" that attaches to verb forms in Gorani and Kurdish, see Haig et al. 2024 [this volume], Mohammadirad 2024 [this volume].

[12] *dehngōn* (cf. Persian *dehqān, dehgān*) shows an unetymological *-n-* throughout this text.

[13] See Korn (2022a) and Korn & Suleymanov (2017) for the functions of the prefixes, the progressive and the verbal system in general.

(8) North Bashkardi (J:46f)

    a. *ei eståd=ī   go   gorg ke*
       to master=PC.3SG say.PST wolf SUB
       'To the blacksmith said the wolf:

    b. *to   bååt-ē   gap=e pošbår mo be-zan-ī*
       you.SG it_is_necessary-2SG talk=EZ support I   MOD-hit.PRS-2SG
       "You have to speak up for me, (...).""

Complex predicates are frequent, e.g. *på b-* 'get up' (7a) and *seråk da-* 'show' (7b). The nominal part can take an argument constructed like a possessor (8b), (9), so that it is positioned within the complex predicate.

There is a continuum of combinations of a transitive verb with an indefinite (or generic) direct object to conventionalised complex predicates, which can be analysed as single predicative items. For instance, *šekåyat kan-* in (13b) could be analysed as a complex predicate 'to complain' or as 'lodge a complaint', thus *šekåyat* a direct object. (8b) could show *gap-e pošbår mo* 'my defense (talk of my support)' as direct object of *zan-*; alternatively, *pošbår mo* 'my support' could be the direct object of the complex predicate *gap zan-* ('talk-hit' is used for 'speak' also in Balochi, vs. Persian *ḥarf zan-*); or one could consider *gap-e pošbår zan-* as a complex predicate 'speak up', in which case *mo* would be the beneficiary.

There is thus a certain amount of subjectivity in the interpretation, which potentially affects the statistics of several categories.

(9) South Bashkardi (A:41)
    hålå to   yeu gap=o seråk=i   yamah adeh-Ø
    now you.SG one talk=SPC showing=EZ we[14]   give.PRS[15]-IMP.2SG
    'Now show us some talk [from the tape recorder]!'

There are also *directional preverbs* such as *or-* in (16e).

In the case of *clause-initial verbs*, it is often unclear whether we are looking at an element postposed to the verb, or at an instance of fronting of the verb. The latter may occur at the beginning of a tale, introducing the tale's characters (10a, 10b)[16] or at the beginning of a new episode, but also in pragmatic contexts that still need to be established (potential examples include (6b); see also the *mihi est* pattern in Section 2.1.1.2).

---

[14]As in Persian, the non-verbal element of a complex predicate can take an argument, which is attached by the ezâfe (for which see Section 2.1.1).

[15]The PRS stem of 'give' is *adeh-* in Gershevitch's SBsh. data; prefixing the imperfective *a-* yields *ādeh-* (Skjærvø 1989a: 848).

[16]See Korn (2020) for features of folk tales in Bashkardi.

(10) South Bashkardi (K:1-5)

   a. *hast=a*   *ya måldår=ē*
      exists=COP.PST.3sg one rich_man=SPC
      'There was a rich man.

(someone in the background:) Hm.

   b. *ya måldår=ē*   *hast=a*
      one rich_man=SPC exists=COP.PST.3SG
      hast=ar=ī   *ya sålål*
      exists=COP.PST.3SG=PC.3SG one shepherd
      There was a rich man, [and] he had a shepherd.

   c. *ī*   *sålål=ī*   *fakīr a*
      this shepherd=PC.3SG poor COP.PST.3SG
      This shepherd of his was poor.

   d. *ī*   *fakīr*   *do tå*   *čuk=ī*   *hast=ī=a*
      DEM1 poor_man two piece child=PC.3SG exists=PC.3SG=COP.PST.3SG
      o ya zã
      and one woman
      This poor man had two children, and one wife.'

2.3 Complex sentences

The subordinator *ke* introduces any kind of subordinate clause – i.e. relative (11), complement (13a), adverbial – as well as quoted speech (8a). The subordinate generally follows the matrix clause. Owing to its clitic nature (see Section 1.3), *ke* is attached to the first stressed element or to the first multi-word constituent if the subordinate precedes the matrix clause, as in (12a), where it even interrupts the noun phrase 'as many guns'.

(11) North Bashkardi (F:24)
    hamå best-ōn   *sorx ke* *bå*   *åteš sorx en*
    DEM2 pebble-PL red SUB with fire red COP.3PL
    'those red-hot pebbles, which are red-hot from the fire'

(12) South Bashkardi (E:33f.)

   a. *har-kader ke tofak=an*   *hat*
      any-extent SUB gun=PC.1PL COP.PST.3SG
      'Insofar as we had guns (any amount of guns that we had),

b. *doulat=ī   a   dā-y   a-bert-om*
State=PC.3SG from hand-HI IPFV-carry.PST-1PL
the State kept taking [them] from our hand.'

(13)   South Bashkardi (D:13f.)

a. *hål   a-xåh-om*
now IPFV-want.PRS-1PL
ke   ra-m   bå   ǰamīat=e   xailī=e   xo
SUB go.PRS-1PL with population=EZ very=EZ REFL
'Now we want (that) we go with a large group of our [people]

b. *dar bandir-abbås šekåyat=ī   kan-om*
in   PN   complaint=SPC do.PRS-1PL
[and] lodge a complaint in Bandar Abbas.'

When *ke* is combined with a nominal (e.g. '[the] time SUB' = 'when'), this neo-conjunction (borrowed or calqued on Persian models) is accentuated and introduces the subordinate clause. However, this is rare in the data.

Indeed, chains of main clauses are often preferred to overt marking of subordination, as is demonstrated by (10) and (16). These passages also show that Tail-Head-Linkage is frequent. Repeated material often appears in a different word order, e.g. with old information placed in front of the verb (10b, 10c, 16b).

2.4 Word order in main clauses

2.4.1 Generalities

Like other Ir. languages, Bashkardi main clauses with NPs as constituents are for their majority SOV. However, other word orders are not rare, and this in spite of the absence of marking of noun phrases (see Section 1.3). In the data used for this project, Bashkardi shows 30% of non-subject elements in postverbal position. This count includes nouns and pronouns (following the WOWA project's concept, pronominal clitics are not taken into account in the discussion and percentages to follow), with or without adpositional elements, but excludes clauses (e.g. complement clauses, which could likewise be considered as direct objects).

Information structure seems to play an important role for word order, and the status of an element as new or old information may override the SOV pattern and also the preferences for the placement of a particular element mentioned below (see also Sections 2.2 and 2.3).

Many of the phenomena discussed below are also found in other Iranian languages, but references are limited to other chapters in this volume.[17]

2.4.2 Direct objects

Overall, 21% of all direct objects are in postverbal position (Table 1). Compared to other Iranian languages in the WOWA sample (cf. Haig et al. 2024 [this volume]), Bashkardi has the highest frequency of nominal postverbal direct objects.

Table 1: Position of direct objects (n / %)

	Preverbal		Postverbal		All positions	
Nominal	198	(77%)	58	(23%)	256	(100%)
Pronominal	30	(94%)	2	(6%)	32	(100%)
All direct objects	228	(79%)	60	(21%)	288	(100%)

2.4.2.1 Nominal direct objects

Nominal direct objects (see also Section 2.2) are predominantly preverbal (4, 7b, 9, 16b, 16e). However, postverbal direct objects such as in (6a) and (14) make up 23%, which is an important percentage, even if significantly less than the overall average of 30% mentioned for postverbal non-subject elements in Section 2.4.1.

(14) North Bashkardi (K:46)
  hamē čürak=e šēr   zar=ī   ya gart=ī
  DEM1 kid=EZ   lion hit.PST=PC.3SG one roar=SPC
  'The lion cub gave a roar.'

2.4.2.2 Pronominal direct objects

Pronominal direct objects only occur under specific pragmatical conditions, since being expressed by a pronoun would imply that the referents are known, in which situation pro-drop applies (see Section 1.3); this is the case for the knife in (6b), the guns in (12b), the goat kid in (15a) and the talk in (17c). When they

---

[17]Percentages in the text to follow are rounded, and refer to North and South Bashkardi taken together.

do occur, they are nearly always in preverbal position, and indefinite pronouns always so (15a, 17a).

(15)   North Bashkardi (G:29)

    a.   *har-čī*   *mon a-g-om:*   *ma-koš-Ø,*
        any-thing I   IPFV-say.PRS-1SG PROH-kill.PRS-IMP.2SG
        'However much I say: "Don't kill [it] (the goat kid)!",

    b.   *å*   *a-koš-i=e*
        DEM2 IPFV-kill.PRS-3SG=PC.3SG
        he will kill it.'

2.4.3 Targets

Targets, a cover-term for elements indicating the end point of an action or event (see Table 2 and Korn 2022b: 90), occur postverbally in more than 50% of all instances, which is significantly more than the 30% found across all non-subject elements. Pronouns are rarely used in these functions; they do not share the postverbal tendency.[18]

Table 2: Postverbalness of Targets[19]

	total n	of which postverbal		pronouns/ of which postverbal		
Goals of motion	102	74	(73%)	2	1	(50%)
Goals of caused motion proper	25	18	(72%)	3	0	(0%)
'put'-expressions	22	8	(36%)	1	0	(0%)
Recipients	35	21	(60%)	16	5	(31%)
Beneficiaries and rec-ben[20]	16	6	(38%)	9	1	(11%)
Addressees	7	2	(29%)	2	0	(0%)
Final states	39	6	(15%)	1	0	(0%)
Sum	246	135	(55%)	34	7	(21%)

---

[18]The non-pronominal instances include nouns and adverbial expressions (e.g. 'here, to this place'), which are common for Targets.

As for all languages discussed in Chapter 1 (cf. Haig et al. 2024 [this volume]), *Goals of verbs of motion* show the highest postverbal percentage. Preverbal instances are often part of a Tail-Head-Linkage chain (see Section 2.3), as in (16b).

(16) South Bashkardi (A:2–8)

    a. *sabåh*   *a-rra-īn*   *gaverx*
        morning ɪᴘꜰᴠ-go.ᴘʀꜱ-1ꜱɢ ᴘɴ
        'In the morning I go to Gaverx.

    b. *gaverx a-rra-īn*   *čūr xom būr a-kan-īn*
        ᴘɴ   ɪᴘꜰᴠ-go.ᴘʀꜱ-1ꜱɢ four date load ɪᴘꜰᴠ-do.ᴘʀꜱ-1ꜱɢ
        I go to Gaverx (When I have arrived in Gaverx...), I load four [loads of] dates.

    c. *a-p-īn*   *ba mahala barå_i amīrī*
        ɪᴘꜰᴠ-come.ᴘʀꜱ-1ꜱɢ to home   for   ᴘɴ
        I come [and bring them] home for Amiri.

    d. *az_bād_e a-rra-īn*   *mahale=i ahmad=i madī*
        after_that ɪᴘꜰᴠ-go.ᴘʀꜱ-1ꜱɢ home=ᴇᴢ ᴘɴ=ᴇᴢ   ᴘɴ
        Then I go the house of Ahmad Mahdi.

    e. *ahmad=i mādī or-gir-īn*   *a-p-īn*   *dah gare=a*
        ᴘɴ=ᴇᴢ   ᴘɴ   up-take.ᴘʀꜱ-1ꜱɢ ɪᴘꜰᴠ-go.ᴘʀꜱ-1ꜱɢ into ᴘɴ=ᴅɪʀ[21]
        I take Ahmad Mahdi [and] I come to Gaverx.'

Targets of expressions meaning 'put' (18) pattern like the overall average. They thus differ from *Goals of caused motion* such as 'bring', 'send', which pattern like Goals of motion (17c).

(17) South Bashkardi (A:47f.)

    a. *to*   *heč*   *a-n-k-en*
        you.ꜱɢ nothing ɪᴘꜰᴠ-ɴᴇɢ-do.ᴘʀꜱ-2ꜱɢ
        'You don't do anything

---

[19]The n here refers to the total number of the category, e.g.: there are 102 goals of motion in the data, of which 74 (73%) are postverbal; 2 of the 102 instances are pronouns, of which one is postverbal.

[20]"rec-ben" is a category for items where it is difficult to decide whether they are recipients or beneficiaries.

[21]It is not clear whether *gare-a* refers to Gaverx or to another place.

    b. *ɣeir=e   gap-an=e   yamah a-čīn-e*
      besides=ᴇᴢ talk-ᴘʟ=ᴇᴢ we   ɪᴘꜰᴠ-collect.ᴘʀꜱ-2ꜱɢ
      but collect our talk,

    c. *o   or-gir-e*
      and up-take.ᴘʀꜱ-2ꜱɢ

      o   a-rr-e   a-bar-e   ba tehrūn
      and ɪᴘꜰᴠ-go.ᴘʀꜱ-2ꜱɢ ɪᴘꜰᴠ-carry.ᴘʀꜱ-2ꜱɢ to ᴘɴ
      take [it] and go [and] bring [it] to Tehran.'

(18)   North Bashkardi (F:32)
     å   wurå dega a-hr-end=eh
     ᴅᴇᴍ there again ɪᴘꜰᴠ-put.ᴘʀꜱ-3ᴘʟ=ᴘᴄ.3ꜱɢ
     'Then they put it (the bread) there (aside).'

Recipients are somewhat Goal-like in terms of placement, while *Beneficiaries* pattern like Goals of 'put'.

Addressees are rare, as they are usually retrievable from the context (see Section 1.3). Where they are expressed (8a), they pattern like the overall average.

A postverbal example of postverbal *final state* ('turns into a stone / is turned into a stone') is 'snakes' in (19), which is even marked by a preposition, recalling a similar pattern in Kurdish (see Haig 2022). However, this position is rare in the data.

(19)   North Bashkardi (K:129)
     hamå mål=ī   kolliya büd-e   vå   mår
     ᴅᴇᴍ2 cattle=ᴘᴄ.3ꜱɢ entirely become.ᴘꜱᴛ-ᴘʀꜰ with snake
     '… [and] that cattle of his had all become snakes.'

2.4.4 Other obliques

Non-subject elements other than direct objects and Targets occur mostly preverbally, but all categories do show postverbal instances (Table 3).

Instrumentals and comitative expressions are predominantly found following the verb (6a), while the preverbal position also occurs (13a).

Locational expressions are found both preceding (13b) and following the verb (5). Even more strongly preverbal are predicative expressions of location ('X is in the house / on the table') and other predicatives ('X is green / Y is my uncle').

The same applies to ablative-like expressions (12b).

Table 3: Postverbalness of other obliques[19]

	total n	of which postverbal		pronouns/ of which postverbal		
Instrument and Comitative	12	8	(67%)	2	2	(100%)
Location	16	6	(38%)	—		
Predicative location	13	2	(20%)	—		
Predicative	50	5	(10%)	7	0	(0%)
Possessum	47	6	(13%)	5	0	(0%)
Ablative, Cause and Stimulus	32	3	(9%)	2	0	(0%)
Time	62	3	(5%)	—		
Other	49	12	(24%)	2	0	(0%)
Sum	281	45	(16%)	18	2	(11%)

Patterning similarly are possessed items, which in the WOWA project refers to the X in the *mihi est* pattern 'to me is/exists X = I have X' (see Section 2.1.1.2).[22]

Temporal adverbials are found in clause-initial position in the vast majority of cases (9, 7a, 13a, 16a, 16d); their placement thus is not so much a question of pre- vs. postverbal.

3 Areal features

As mentioned in Section 2.4.1, Bashkardi shares the basic SOV order and other features with other Iranian languages. Its basically head-initial structure of nominal and adpositional phrases is shared with Persian, which obviously has influenced all other languages spoken in the country (and beyond). In other respects, however, Bashkardi is closer to Balochi, by which it is geographically surrounded (and with which it must surely have been in contact for centuries), for instance in the preservation of (some type) of (post-)ergativity. Bashkardi differs from both Persian and Balochi (but somewhat agrees with late Middle Persian) in the absence of marking of arguments, since the prepositions and postposed elements mentioned in Section 2.1 are used only sporadically to mark the direct or indirect object or the Goal of movement.

---

[22]The possessor in the *mihi est* construction could alternatively be interpreted as recipient or beneficiary, which would affect the statistics.

Some of the features just mentioned are discussed in Korn (2022a), which also suggests the possibility that only (some of) North Bashkardi is of the Persian sub-branch (South West Iranian) to which Bashkardi has been held to belong, while (some of) South Bashkardi could be a Balochi dialect historically (see Section 1.1). In this case, both head-final and head-initial noun phrase structures could be inherited in Bashkardi, contributing to the variation found in the data.

Seeing that word order is rather free in many Iranian languages (including colloquial Persian, but excluding more strictly verb-final standard Persian), I assume the same freedom for their Middle Iranian predecessors. I also think that a tendency towards the postverbal position of Goals of verbs of movement (and maybe some other types of Targets) could be inherited from Middle Iranian.[23]

There is a rather far-reaching agreement of Bashkardi word order with the South Balochi data investigated in Korn (2022b), while there are also some points of difference. A systematic comparison of the two languages will be the topic of a separate article, which will also discuss the influence of weight, flagging, etc. to the placement of non-subject elements.

Abbreviations

COP	copula	NP	noun phrase
DEF	marker of definiteness (see 2.1.2)	PC	pronominal clitic (see 1.3)
DEM	demonstrative pronoun (1: proximal, 2: distal)	PL	plural
		PN	name
DIR	directional clitic (see 2.1.3.2)	POSS	possessor clitic (see 2.1.1.2)
EZ	ezâfe (see 2.1.1)	PRF	perfect (participle)
HI	hiatus-bridging consonant	PROH	prohibitive prefix
		PRS	present stem
IMP	imperative	PST	past stem
IPFV	imperfective prefix	REFL	reflexive pronoun
MOD	modal prefix (for subjunctive and imperative)	SBsh.	South Bashkardi
		SG	singular
		SPC	marker of specificity (see 2.1.2)
n	number of instances		
NBsh.	North Bashkardi	SUB	subordinator (see 2.3)
NEG	negation	TAM	tense-aspect-mood

---

[23]Cf. Korn (2022b: 122), and see Jügel (2022) for a study of word order in Middle Iranian. See also Nourzaei & Haig 2024 [this volume]: Section 6 for discussion.

Acknowledgements

References

Barbera, Gerardo. 2005. *Lingua e cultura a Minab (Iran sud-orientale). Profilo grammaticale, testi e vocabolario*. Naples: Università degli Studi di Napoli "L'Orientale." (Doctoral dissertation).

Barbera, Gerardo. 2020. A Minâbi text and grammatical sketch. In Mauro Maggi & Mohsen Ashtiany (eds.), *A turquoise coronet: Studies in Persian language and literature in honour of Paola Orsatti*, 1–13. Wiesbaden: Reichert.

Barbera, Gerardo. 2023. The cannibal sister: Linguistic annotation of a South Bashkardi folktale. *Folia Orientalia* 60. 145–187. DOI: 10.24425/for.2023.148422.

Barbera, Gerardo. 2024. A fool's story in a South Bashkardi sub-dialect text. In Gerardo Barbera, Matteo De Chiara, Alessandro Del Tomba, Bhikkhunī Dhammadinnā, Federico Dragoni & Paola Orsatti (eds.), *Siddham. Studies in Iranian philology in honour of Mauro Maggi*, 1–18. Wiesbaden: Reichert.

Gershevitch, Ilya. 1959. Travels in Bashkardia. *Journal of the Central Asian Society* 46. 213–225. DOI: 10.1080/03068375908731669.

Haig, Geoffrey. 2022. Post-predicate constituents in Kurdish. In Yaron Matras, Geoffrey Haig & Ergin Öpengin (eds.), *Structural and typological variation in the dialects of Kurdish*, 335–377. Cham: Springer International Publishing. DOI: 10.1007/978-3-030-78837-7_8.

Haig, Geoffrey, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils Schiborr. 2024. Post-predicate elements in the Western Asian Transition Zone: Data, theory, and methods. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 3–54. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266331.

Jügel, Thomas. 2022. Word order variation in Middle Iranic: Persian, Parthian, Bactrian, and Sogdian. In Hiwa Asadpour & Thomas Jügel (eds.), *Word order variation – Semitic, Turkic and Indo-European languages in contact*, 39–62. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110790368-003.

Korn, Agnes. 2015. Voices from far away: Bashkardi material at the Ancient India and Iran Trust. *INDIRAN: Newsletter of the Ancient India and Iran Trust* 10. 4–5. https://indiairantrust.files.wordpress.com/2015/07/indiran10forwebsite.pdf.

Korn, Agnes. 2017. Notes on the nominal system of Bashkardi. *Transactions of the Philological Society* 115. 79–97. DOI: 10.1111/1467-968X.12087.

Korn, Agnes. 2020. Once upon a time: Telling a story in Bashkardi. *INDIRAN: Newsletter of the Ancient India and Iran Trust* 14. 8–9. https://www.indiran.org/wp-content/uploads/2020/11/INDIRAN-14-web-version.pdf.

Korn, Agnes. 2021a. A Bashkardi version of the Chain Tale ATU 2034: The Jackal Retrieves His Tail / Цепочная сказка ATU 2034: Версия башкарди «Шакал вновь обретает свой хвост». *Rodnoj jazyk / Родной язык* 2021/1. 368–383. DOI: 10.37892/2313-5816-2021-1-368-383.

Korn, Agnes. 2021b. Greeting a guest in Bashkardi. In Irina Nevskaya, Hatice Şirin & Ferruh Ağca (eds.), *Marcel Erdal Armağanı / Marcel Erdal's Festschrift: Journal of Turkish Studies / Türklük Bilgisi Araştırmaları, Special edition vol. I*, 301–323. Cambridge, MA. https://shs.hal.science/halshs-03671438.

Korn, Agnes. 2022a. Areal features in the languages of South Iran: Focus on Balochi and Baskhardi. Paper presented at the *Iranian Languages and Linguistics Lecture Series*, Toronto. https://halshs.archives-ouvertes.fr/halshs-03622909.

Korn, Agnes. 2022b. Targets and other post-verbal arguments in Southern Balochi: A multidimensional cline. In Hiwa Asadpour & Thomas Jügel (eds.), *Word order variation – Semitic, Turkic and Indo-European languages in contact*, 89–128. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110790368-005.

Korn, Agnes & Ilya Gershevitch. 2023a. Bashkardi (Northern): Recordings by Ilya Gershevitch (1914–2001), transcription and analysis by Agnes Korn. In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation*. Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

Korn, Agnes & Ilya Gershevitch. 2023b. Bashkardi (Southern): Recordings by Ilya Gershevitch (1914–2001), transcription and analysis by Agnes Korn. In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation*. Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

Korn, Agnes & Maryam Nourzaei. 2019. Notes on the speech of the Afro-Baloch of the southern coast of Iran. *Journal of the Royal Asiatic Society* 29. 623–657. DOI: 10.1017/S1356186319000300.

Korn, Agnes & Murad Suleymanov. 2017. Contact-Induced Progressives in Iranian Languages: A Typological Comparison of Caucasian Tat and Bashkardi. Paper presented at the Conference "Historical Linguistics of the Caucasus", Paris. https://hal.archives-ouvertes.fr/hal-01516383.

Mohammadirad, Masoud. 2024. Zagros region: The Kurdish-Gorani continuum. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 245–279. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266347.

Nourzaei, Maryam & Geoffrey Haig. 2024. Balochi: A cross-dialect investigation of post-verbal elements. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 121–154. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266337.

Pelevin, Mikhail. 2010. Materials on the Bandari dialect. *Iran and the Caucasus* 14. 57–78. DOI: 10.1163/157338410X12743419189388.

Seddiqi Nejad, Bakhtiar. 2010. *Bar-resī-ye touṣīfī-ye sāxt-vāže-ye gūyeš-e bašākardī: A descriptive study of Bashakardi dialect morphology*. Zahedan: University of Sīstān-va Balūčestān 1389 h.š. (MA thesis).

Skjærvø, Prods O. 1989a. Baškardi. In *Encyclopædia Iranica*, vol. III, 846–850. DOI: 10.1163/2330-4804_EIRO_COM_6720.

Skjærvø, Prods O. 1989b. Languages of Southeast Iran: Lārestānī, Kumzārī, Baškardī. In Rüdiger Schmitt (ed.), *Compendium linguarum iranicarum*, 363–369. Wiesbaden: Reichert.

Voskanian, Vardan & Vahe Boyajian-Sureniants. 2007. Lexical gleanings from Bašākard. *Iran and the Caucasus* 11. 121–125. DOI: 10.1163/157338407X224950.

Chapter 6

Kholosi

Maryam Nourzaei

Uppsala University

This chapter studies the word order configuration of Kholosi, an Indo-Aryan out-
lier spoken in the Southwest of Iran. Kholosi shows considerable influence of Per-
sian and other western Iranian languages, including phonology and morphology,
but also in syntax: It exhibits regular OV, but predominantly post-posed goals in
a manner that matches that of the neighbouring Iranian languages. Kholosi ap-
pears to have converged more closely with Iranian than other Indo-Aryan outliers
in Iran have (e.g. Jadgali). The Kholosi case illustrates how languages that have
shifted comparatively recently into the Western Asian Transition Zone can adapt
their syntax to match the profile of neighbouring languages.

1 Introduction

Although Kholosi is spoken in Iran, it belongs to the Indo-Aryan branch of Indo-
European, and is thus separated from the core location of its closest relatives by
hundreds of kilometres. This chapter provides the first available analysis of word
order in Kholosi, and explores the respective roles of inheritance and contact in
shaping its syntax. Kholosi is predominantly verb final, but exhibits mixed ad-
positional typology, with age-graded variability: younger speakers mix Iranian
prepositions and Indo-Aryan postpositions, while the older speakers produce
only postpositions. Unlike its Indo-Aryan relatives (Dahl & Stronski 2016), most
of which exhibit some form of split ergativity, Kholosi shows accusative align-
ment throughout, possibly through the influence of Persian and other contact
languages lacking ergativity. The data for the present study are extracted from
two tales which are available in the WOWA corpus (see full details in Nourzaei
2022a), and were supplemented by texts from my on-going analysis.

Kholosi is spoken mainly in two villages, Kholos and Gotav, in Hormozgan province, Iran (see Figure 1). Additional small speech communities are found in other regions such as Bastak, Jenah, Bandar Abbas, Bandar Lenge and Bandar Khamir. In addition, Kholosi is also spoken by Kholosi people living outside of Iran for example in Bahrein, Doha, Dubai, and Abu Dhabi. The total population of the Kholosi speakers is uncertain. However, based on recent field studies by myself, the number of speakers is estimated as at least 2,300.

Figure 1: Location of Kholosi Speakers (from Nourzaei 2023)

There is another Indo-Aryan language spoken in Iran, Jadgali, located mainly in Chabahar, Dashtiyari, and Polan regions in Sistan and Balochistan province (Barjasteh Delforooz 2008). The distance between the Kholosi community to the Jadgal communities is more than 600 km. The Kholosi and Jadgali speakers have not traditionally had any contact and only became acquainted with each other recently (see Nourzaei 2023). The location of the Jadgali speech communities is indicated in Figure 2.

The question of how the Kholosi reached their current location, and what their historical connection to the Jadgali speech community may have been, remains controversial. There are four different accounts: (a) they directly moved

to their present homeland from India (see Anonby & Mohebbi Bahmani 2016); (b) they moved here via Dashtiyari in Balochistan; (c) they migrated to their present homeland from Sindh during the Safavid dynasty; or (d) they migrated from the Makoran coast as a separate group, distinct from the Jadgal.

Figure 2: Location of Kholosi and Jadgali speakers (taken from Nourzaei 2023)

The area where Kholosi is spoken is linguistically highly diverse, and Kholosi speakers are multilingual. Contact languages include two different language families: Indo-European (Iranian) and Semitic. Kholosi speakers in Iran are in direct contact with Bastaki, Lārestāni (both West Iranian) as local vernaculars, and Persian as the official language of Iran via TV, Radio, and education, and Arabic as a liturgical language via the Koran and Islamic literature. Likewise, Kholosi speakers outside of Iran are in contact with Arabic as an official language (see also Anonby & Mohebbi Bahmani 2016: 2). In addition, Kholosi would have been in contact with other languages during the migration of the Kholosi speakers to their current location, but the identity of these languages has not been established with any certainty.

Kholosi is used as the first language in the Kholosi community. Parents speak Kholosi with their children to a large extent. In Gotav village, however, there is a tendency for the Kholosi residents to use Bastaki among themselves and with their children. In Kholos, many parents speak with their children in Persian to prepare them for school. Outside these two villages, e.g., Buchir, it is observed that Kholosi speakers have lost their language and have switched to

Bastaki/Lārestāni (known as Achomi, Khodemuni), which are the vernacular languages in these regions. In Kholosi families with an exogamic marriage pattern, the Kholosi parents do not use Kholosi as their first language. Instead, the common language between parents and children at home is Bastaki, Lārestāni, or even Persian. Among Kholosi speakers living abroad, a tendency to use Arabic at home is reported. Note that some of the Kholosi speakers switched to Arabic.

Kholosi lacks an established writing system, and is generally restricted to oral domains. For example, speakers use standard Persian when texting each other via cell phone or writing letters. There are also no TV or radio programmes in Kholosi. The language of teaching is Persian; however, if the teacher is a Kholosi speaker, Kholosi can be used in the classroom. For religious instruction, Arabic and Persian are used. In the past, Kholosi served as a language of religious instruction in Islamic schools; however today, only Persian is used. Their religious leaders (Mullah) use Persian for their sermons after Friday and Ids (festival) prayers.

1.1 Previous studies

Kholosi is an under-documented language (see Nourzaei 2022a). The first paper (a questionnaire-based investigation) on Kholosi was published by Anonby & Mohebbi Bahmani (2016), which confirmed that Kholosi is an Indo-Aryan language. In addition, Anonby et al. (2019) published a short paper in Encyclopædia Iranica. The first Kholosi text to be transcribed, glossed and translated into English in FLEX has been published by myself, and includes a sociolinguistic study of Kholosi Nourzaei (2023). I have also written a grammar sketch (Nourzaei 2024), and an outline of the Kholosi nominal and pronominal system is under review. Arora & Etebari (2021), Nourzaei (In press) published online a list of Kholosi lexical items, based on English words presented to one speaker via the medium of Persian translation and through an online messenger service. The present study is the first contribution to the word order in Kholosi, with special focus on postpredicate elements.

1.2 Data and methodology

The background analysis for this paper is based on the ongoing documentation described above, while the quantitative data stems from two tales entitled "Untidy fox," recorded from a 32-year-old female speaker from Kholos in 2020, and "Prophet Musa," recorded from a male speaker from Kholos in 2017 and available in the WOWA database (Nourzaei 2022b). I have also included data from

a growing corpus of Kholosi narrative speech including folktales, biographical tales, and procedural texts spoken by one male and two female Kholosi speakers from Kholos and Gotav. The speakers have different social backgrounds and are between 32 and 82 years of age. All speakers are fluent in Persian and vernacular languages such as Bastaki and Lārestāni, and can read Arabic very well. One of the female speakers is an Arabic teacher at an Islamic school in Kholos, and the male speaker is fluent in Arabic. The texts were recorded in WAV and MP3 format. They were then imported into ELAN software,[1] transcribed phonemically, and double-checked by the linguistic consultants. After that, a morpheme-by-morpheme glossing was carried out using FLEx software.[2] Finally, a free clause-by-clause translation of the texts was produced.[3]

1.3 Word order in Indo-Aryan

Like the Iranian languages, the Indo-Aryan languages show an OV order in the clause. Noun phrases across all of IA are left-branching (Modifier + N) Masica (1991: 370). Examples from Sindhi, one of Kholosi's presumably closest Indo-Aryan relatives, illustrate this feature: *vada vātō* 'big mouth'; *sāfu dīle* 'pure heart'; *čau darō* 'four doors' and *ghara ǰō dhaṇī* 'house-of the master' i.e. the master of the house (Trumpp 1872: 88, 119). Masica (1991: 218) states that across Indo-Aryan, primarily nouns, pronouns, adjectives, and in some languages, also certain numerals and adverbs may inflect for gender, number, case, and definiteness. While case is found in all IA languages, gender is not universal, and a grammatical marker for definiteness characterizes only certain languages. Case, number and gender are noted for Sindhi (Trumpp 1872: 119, 131), but reliable information on definiteness is not available. Masica (1991: 373) mentions that for all VP arguments, e.g. the direct and indirect objects, goals and source of motion, all occur to the left of the main verb stem (except in aberrant Kashmiri). Examples (1-2) from Gujarati and Sinhalese demonstrate preverbal Goal and direct object in turn, glosses added.

(1) Gujarati (Masica 1991: 352)
     māre/mane ghɛr   javānũ che
     1SG.OBL   home go.INF COP.3SG
     'I have to go home.'

---

(2)   Sinhalese (Masica 1991: 333)
      mama bat kanavā
      1SG   rice eat.PRS
      'I eat rice.'

We are unaware of any corpus-based investigation of post-posed elements in Indo-Aryan that might assist in reconstructing the ancestor word order of Kholosi. Preliminary consideration of available material, such as Liljegren (2016) on Palula (Indo-Aryan, Dardic) indicate a dominant head-final VP, and in the experimental investigation of Hindi word order by Patil et al. (2008), only verb-final variants are considered. In the absence of evidence to the contrary, I will therefore assume that the Indo-Aryan relatives of Kholosi are consistently verb-final, from which we can provisionally infer that the ancestor language of Kholosi was likewise verb-final.

2 Some elements of Kholosi grammar

2.1 Alignment

Some basics are needed as background information for the discussion which follows. Kholosi exhibits accusative alignment with all verbal categories, as in Persian. The subject is canonically in the nominative case and the object in the oblique case. The verb agrees with the subject both in number and person. Note that in the past domain, transitive and intransitive verbs have different sets of endings, which presumably reflects an earlier ergative alignment in Kholosi. The following examples present accusative alignment for transitive and intransitive verbs in the present (3-4) and past (5-6) domains.

(3)   Kholosi (Nourzaei 2022b, B,0234)
      sandūǧ=e čūb-ī   deres   ker-aw
      box=INDV wood-ADJV complete do.PRS-3SG
      'He makes a wooden box.'

(4)   Kholosi (Nourzaei 2022b, B, 0160)
      yak hāro=e   denyā t=eǰ-aw
      one boy=INDV world to=come.PRS-3SG
      'A boy will be born (lit. comes to the world).'

(5) Kholosi (Nourzaei 2022b, A, 0019)
hoko roz robā šamšer=es   pale pord-ū
one day fox   sword=PC.3SG PREV search.PST-3SG
'One day the fox, (he) searched for his sword.'

(6) Kholosi (Nourzaei 2022b, A, 0005)
robā nōk-ō   dar he   sere   mā vađ-ō tʰo
fox.M small-M in textscprox house in   big-M become.PST.3SG
'The small fox grew up in this house.'

2.2 Morphological case system

Kholosi shows three morphological cases: Direct (unmarked), oblique *=ke*, and genitive *=ǰō/ǰī* cases.[4] The genitive case is partially replaced with Iranic Ezafe construction (see for more details Nourzaei 2024, and Nourzaei In press). Kholosi has Differential Object Marking, though the exact nature of the triggering factors has not yet been established. Indefinite and inanimate objects such as 'wooden box' in (3) are unmarked for case, while other objects can be marked with *=ke*, as in (25).

2.3 Person marking clitics

Similar to Iranic spoken in this region, e.g. Koroshi, Kholosi has a full set of person-marking clitics, which have different functions, e.g., possession, direct and indirect object markings. The person-marking clitics in Kholosi are presented in Table 1. The person-marking clitics attach to various hosts: nouns, *kolāh=ās* 'his hat,' adverbs, *roz=e baad=os* 'its next day,' postposition, *ag=es* 'to him,' relational nouns, *ak=es* 'his front,' verbs, *'čaī=ves* 'says to him'. Note, some of these forms are copied from Iranic, such as 2nd person *=ū* and 3rd person *=es/eš*.

2.4 Morphological gender marker

In other respects, Kholosi differs from Iranic. This includes a grammatical gender distinction in some verb endings - in particular the 1st, 2nd and 3rd persons singular - numbers, and adjectives (preferably Indo-Aryan), comparative suffix

---

[4]It is currently unclear whether these 'case markers' should be considered to be some form of particle (i.e. postpositions), or phrasal affix, or clitic. They are provisionally treated here as clitics.

Table 1: Person-marking clitics in Kholosi

SG	1st	*=oe/=e/oy/=ī/yāe*
	2nd	*=o/ū/āv/āo/*
	3rd	*=ās/yās/os/es/eš*
PL	1st	*=om=ām=mām/em*
	2nd	*=om=ām*
	3rd	*=ān/on/en*

-r, imperfective suffix *-d*, and genitive suffix *-ǰ*, which distinguish masculine and feminine gender. Masculine adjectives are marked with (*o/ō*) and feminine adjectives with (*ī*). In (7), *-ō* on the adjectives *gahr* 'red' agrees with the river Nile, which is masculine. Gender agreement can be within the NP, as in (8), or between subject and predicate, as in (7); the details remain to be elucidated.

(7) Kholosi (Nourzaei 2022b, B, 0542)
    rūd=e   nīl   gahr-ō tʰ=ū
    river=EZ Nile.M red-M become.PST=COP.PRS.3SG
    'The river Nile turned red.'

Similarly, in (8), *-ī* on the adjective *kalt* 'big' agrees with 'dragon', which is feminine.

(8) Kholosi (UP)
    aždehā=ye   kalt-ī
    dragon.F=EZ big-F
    'a big dragon'

Gender is also marked directly on some nouns, with phonologically comparable suffixes to those outlined above: *nonō* 'grandfather' and *nonī* 'grandmother'; *čōrkō* 'boy' and *čōrkī* 'girl'; *hārrō* 'boy' and *herrī* 'girl'.

3 Word order in the NP and the clause

3.1 Adjective/noun

In the noun phrase, Kholosi has adopted some Persian syntactic features. For example, some adjectives may follow the noun and be linked to it via the 'ezafe,' as in (9), or occur without an 'ezafe', as in (10).

(9) Kholosi (Nourzaei 2022b, A, 0044)
moškel=e   vađ-o
problem=EZ big-M
'a big problem'

(10) Kholosi (Nourzaei 2022b, A, 0005)
rōbā nōk-ō
fox   small-M
'the small fox'

3.2 Possessor/possessed

Kholosi also follows the general Persian pattern in that the possessed precedes possessor, either with an intervening 'ezafe' (EZ), as in (11), or without, as in (12). Note the use of the genitive marker on the possessor in (12).

(11) Kholosi (UP)
laškar=e ferawn
army=EZ pharaoh
'pharaoh's army'

(12) Kholosi (Nourzaei 2022b, A, 0148)
māv   rūbāh kočūlū=ǰo
mother fox   small=GEN.M
'mother of small fox'

However, the data demonstrate some variation, and the possessor can also precede the possessed, as (13), which also illustrates the expected Indo-Aryan genitive marking of the possessor.

(13) Kholosi (Nourzaei 2022b, B, 0266)
hazrat   mūsā=ǰī   dođā
prophet Musa=GEN.F sister
'the sister of the prophet Musa'

Possession and similar relations have also been attested with juxtaposition, as *māre šahr* 'people of the city' which have no marking of possession in (14).

(14) Kholosi (Nourzaei 2022b, B, 0549)
*tamām=e sāher-ēn=ke   **māre** šahr daʔvat dī-yaw*
all=EZ   magician-PL=OBL people city invite give.PRS-3SG
'He invites all the magicians [and] people of the city.'

3.3 Demonstrative/noun

Demonstrative pronouns precede the head nouns without any linker.

(15)   a.   Kholosi (Nourzaei 2022b, A, 0005)
         he   *sere*
         PROX house
         'this house'
       b.   Kholosi (UP)
         ho   *motor*
         DIST car
         'that car'

3.4 Numeral/noun

In Kholosi, numerals precede head nouns. The head nouns show number and gender agreement with only Indo-Aryan numerals, as in the following examples.

(16)   a.   Kholosi (Nourzaei 2022b, B, 0160)
         hēk-ō hār-ō
         one-M boy-M
         'a boy'
       b.   Kholosi
         bahr-ā hēr-ā
         two-M boy-M.PL
         'two boys'

(17)   a.   Kholosi (UP)
         hīk-ī   čōrk-ī
         one-F girl-F
         'a girl'
       b.   Kholosi (UP)
         bahr-ī čōrk-ī-ū
         two-F girl-F-PL
         'two girls'

3.5 Adpositions

Kholosi has a mixed typology with respect to adpositions, with prepositions, postpositions, and combinations of the two. The most widely attested prepositions are obvious borrowings from Iranic, while postpositions are Indo-Aryan.

(18) shows a combination of an Iranic preposition and an Indo-Aryan postposition and (19) just Indo-Aryan postposition.

(18) Kholosi (Nourzaei 2022b, A, 0032)
az mān=ās **tāv** soāl ka-ī
from mother=PC.3SG from question do.PST-3SG
'he asked of his mother'

(19) Kholosi (Nourzaei 2022b, A, 0034)
mov=ās gen=ās **mā** pēr-ī
mother=PC.3SG room=PC.3SG in look.PST-3SG
'His mother looked inside his room.'

There is a tendency to use the Iranic preposition *dāxel* as a postposition in Kholosi, with the same meaning, as in (20)

(20) Kholosi (UP)
ham=e atte **dāxel** māyexamīr kaṛ-e
EMPH=PROX dough inside yeast do.PRS-2SG
'You add yeast to (lit. in) this dough.'

3.6 Auxiliary/main verb

Independent auxiliaries expressing tense, aspect, mood, or voice, are not attested in the present data; TAM categories are expressed by suffixes on the verb. There are no TAM prefixes that would parallel the widely-attested aspectual and modal prefixes of the Iranic system (Persian *mi-* and *be-*, for example). However, modal verbs such as 'want' precede the main verbs, as in Iranic (32), and this could well be considered a result of contact influence, in line with the claim that clause (or predicate) combining is a favoured domain for contact influence (Haig 2001).

3.7 Complement clause/matrix verb

The subordinators *ǰō* and *ke* 'that' introduce various subordinate clauses, including relative, complement and adverbial - as well as quoted speech. The particle *ke* is an Iranic borrowing. Similarly, to Persian, the complement clause follows the matrix verb in Kholosi, and complementizers generally occur in the initial position within the complement clause. In these respects, Kholosi entirely matches Persian and other West Iranian languages.

(21)   Kholosi (Nourzaei 2022b, B, 0457)
    čī-ya   jŏ   hat=ī   mā lat=e
    say.PRS-3SG CMP hand=PC.1SG in   stick=COP.PRS.3SG
    'He says, "There is a stick in my hand".'

(22)   Kholosi (Nourzaei 2022b, B, 0247)
    pesa-vān   jŏ   bale yak sandūġ=e hat
    see.PST-3PL CMP yes one box=INDV COP.PRS.3SG
    'They saw that, yes, there is a box [on the water].'

3.8 Light verb complements

Kholosi follows the widespread Iranic pattern of combining non-verbal elements with light verbs, in that order, to create complex predicates. The attested light verbs are 'become,' 'come,' 'do,' and 'give' as in examples (23) and (24).

(23)   Kholosi (Nourzaei 2022b, A, 0166)
    ferawn peǰal tāv bīdār tʰī-yaw
    pharaoh sleep from awake become.PRS-3SG
    'The Pharaoh gets up (lit. wakes up from sleep).'

4 VP constituents in the Kholosi WOWA data

In this section we exemplify the main combinations of non-subject constituents and verb, based on the quantitative analysis of the WOWA Kholosi data set (Nourzaei 2022b). An overview of the findings is presented in Table 3 below. Kholosi makes extensive use of pro-drop in different syntactic functions, so pronominal arguments are infrequent in the data. In line with general practice in the WOWA framework (see Haig et al. 2024 [this volume]), we will not consider the position of subjects.

4.1 Direct objects

In Kholosi, generally nominal direct objects occur in pre-verbal position, as in the following example. They may be separated from the verb by other constituents, as in (24), but are only very rarely post-posed after the verb.

(24) Kholosi (Nourzaei 2022b, A, 0139)
    šamšīr=os   *o=te*   *āvīzān ker-a-va*
    sword=PC.3SG there=to hang   do.PRS-3SG-SBJV
    'He should hang his sword in the other direction.'

    In the Kholosi data used for this project, only 2% of all direct objects are in postverbal position, which makes it even lower than obtained for most Iranian languages, which range between 3–10%. This may be a reflex of the Indo-Aryan origins, but the differences are certainly not great.

    Free pronouns as direct objects are rare in Kholosi (14 examples), of which all are pre-verbal, as in (25). Direct objects (and other objects) may also be expressed through a clitic pronoun on the verb, which then follows the verb, but as it is part of the same phonological word as the verb, these are not counted as 'post-verbal' in line with standard practice for WOWA. An example of a clitic object is provided in (26).

(25) Kholosi (UP)
    meskīn=e golī pʰepī=yāe   *mo=ke*   *vađ-o ka-ī*
    poor=EZ goli aunt=PC.1SG 1SG=OBL big-M do.PST-3SG
    'My poor aunt, Goli, raised me.'

(26) Kholosi (Nourzaei 2022b, B, 305)
    ferawn   *konǰān-do=sū*   *čembī dīv=es-ī-ya*
    pharaoh want.PST-IPFV=COP.PST.3SG kiss   give.PRS=PC.3SG-3SG-SBJV
    'Like this pharaoh was about to kiss him.'

4.2 Copula complements

Copulas are generally clitics which cliticize to the copular complement, and are thus generally clause-final, as in (27-28). This is in line with a near-universal structure for copular constructions across the Western Asian Transition Zone, though with exceptions at the western periphery (see Haig & Khan 2019, Haig et al. 2024 [this volume], Mohammadirad 2024 [this volume]).

(27) Kholosi (UP)
    pon=āe=ǰī   *mīṛī=ya*
    father=PC.1SG=GEN wife=COP.PRS.3SG
    'She is my father's wife.'

(28) Kholosi (UP)
čekada   zaīfā=e   čang-ī=ū
how_much woman=INDV kind-F=COP.PST.3SG
'She was a very kind woman.'

Note that in the present data, one example of a postposed complement of copular verb is attested (29):

(29) Kholosi (Nourzaei 2022b, B, 0230)
*hat   **ǰīnd-o***
COP.PRS.3SG alive-M
'He is alive.'

4.3 Goal/verb

In Kholosi, the largest number of post-verbal elements are goals of caused motion, such as 'throw', 'put,' and 'hit' (see Table 3) as in (30-31).

(30) Kholosi (Nourzaei 2022b, B, 0193)
hazrat=e   mūsā=ī   māv   petr=es=ke   lār-aw   tanūr mā
prophet=EZ musa=GEN mother boy=PC.3SG=OBL throw.PRS-3SG tanur in
'The Prophet Musa's mother throws her son into the Tanur.'

(31) Kholosi (UP)
kolan   lī-yāv=es   fer   mā
you_know put.PRS-1PL=PC.3SG oven in
'You know, we put it into the oven.'

Similarly, goals of simple motion verbs such as 'go,' 'come,' 'fall,' are often post-verbal (see Table 3).

(32) Kholosi (Nourzaei 2022b, A, 0024)
*ho   konǰān-do=sū   ven-a-va   **rafīk-en=des**   akaya*
DIST want.PST-IPFV=COP.PST.3SG go.PRS-3SG-SBJV friend-PL=PC.3SG front
'He wanted to go to his friends.'

(33) Kholosi (Nourzaei 2022b, B, 0366)
*eč-aw   **yek ǰā=e***
come.PRS-3SG one place=INDV
'He comes to a certain place.'

4.4 Recipient/verb

Similarly to goals, the majority of recipients also follow the verb, as in (34-35), though the absolute number of recipients in the data is quite low, cf. Table 3.

(34) Kholosi (Nourzaei 2022b, B, 0394)
 *tʰīo ponǰ=ās   dī-aw **payambar mūsā=ke***
 daughter REFL.GEN=PC.3SG give.PRS-3SG prophet Musa=OBL
 'He gives his daughter to the prophet Musa.'

(35) Kholosi (Nourzaei 2022b, B, 0205)
 *xebar dī-yen hazrat=e **mūsā=ī** **māo=ke***
 news do.PRS-3PL prophet=EZ Musa=GEN mother=OBL
 'They informed the Prophet Musa's mother.'

4.5 Addressee/verb

Nominal addressees in Kholosi precede the verb as in (36), but addressees expressed by a person marking clitic attach directly to the verb as in (37), or attach to a postposition as in (38).

(36) Kholosi (UP)
 am=te xodā=ke ča-yen
 this=at God=OBL say.PRS-3PL
 'At this point they said to God...'

(37) Kholosi (Nourzaei 2022b, B, 0520)
 payāmbar hazrat=e mūsā xayle kām-hā vet-ī=yās
 prophet prophet=EZ Musa very thing-PL say.PST-3SG=PC.3SG
 'The Prophet Musa said to him a lot of things.'

(38) Kholosi (Nourzaei 2022b, B, 0172)
 tā yek nafar manī=os   čī-yaw   ag=es
 till one person meaning=PC.3SG say.PRS-3SG to=PC.3SG
 'Till a person says its meaning to him.'

4.6 Complements of 'become'

In Kholosi complements of 'become' are preverbal, as in the following examples.

(39) Kholosi (Nourzaei 2022b, B, 0334)
ferawn motmaen tʰo   ǰo...
pharaoh relaxed   become.PST.3SG CMP
'The Pharaoh became relaxed that...'

(40) Kholosi (Nourzaei 2022b, B, 0483)
heždehā=te tʰī-aw
dragon=to become.PRS-3SG
'It will turn into a dragon.'

4.7 Place constituents and place constituents of a copular verb

Locative constituents of a copular verb i.e., ('X is in the garden') (41) precede the predicate.

(41) Kholosi (daily conversation)
Hasan dar sere   mā=e
hasan in   home in=COP.PRS.3SG
'Hasan is at home.'

4.8 Other obliques

Most other obliques such as instruments, ablatives, comitatives, and beneficiaries appear before the verb; see the figures for 'other' in Table 3 below. The following examples demonstrate various other semantic relations.

(42) Kholosi (Nourzaei 2022b, A, 0003)
hok-ō xānavāda hīnkī sānda zendegī kar-d=ayaū
one-M family   each other life   do.PRS-IPFV=COP.PRS.3PL
'A family used to live with one another.'

(43) Kholosi (Nourzaei 2022b, A, 0065)
xers bā   šamšīr=os   be havā=te be   zarba
bear with sword=PC.3SG with air=at   with beat
čīn-d=ū
beat.PRS-IPFV=COP.PST.3SG
'The bear was striking with his sword into the air.'

(44) Kholosi (Nourzaei 2022b, B, 0162)
yak mangāl=e az   ġawm=e banī   esrāīl tāv boland tʰo
one fire=INDV from tribe=EZ sons_of Israel from upright become.PST.3SG
'A fire rose from the people of Israel.'

5 Summary

So far, I have discussed various word order parameters, and the post-verbal place-ment of different constituents related to the verb. This historically head-final language shows some shifts towards head-initial syntax, leading to some degree of inconsistency in head-directionality (see Dryer 1992). We also note the occa-sional use of mixed head-final and head-initial constructions involving borrowed prepositions combined with inherited post-positions, see (18). The head-initial configurations which have been illustrated in this chapter are given in Table 2.

Table 2: Head-initial configurations

Head	Complement
Noun	Adjective
Possessed	Possessor
Matrix clause	Complement clause
Complementizer	Complement clause
Verb	Goal
Verb	Recipient

Within the VP, most kinds of verbal arguments remain consistently pre-verbal (90% or more), with the sole exception of goals and recipients. Table 3 shows the relevant figures for the most important roles, as identified in the WOWA framework (see Haig et al. 2024 [this volume]). Note that "Goals" subsumes both goals of verbs of motion and of caused motion (both of which are predominantly post-verbal in Kholosi).

Compared to its Indo-Aryan relative Jadgali, Kholosi exhibits greater influence of Iranic. Kholosi has adopted some features of Persian noun phrases, for exam-ple post-posed adjectives, and some prepositions, while Jadgali retains the Indo-Aryan structures. Jadgali also retains its split ergative alignment, while Kholosi shares the same alignment (accusative) with Persian, and makes more extensive use of person-marking clitics. Overall, it seems that Kholosi has undergone a greater degree of syntactic convergence with Iranian languages. Kholosi never-theless maintains grammatical features that distinguish it from (neighbouring) Iranic. These include a grammatical gender distinction on some numerals, verb endings, and adjectives (if the host item is Indo-Aryan), and also postpositions, which are a prominent feature of Indo-Aryan languages. Note that its morpholog-ical gender system is not strong as such (see Nourzaei Nourzaei In press). Unlike Iranian languages, Kholosi lacks any prefixal TAM elements.

Table 3: Percentages of post-predicate placement of different constituents in Kholosi

	Total number of tokens	preverbal	post-verbal	% post-verbal
Ablative	22	22	0	0%
Addressee	14	13	1	7%
Become	14	13	1	7%
Comitative	13	13	0	0%
Copular expression	41	39	2	5%
Direct object	67	65	2	3%
Definite direct object	88	86	2	2%
Goal	56	21	35	62%
Static location	43	42	1	2%
Other	128	126	2	2%
Possessed	12	12	0	0%
Recipient	8	2	6	75%
Totals	**506**	**454**	**52**	

At the level of the clause, Kholosi has adapted to the profile of the Iranian languages of Western Asia: Modal verbs precede main verbs, subordinate clauses follow matrix clauses, and complementizers occur clause-initially. One of the most salient features of these languages is the combination of near-categorical OV with a high frequency of post-verbal Goals. Kholosi likewise exhibits this combination (>90% OV, and >60% VG, cf. Table 3). Note, however, that rates of post-verbal Goals do not reach the levels found further westward in the Iranian languages of Mesopotamia. But Kholosi does share with several Iranian languages the spread of post-verbal placement to include recipients, entirely in line with the sequence predicted in Haig et al. (2024 [this volume]). Other types of constituents (such as locatives, copula complements, adverbs, addressees) show significantly lower rates of pos-verbal position.

Although Kholosi retains abundant evidence of its Indo-Aryan origins in the lexicon and in morphology, it exhibits syntactic convergence with areally contiguous languages, which has rendered it significantly different from its Indo-Aryan relatives located outside the Western Asian Transition Zone. The Kholosi case thus has considerable implications for assessing the role of contact versus inheritance in predicting word order.

Abbreviations

ADJV	adjectivizer	PC	person-marking enclitic
CMP	complementizer	PL	plural
COP	copula	PREV	preverb
DIST	distal	PROX	proximal deixis
EMPH	emphasis	PRS	present
EZ	ezafe particle	PST	past
F	feminine	REFL	reflexive pronoun
GEN	genitive	SBJV	subjunctive
IA	Indo-Aryan	UP	Unpublished text
IPFV	imperfective	WATZ	Western Asian Transition Zone
INDV	individuation clitic		
M	masculine	WOWA	Word Order in Western Asia
OBL	oblique case		

Acknowledgements

References

Anonby, Erik & Hassan Mohebbi Bahmani. 2016. Shipwrecked and landlocked: Kholosi, an Indo-Aryan language in South-West Iran. In Jila Ghomeshi, Carian Jahani & Agnes Lenepveu-Hotz (eds.), *Further topics in Iranian linguistics: Proceedings of the 5th International Conference on Iranian Linguistics, held in Bamberg on 24-26 August 2013,* Cahiers de Studia Iranica 58, Paris, 13–36.

Anonby, Erik, Hassan Mohebbi Bahmani, Maryam Nourzaei & Mortaza Taheri-Ardali. 2019. Kholosi. *Encyclopædia Iranica* XVI/5. 541–542.

Arora, Aryaman & Ahmed Etebari. 2021. *Kholosi dictionary*. https://aryaman.io/kholosi/Kholosi_Dictionary.pdf.

Barjasteh Delforooz, Behrooz. 2008. A sociolinguistic survey among the Jadgal of Iranian Balochistan. In Carina Jahani, Agnes Korn & Paul Titus (eds.), *The Baloch and others: Linguistic, historical and socio-political perspectives on pluralism in Balochistan*, 23–43. Wiesbaden: Reichert.

Dahl, Eystein & Krzysztof Stronski. 2016. Ergativity in Indo-Aryan and beyond. In Eystein Dahl & Krzysztof Stronski (eds.), *Indo-Aryan ergativity in typological and diachronic perspective*, 1–38. Amsterdam: Benjamins. DOI: 10.1075/tsl.112. 01dah.

Dryer, Matthew S. 1992. The Greenbergian word order correlations. *Language* 68. 81–138.

Haig, Geoffrey. 2001. Linguistic diffusion in present-day East Anatolia: From top to bottom. In A. Aikhenvald & R. M. W. Dixon (eds.), *Areal diffusion and genetic inheritance: Problems in comparative linguistics*, 195–224. Oxford: Oxford University Press.

Haig, Geoffrey & Geoffrey Khan. 2019. Introduction. In Geoffrey Haig & Geoffrey Khan (eds.), *The languages and linguistics of Western Asia: An areal perspective*, 1–29. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110421682-001.

Haig, Geoffrey, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils Schiborr. 2024. Post-predicate elements in the Western Asian Transition Zone: Data, theory, and methods. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 3–54. Berlin: Language Science Press. DOI: 10.5281/ zenodo.14266331.

Liljegren, Henrik. 2016. *A grammar of Palula*. Berlin: Language Science Press.

Masica, Colin P. 1991. *The Indo-Aryan languages*. Cambridge: Cambridge University Press.

Mohammadirad, Masoud. 2024. Zagros region: The Kurdish-Gorani continuum. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 245–279. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266347.

Nourzaei, Maryam. 2022a. *Documentation of Kholosi as an endangered language.* Bamberg. https://www.elararchive.org/dk0659/.

Nourzaei, Maryam. 2022b. Kholosi (Kholos). In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* Bamberg: University of Bamberg. multicast.aspra.uni-bamberg. de/resources/wowa/.

Nourzaei, Maryam. 2023. On orality and the sociolinguistic situation of the Kholosi community. *Orientalia Suecana* 72. 47–68.

Nourzaei, Maryam. 2024. Kholosi. In Charles Häberl (ed.), *Language diversity in Iran: New texts and perspectives from non-Iranian languages.* Berlin: De Gruyter Mouton. DOI: 10.0000/9783110641578.

Nourzaei, Maryam. In press. On nominal and pronominal morphosyntax in Kholosi. *Studia Iranica.*

Patil, Umesh, Gerrit Kentner, Anja Gollrad, Frank Kügler, Caroline Féry & Shravan Vasishth. 2008. Focus, word order and intonation in Hindi. *Journal of South Asian Linguistics* 1. 55–72.

Trumpp, Ernst. 1872. *Grammar of the Sindhi language.* London: Trübner.

Chapter 7

Post-predicate elements in modern colloquial Persian: A multifactorial analysis

Mohammad Rasekh-Mahand[a], Elham Izadi[a], Mehdi Parizadeh[a], ⓘ Geoffrey Haig[b] & Nils Schiborr[b]

[a]Bu-Ali Sina University, Hamadan [b]Bamberg University

We investigate post-verbal elements in contemporary spoken Persian, based on the HamBam corpus (Haig & Rasekh-Mahand 2022), and compare the results with Frommer (1981). We apply two multi-variate analyses to the HamBam data (logistic regression, gradient boosting), which suggest semantic/syntactic role (e.g., Goals, direct objects) is the primary predictor of post-verbal placement; other factors, such as weight, are marginal. Our findings confirm those of Frommer (1981) for the least formal spoken registers of Persian (>80% rates of post-verbal Goals). However, we detect a shift in register distribution in today's spoken language compared to the late 1970s.

1 Introduction

Persian has some odd features regarding its word order typology. It has prepositions (though the object marker is an enclitic *=rā*), and post-nominal adjectives, genitives and relative clauses. These are features generally associated with the head-initial languages (Dryer 1992). Yet the verb occurs in the final position of the clause, especially in written and formal registers (Faghiri et al. 2014: 220, Haig & Rasekh-Mahand 2019, Faghiri & Samvelian 2020). In this respect, Persian word order is disharmonic (Hawkins 2008), showing a mixture of head-initial and head-final features. A second aspect of disharmonic word order is that while direct objects are fairly consistently pre-verbal (OV), certain other kinds of constituents

may follow the verb. The first and systematic study of post-verbal phenomena in Persian is Frommer's (1981) dissertation. This pioneering study focused exclusively on the syntax of less formal Persian ('Informal Persian', IP), was based on a corpus graded according to levels of formality within IP, and employed statistical analyses rigorous for its time. We summarize Frommer's main findings in Section 2.

In this paper, we take another look at post-predicate phenomena in modern colloquial Persian, using data made available through the HamBam corpus (Haig & Rasekh-Mahand 2022) which contains annotated recordings of contemporary spoken Persian.[1] In order to facilitate comparison with the other data-sets from WOWA (Haig et al. 2024 [this volume], Section 3) we have selected a sub-set of texts from HamBam and created a data-base conforming with the WOWA format, online available as Izadi (2022). On the basis of this data, we are able to compare colloquial spoken Persian of today with Frommer's (1981), data compiled in the late 1970's, allowing us to address the question of whether Frommer's (1981) findings still hold after 40 years. In Section 2, a summary of Frommer's (1981) study and his findings is provided. In Sections 3 and 4, we present our findings based on data from HamBam, pursuing both a qualitative and quantitative approach, the latter testing the effects of predictor variables, including weight, role, flagging, animacy, and register. Section 5 compares our findings with Frommer's (1981), identifying a hitherto undetected shift in register differentiation in spoken Persian of the late 1970's, and today's language. Section 6 summarizes the main findings.

2 Frommer (1981)

Post-predicate elements in Persian have generally received only passing attention, particularly as they are generally considered to be a feature of informal spoken language, thus apparently lacking systematicity. However, Lazard (1957: 201–205) had already observed that in the colloquial spoken language, spatial Goals were normally post-posed, but the phenomenon did not attract more systematic investigation until Frommer's (1981) dissertation. In his study, Frommer focussed entirely on post-verbal elements, conducting a systematic investigation across several registers of what he refers to as "Informal Persian" (IP). The first subcorpora of his corpus consists of informal conversations in a home setting recorded by an in-group member (2595 clauses). The second part is spoken, but more formal, based on broadcasts from Radio *Payām* (1068 clauses). The third

---

[1]https://multicast.aspra.uni-bamberg.de/resources/hambam/

part is from the dialogue parts of two plays written by a famous Persian writer, Sādeq Čubak (1670 clauses), and the last part is a children's story *Kuti o Muti*, adapted for radio broadcasting (451 clauses). It is important to note that Frommer's (1981) research does not consider formal written Persian (e.g. academic, or conservative journalistic texts), but only different registers within Informal Persian (IP), as opposed to the highly formalized written language. His guiding assumption is that while verb-finality is quite strictly maintained in "Informal Persian", IP differs because it frequently permits constituents to occur post-verbally. The aim of Frommer's (1981) work is thus to elucidate the nature and function of the post-verbal elements in IP (Frommer 1981: 58–59).

Frommer (1981) noted that among the post-predicate elements, the "Goals" of verbs of motion and caused motion were among the most frequently postposed elements. Frommer used the term 'destination', which includes physical places, pro-forms (*jā* 'place'; *injā* 'here'; *unjā* 'there'; *kojā* 'where'), and abstract, or what he calls quasi-destinations (e.g. *raft xarid* 'went shopping'). This usage is close enough to the WOWA term 'Goal', which we will adopt here throughout. The following examples illustrate Goals, with the assumed canonical pre-predicate placement in (1) contrasting with the post-predicate placement in (2) (post-predicate elements are in bold through the paper):

(1) Colloquial New Persian (constructed)
     mi-xā-d   *be madrese be-r-e*
     IND-want.PRS-3SG to school   SBJV-go.PRS-3SG
     'He wants to go to school.'

(2) Colloquial New Persian (constructed)
     mi-xā-d   *be-r-e*   **be madrese**
     IND-want.PRS-3SG SBJV-go.PRS-3SG to school
     'He wants to go to school.'

Turning to Frommer's (1981) actual data, Table 1, adapted from Frommer (1981: 127), shows overall rates of clauses with post-predicate elements (V-X), and the respective proportions of Goals and non-Goals among the post-predicate elements. We have merged the data from the two play scripts because they do not differ significantly from each other.

As the table (5) shows, post-predicate elements (V-X) are overall more frequent in spoken form compared to written, with finer distinctions obtaining within the two written and two spoken sources. An important difference is that while the majority of post-predicate elements in written form are Goals, in spoken genres,

Table 1: Overall frequency of post-predicate elements and non-destination elements Frommer 1981: 127

Genre	Clause	V-Goal		V-non-Goal		V-X	
		N	%	N	%	N	%
Casual spoken	2595	168	6.5	262	10.1	430	16.6
Radio *Payām* (spoken)	1068	15	1.4	123	11.5	138	12.9
Two plays (dialogues, written)	1670	78	4.7	18	1.1	96	5.7
Children's story (written)	451	14	3.1	4	0.9	18	4.0

a much wider range of post-predicate elements is attested, and Goals only make up less than half (Frommer 1981: 128). Thus, a major distinction between the two spoken and the written registers is that the former tolerates a much wider range of post-predicate elements.

While Table 1 indicates the number of clauses with post-predicate elements, Table 2 indicates the percentage of Goals which are post-predicate in the different registers:

Table 2: Overall frequency of post-predicate Goals (Frommer 1981: 131)

Genre	Total Goals	V-Goals	Percentage
Casual spoken	203	168	82.8%
Radio *Payām* (spoken)	38	15	39.5%
Two plays (dialogues, written)	149	78	52.3%
Children's story (written)	26	14	53.8%

The score for spoken casual data is very high compared with other three forms, so it is undeniably the case that casual speech favours higher rates of post-verbal Goals (Frommer 1981 :131). However, even among the other three registers, around 50% of all Goals are post-verbal, indicating that the phenomenon of post-verbal Goals cannot be explained solely through reference to sloppy speech in informal conversational registers. Rather, it must be considered a genuine feature of vernacular Persian, evident even in (less formal) written language. Frommer also identifies a relationship between word order and flagging: Goals in post-predicate position are more likely to lack the normal prepositional flagging: "prepositionless destinations are more post-posable" (Frommer 1981: 132). He explains that since prepositionless Goals are associated with casual style, and VX

Goals are too, these two casual features are linked together making the post-predicate Goals more prepositionless (Frommer 1981: 183).

The predicates occurring with post-predicate Goals reveal a sensitivity to individual lexical verbs. post-predicate Goals occur mainly with two motion verbs, *raftan*, 'to go' and *āmadan*, 'to come' and two caused-motion verbs, *gozāštan*, 'to put' and *bordan*, 'to carry'. Table 3 shows the frequency:

Table 3: Overall frequency of post-predicate Goals with specific predicates (Frommer 1981: 133)

Predicate	Total Goals	goal-V	V-Goal	Percent of V-Goal
raftan (to go)	93	12	81	87.1%
āmadan (to come)	35	3	32	91.4%
gozāštan (to put)	17	0	17	100%
bordan (to carry)	6	2	8	75%

The tokens for other predicates in Frommer's (1981) data are too few to infer plausible conclusions. But for the verbs in Table 3, the tokens are sufficient to illustrate the strength of the post-predicate tendency, which is close to categorical. Frommer (1981: 172) summarizes his findings for the casual genre in the form of the following hierarchy for post-posability:

(3) The hierarchy of post-posability (Frommer 1981: 172):
     Goal (without preposition) > Goal (with preposition) > PP (non-Goal, including IO) > DO (with *rā*) and ADV (without preposition) > SU > DO (without *rā*)

Table 4 shows the frequencies of post-predicate arguments in the casual spoken register in Frommer's data, which underly the hierarchy of post-posability.

Frommer (1981: 135) also analyzed the effects of information structure on post-posing. He distinguished between focus (the constituent that conveys new information or asks for information as a wh-element and normally is the intonation center of the clause) and non-focused, old, background information as two main parts of information structure of the clause. He analyzed only the casual spoken data for this feature. Table 5 shows the statistics of non-focused post-predicate elements.

As the table shows, post-predicate constituents are generally non-focused (given) information - but this does not hold for Goals. Thus, while post-predicate

Table 4: Post-predicate elements hierarchy (Frommer 1981: 172, casual spoken genre only)

Constituent type	Total	VX	% VX
Goals with prepositions	134	117	87.3%
Goals without prepositions	69	51	73.9%
Prepositional arguments (not Goals)	526	95	18.1%
Objects with *rā*	224	21	9.4%
Adverbs without prepositions	1270	96	7.6%
Subjects	1083	52	4.8%
Objects without *rā*	422	6	1.4%
Total	3728	438	12%

Table 5: The frequency of non-focused elements in post-predicate position (Frommer 1981: 137)

Constituent type	% non-focused
Goals without prepositions	8.5%
Goals with prepositions	15.7%
Prepositional arguments (not Goals)	77%
Objects with *rā*	83.3%
Adverbs without prepositions	89%
Subjects	90.4%
Objects without *rā*	100%

position strongly disfavors new information, for Goals, this constraint is neutralized, with the vast majority of post-predicate Goals being in focus. It can be provisionally concluded that focus versus non-focus is not relevant for the placement of Goals, though it is clearly relevant for other constituents. Frommer (1981) explored the effect of other factors, e.g., clause type (main or subordinate), verb type (simple or complex) and heaviness, but he found no significant effects, at least in spoken language genres.

Frommer (1981: 179–181) summarizes his main findings as follows:

(a) Post-predicate placement is markedly prevalent in informal spoken Persian and less frequent in formal written Persian.

(b) Goals are the most frequent elements in post-predicate position, and more than 80 percent of them are post-posed in casual speech.

(c) Goals are mainly new information in post-predicate position, contrary to other post-posed elements.

(d) Grammatical weight has no significant effect on post-posing elements.

Following Haiman (1980: 532), Frommer (1981: 182) postulates that putting Goals in post-predicate position is related to iconicity of sequence, asserting that order of elements in language mirrors order of appearance in experience. Hence, Goals are the endpoint of a motion and appearing in final position reflects their nature (see also Haig 2022, for related claims). Finally, Frommer (1981: 183) asks if post-predicate phenomenon represents an ongoing change: Is Persian fully grammaticalizing this position? Could it be a sign of changing from SOV to SVO? Or is the VX variability a stable situation? As Frommer (1981) recognized, his data could not resolve these questions, but forty years later we are in a better position to address them.

3 Post-predicate elements in the HamBam corpus

The data of this section come from the HamBam corpus (Haig & Rasekh-Mahand 2022), a collection of annotated recordings of contemporary spoken Persian. All figures cited here stem from a data set extracted from HamBam, and analysed in the WOWA framework (Izadi 2022). The texts gathered in this corpus are predominantly monological in nature, and represent colloquial, unscripted spoken Persian. They have been broadly categorized into informal (recordings made in private homes, between kin and friends, concerned with oral history and various anecdotes), and more formal speech (e.g. radio interviews and podcasts), designed for public broadcasting. This broad two-way distinction does not readily map onto Frommer's (1981) four-way distinction; we discuss the issue of register in Section 5. The speakers are of both genders, various ages, different educational levels and occupations. Table 6 shows the total number of analyzed tokens and the rate of post-predicate elements.

It is important to note that in keeping with the WOWA data-base structure (see Haig et al. 2024 [this volume], Section 3), we have only considered non-subject constituents, hence the number of non-classified tokens is high, since there are many sentences which contain just a subject (see below for a discussion on subjects' status). In addition, in some clauses more than one token is analyzed.

Table 6: Frequency of post-predicate elements in HamBam corpus (figures based on Izadi 2022)

Total number of clause units identified	3220	100%
Number of analyzed tokens	1625	50.5%
Number of clauses lacking a classifiable token	1595	49.5%
Rate of post-predicate elements (all roles) among analyzed tokens	413	25.4%

This means that "number of tokens" means the number of analyzed constituents, and should not be confused with number of clauses (which is the unit used in several other studies). As Table 6 shows, one out of four tokens analyzed occurred after the verb.

WOWA employs a finer-grained, and slightly different classification of constituent types than that used in Frommer (1981), and for the comparison we adapt the WOWA system. Table 7 shows the overall frequency of post-predicate elements by role, including nominal and pronominal tokens.

Table 7: Post-predicate elements of different roles based on HamBam corpus

Constituent type	Total	VX	% VX
Caused Goal	60	55	91.7%
Goal	206	167	81.1%
Direct object (DEF+INDEF)	437	18	4.1%
Locative	146	29	19.9%
Ablative (source)	50	5	10%
Other (non-classifiable)	315	98	31.1%
Comitative	46	12	26.1%
Instrument	28	4	14.3%
'become' complement	24	6	25%
Addressee	69	8	11.6%
Benefactive	11	3	27.3%
Recipient+benefactive	13	2	15.4%
Copula complement	205	5	2.4%
Stimulus	4	1	25%
Recipient	11	0	0%

In the following sections, we first discuss non-direct objects in Section 3.1, direct objects in Section 3.2, and briefly touch on subjects in Section 3.3. For some of the roles in Table 7, the number of tokens or the number of post-predicate tokens is too small to gain a reliable conclusion, so they are not considered further here.

3.1 Non-direct objects

3.1.1 Goals and caused Goals

It is clear from Table 7 that Goals of verbs of caused motion and motion behave fundamentally differently from all other roles. The frequency of post-predicate Goals in our data (collapsing caused-motion and simple Goals) is around 83%, which is more than three times higher than any other single role, ignoring the 'unclassified' category for a moment. This confirms the special role of Goals already identified for Persian by Frommer (1981), and since confirmed in other studies on post-predicate elements in Iranian and neighboring languages (Haig et al. 2024 [this volume], Jahani 2018, Stilo 2018, Korn 2022).

Furthermore, the figure of around 80% matches the figure for post-predicate Goals in Frommer's (1981) casual spoken data, provided in Table 4 above. It is also replicated in another corpus of spontaneous spoken Persian, (Haig 2017). This suggests that the approximately 80% level for post-predicate Goals is a fairly stable linguistic variable for spoken Persian, which has not varied significantly over the last 40 years; we turn to this in Section 5 below; in the meantime, we provide illustrative examples of simple and caused Goals from our data.

Goals are the arguments of motion verbs (e.g., *go, come*) and caused Goals are the arguments of caused motion verbs (e.g., *put, bring, send, carry*). The data in Table 7 suggest that Goals of caused motion are more likely to be post-predicate than simple motion Goals, but a Fisher Exact test yields a p-value of 0.0504, which is only borderline significant. Examples (4) and (5) are sentences with caused motion verbs and post-predicate Goals, while (6-7) illustrate the much rarer pattern with pre-verbal Goals:

(4) Colloquial New Persian (Izadi 2022: J, 1226)
    *rad kard-e   bud-am **tuy=e čub***
    send do.PST-PTCPL be.PST-1SG in=EZ wood
    'I had stuck it into wood.'

(5)   Colloquial New Persian (Izadi 2022: J, 1299)
      *in   rā be-gozār   **ruy=e sar=at***
      this RA IMP-put.2SG on=EZ head=2SG
      'Put this on your head.'

(6)   Colloquial New Persian (Izadi 2022: J, 1288)
      tuy=aš   āb   rixt-e   bud-im
      inside=3SG water pour-PTCPL be.PST-1PL
      'We poured water into it.'

(7)   Colloquial New Persian (Izadi 2022: F, 0836)
      in   pāy=aš   rā kuče bo-gzār-ad
      this foot=3SG RA alley SBJV-put.PRS-3SG
      '(If) he puts his foot in the alley (i.e. If he goes out.).'

Frommer (1981: 132) suggests verb-specific effects here: *rixtan* 'to pour' behaves differently, for example compared with *gozāštan* 'to put', where for the former the Goals are not post-posed, but for the latter, all of the Goals appear in post-verbal position. Table 8 shows the most frequent verbs of caused motion in our corpus. The Goals of *āvordan* 'bring' categorically appear after the verb and for two other frequent verbs, just one token appears pre-verbally.

Table 8: Overall frequency of post-predicate caused Goals with specific predicates

Caused motion predicate	Total	VX	Percent of total VX
āvordan (to bring)	11	11	100%
gozāštan (to put)	12	11	92%
bordan (to carry)	11	10	91%

Examples (8-11) illustrate Goals of simple motion, the first two post-verbal and the second two examples pre-verbal:

(8)   Colloquial New Persian (Izadi 2022: C, 0256)
      *raft-am   **doktor***
      go.PST-1SG **doctor**
      'I went to (the) doctor.'

(9)   Colloquial New Persian (Izadi 2022: P, 1849)
    parid-and   **ruy=e   miz**
    jump.PST-3PL over=EZ table
    'They jumped onto the table.'

(10)   Colloquial New Persian (Izadi 2022: C, 0237)
    doktor raft-am
    doctor go.PST-1SG
    'I went to (the) doctor.'

(11)   Colloquial New Persian (Izadi 2022: ZB, 3016)
    tu harf-hā=yaš,   sohbat-hā=yaš be injā resid
    in speech-PL=3SG talk-PL=3SG   to this reach.PST.3SG
    'In his speech, his talk reached to this point.'

The overall frequency of post-verbal Goals is closely matched by the frequencies of post-verbal Goals associated with the two most frequent motion verbs; see Table 9.

Table 9: Overall frequency of post-predicate Goals with specific predicates

Motion Predicate	Total	VX	Percent of total VX
raftan (to go)	84	69	82%
āmadan (to come)	32	26	81%

We conclude that in spoken contemporary Persian, for Goals of motion and caused motion verbs the default position is post-verbal.

3.1.2 Local roles, excluding Goals: Locative and Source

Apart from Goals, some other roles referring to location such as Locatives and Source, are also relatively frequently postposed. About 20% of Locatives appear in post-predicate position, illustrated in (12) and (13), while Source is much less frequently postposed (about 10%), see (14)

(12)   Colloquial New Persian (Izadi 2022: W, 2447)
    tavaqqof dāšt-e   ast   tu Andimešk
    stop   have.PST-PTCPL be.PRS.3SG in Andimeshk
    'He stopped in Andimeshk.'

(13) Colloquial New Persian (Izadi 2022: W, 2487)

*šahid   šod   **tu jebhe***
martyr become.PST.3SG in war
'He died as a martyr in war.'

(14) Colloquial New Persian (Izadi 2022: Q, 1912)

*dast=aš   rā greft   **az   man***
hand=3SG RA take.PST.3SG from me
'He took his hand from me.'

3.1.3 Non-local obliques: Instrument, comitative, stimulus

Among the general obliques, Comitatives are more frequent than Instrument and
Stimulus roles in post-predicate position. Out of 46 tokens of Comitatives (15,
16 and 17), 12 tokens (26%) are post-posed. Out of 28 tokens of Instruments, 4
are postposed (18), and there are only two postposed Stimulus tokens (19). The
following examples illustrate these roles.

(15) Colloquial New Persian (Izadi 2022: N, 1694)

*ke yeki be-š-im   **bā   ham-digar***
that united SBJV-become.PRS-1PL with each-other
'That we become united with each other.'

(16) Colloquial New Persian (Izadi 2022: N, 1699)

*bāz   zendegi mi-kon-am   **bā=hāšun***
again life   IND-do.PRS-1SG with=3PL
'I live with them again.'

(17) Colloquial New Persian (Izadi 2022: P, 1803)

hatta bā   doxtar-hā rābet=aš   xeyli jāleb bud
even with girl-PL   relation=3SG very good be.PST.3SG
'Even his relationship with the girls was good.'

(18) Colloquial New Persian (Izadi 2022: K, 1365)

*ba'd ešāre kard   **bā   dast***
then refer do.PST.3SG with hand
'Then he indicated with his hand.'

(19) Colloquial New Persian (Izadi 2022: P, 1819)

*hasudi na-kon-and   **be ham-digar***
envy   NEG-do.PRS-3PL to each-other
'They are not jealous of each other.'

3.1.4 Other roles

This group consists of tokens which are not classifiable in other groups. Mostly they are adverbs of time and manner, or various unclassified constituent types. Overall, preverbal position is preferred for this heterogenous group, but post-verbal position is also possible (20).

(20)   Colloquial New Persian (Izadi 2022: P, 0005)
    xābid-am   **tā sā'at=e do**
    sleep.PST-1SG till hour=EZ two
    'I slept till 2 o'clock.'

(21)   Colloquial New Persian (Izadi 2022: P, 1813)
    hatta vasat=e   *kelas har*   *nim saat jāy=aš*   *rā avaz*
    Even middle=EZ class every half hour place=3SG RA change
    mi-kard
    IPFV-do.PST.3SG
    'Even in class, he changed his seat every half hour.'

(22)   Colloquial New Persian (Izadi 2022: O, 1763)
    guš-hā=yaš bā   *māsk kār*   *mi-kard*
    ear-PL=3SG with mask work IND-do.PST.3SG
    'His ears worked (despite being) with the mask.'

3.1.5 Addressees

Addressees of speech verbs appear mostly in pre-predicate position (23), with around 11% post verbal (24):

(23)   Colloquial New Persian (Izadi 2022: O, 1747)
    ba'd be āqāh=e   *goft-am*
    then to man=DEF tell.PST-1SG
    'Then I said to the man.'

(24)   Colloquial New Persian (Izadi 2022: F, 0630)
    vali na-gu   **be rezā**
    but NEG-tell.2SG to Reza
    'But, don't tell Reza.'

3.1.6 Become-complements

Complements of *become* have been identified as candidates for post-verbal position in Iranian languages (see Korn 2022 for a discussion). Our data contain 24 tokens, six of which (25%) occur post-verbally (25), while the majority is pre-verbal (26):

(25)   Colloquial New Persian (Izadi 2022: ZC, 3104)
   dah ruz šod   ***davāzdah ruz***
   ten day became.PST.3SG twelve   day
   '(The promised) ten days become twelve days.'

(26)   Colloquial New Persian (Izadi 2022: T, 2227)
   xalāban šod
   pilot   become.PST.3SG
   'He became a pilot.'

3.1.7 Benefactive

Both Recipients and Benefactives have been claimed to pattern similarly to Goals in some languages (Haig et al. 2024 [this volume]). In our data, all Recipients are preverbal, and the majority of Benefactives likewise, though the absolute number of tokens is low (we include under 'Benefactives' tokens that are ambiguous between a Recipient and Benefactive reading, coded as "rec-ben" in WOWA). Of the 24 tokens of Benefactives, five were post-verbal (27).

(27)   Colloquial New Persian (Izadi 2022: M, 1537)
   in   yek pitzā āvar-d   ***barāy-e mādar-e   man***
   This one pizza bring-PST.3SG for-EZ   mother-EZ I
   'He brought a pizza for my mother.'

3.1.8 Summary: Non-direct objects

With regard to the non-direct objects position in our data, the first and expected observation is that Goals and Goals of caused motion verbs are distinct from all other roles, and appear in post-predicate position near categorically. However, it is also important to note that the second most likely post-posed argument after Goals are actually locations (around 20%). This suggests a general principle of constituents indicating spatial location (either static (loc) or dynamic (Goals) are

more likely to be post-predicate than any others. It may also be linked to the feature of +/- humanness; this possibility is explored in Section 4 below. These findings question the validity of lumping Addressees and Recipients with Goals into a meta-role "Target" (Asadpour 2022). The Persian data suggest that Addressees are actually less likely to be post-predicate than, for example, locations, while Recipients are categorically pre-verbal. Thus spoken Persian provides little support for the assumption of a meta-role that would encompass Goals, Recipients, and Addressees. Rather, they reinforce the special status of Goals, in opposition to all other constituent types.

3.2 Direct objects

The first point about post-predicate direct objects is that they are overall very infrequent. As Table 10 shows only 18 tokens (about 4%) of direct objects of different kinds appear post-verbally, demonstrating that spoken Persian is fairly consistently OV. The frequency of different kinds of direct objects is provided in Table 10, further distinguishing animacy, definiteness, and noun vs. pronoun.

Table 10: Post-predicate Direct Objects in HamBam corpus

Direct objects	Total	VX	Percent
Nominal, all	372	17	4.6%
Nominal, human	63	5	7.9%
Nominal, animate	5	0	0%
Nominal, inanimate	285	11	3.9%
Nominal, indefinite	204	7	3.4%
Nominal, definite	168	10	6%
Pronominal (1, 2, 3)	44	1	2.3%
DO with RA	240	12	5%
DO without RA	197	6	3%

Although the absolute number of direct objects in post-predicate position is low, the findings suggest that most of the post-predicated direct objects are human, and definite ones appear more freely in post-verbal position compared to indefinite ones. When pronominal, they appear rarely in post-predicate position, and direct objects with =rā move more freely to post-verbal position compared to those without =rā. The following are examples of direct objects in post-predicate position:

(28) Colloquial New Persian (Izadi 2022: F, 0797)
*in   ke   māšin zad   **Mehrdad rā***
this that car   hit.ᴘsт.3sɢ mehrdad   ʀᴀ
'When the car hit Mehrdad.'

(29) Colloquial New Persian (Izadi 2022: Q, 1919)
*faqat did-am   **yek daste mu***
Just   see.ᴘsт-1sɢ one bunch hair
'I just saw a bunch of hair.'

(30) Colloquial New Persian (Izadi 2022: ZB, 3034)
*tu=ye sohbat-hā=yāš bargašt   az=am   porsid   **esm-ā   rā***
in=ᴇᴢ talk-ᴘʟ=3sɢ   return.ᴘsт.3sɢ from=1sɢ ask.ᴘsт.3sɢ name-ᴘʟ ʀᴀ
'During his talk, he asked the names from me.'

(31) Colloquial New Persian (Izadi 2022: ZA, 2939)
*va motasefāne   jav   gereft   **man rā***
and unfortunately excitement take.ᴘsт.3sɢ I   ʀᴀ
'And, unfortunately I was excited.'

3.3 Subjects

Up to now we have analyzed non-subject roles (see Table 6), because these are
coded in the WOWA data set (Izadi 2022). In order to consider subjects, we turned
to the full HamBam corpus. We extracted 843 tokens of nominal and pronominal
subjects, from which 27 were post-posed (3%). Some examples of post-predicate
subjects are as follows:

(32) Colloquial New Persian (Izadi 2022: F, 0657)
*bord=eš   **āqā=he***
take.ᴘsт=3sɢ man=ᴅᴇғ
'The man took it.'

(33) Colloquial New Persian (Izadi 2022: F, 0754)
*hiči   um-ad   **doktor=e***
anyway come.ᴘʀs-3sɢ doctor=ᴅᴇғ
'Anyway, the doctor came.'

The post-predicate subjects were all definite, and can reasonably be classified
as afterthoughts: the speaker has already established the reference, which is thus

presumably active in the listener's mind, and the afterthought simply re-confirms the given status of the referent. Post-posing of subjects is therefore overall very seldom in our data (see Skopeteas 2024 [this volume], on the information status and prosody of post-verbal elements in Persian).

4 A multivariate analysis of post-verbal syntax in contemporary Persian

Having introduced and illustrated individual factors identified in Frommer (1981) and the more recent data from WOWA/HamBam (Izadi 2022; Haig & Rasekh-Mahand 2022), in this section we apply two different methodologies that control for the interactions of individual factors in order to assess their respective impact in driving post-verbal placement in spoken Persian. For these purposes, we analyse the full data set in Izadi (2022); Frommer's (1981) actual corpus data are unfortunately not available.

For the first analysis (Section 4.1), we run a series of generalized logistic regression models; in a second step (Section 4.2), we implement methods from the machine learning toolbox, namely a gradient boosting machine (GBM, Friedman 2001) and, for the purposes of illustration, a classification tree. In both approaches, the response variable is positioning (*pre-verbal* vs. *post-verbal*).

4.1 Logistic regression analysis

We run four generalized logistic regression models predicting post-verbal placement, one for each of the following roles or groupings of roles:

   (i) direct objects,

  (ii) Goals,

 (iii) locations and sources, and

  (iv) various other obliques (incl. addressees, recipients, beneficiaries, and instrumentals).

The preceding sections have already confirmed that role is the primary factor in determining post-verbal placement, but also that the relationship differs substantially between roles. It is for this reason that we deem running separate logistic regression models for each role (or role combination) prudent, as doing so enables us to identify any pertinent associations within individual roles more clearly.

The following five predictors are implemented for each role: register (*public* vs. private), form (*nominal* vs. pronominal), the presence of flagging (*none* vs. case marking for objects and prepositions for other roles), humanness (*non-human* vs. human), and phrase weight (measured in characters, roughly equivalent to phonological weight; for details on quantifying weight in the WOWA data, see Haig et al. 2024 [this volume], Section 4).

In the model summaries in Tables 11-14, the values of e^β (the log odds) for each of the predictors assess of how much each of the predictors in the model affects the likelihood of the response variable yielding one or the other outcome. Log odds above 1 indicate higher odds of a post-verbal outcome, while values below 1 do the same for pre-verbal outcomes, both under conditions that all other predictors are held at their respective reference levels (i.e. the ones in italics above, and a theoretical value of 0 for phrase weight).[2] For instance, in Table 11, which summarizes the model outcomes for direct objects, the log odds for the presence of flagging (i.e. case-marking) are e^β = 2.52, meaning a case-marked direct object has 2.52 times higher odds of being post-verbal compared to a direct object without case-marking, relative to the base odds of a post-verbal outcome overall (which can be found in the row labelled "intercept", here e^β = 0.05). However, the model deems this prediction to likely be a matter of chance with a probability of p = 0.098, and it should therefore not be taken as evidence for a causal correlation. For the purposes of this analysis, we set the threshold for significance at $p < 0.05$.

Table 11 shows the model results for direct objects, Table 12 for Goals, Table 13 for locations and sources, and Table 14 for other obliques. With the exception of register for locations/sources (e^β = 8.12 times higher odds of post-verbal positioning for private register, with $p < 0.05$) and flagging for other obliques (e^β = 0.02 times lower odds for PPs, with $p < 0.01$), none of the predictors in any of the four models pass the threshold for statistical significance. As such, what little variation in positioning there is for direct objects and Goals cannot be adequately explained by register, humanness, form, the presence of flagging, or phrase weight.

4.1.1 Discussion

These results largely confirm observations of the preceding sections, but reveal additional subtleties. For two of the roles tested here, direct objects and Goals, position relative to the verb is essentially predictable from the nature of the role

---

[2] Note that these values are on a logarithmic scale, i.e. log odds from 1 to 0 for negative outcomes map onto log odds from 1 to infinity for positive outcomes.

Table 11: Logistic regression model for direct objects

model coefficients		e^β	β	SE	z-val.	p-val.
(intercept)		0.05	-3.06	0.88	-3.49	< 0.001
register	= private	1.00	0.00	0.66	0.01	0.996
form	= pronominal	0.16	-1.83	1.15	-1.60	0.110
flagging	= case-marked	2.52	0.92	0.56	1.65	0.098
humanness	= human	1.56	0.45	0.70	0.64	0.521
weight	per character	0.94	-0.06	0.06	-1.13	0.259

deviance residuals				
min.	lower	median	upper	max.
-0.49	-0.33	-0.25	-0.21	2.87

model evaluation		
observations	434	(17 post-verbal)
null deviance	143.48	on 433 degrees of freedom
resid. deviance	137.77	on 428 degrees of freedom

itself. Factors such as phrase weight, whose importance has been stressed repeatedly in the literature on word order variation, appear to have no consistent effect on position relative to the verb for these two roles in our spoken New Persian data. Furthermore, there is no effect of public versus private registers of spoken language.

For other roles, we find slight effects of flagging, such that absence of a preposition favours post-verbal placement for other obliques, in partial confirmation of one of Frommer's (1981) observations mentioned in Section 2. However, somewhat surprisingly, this effect is absent with non-Goal spatial relations (locations and sources). With locations and sources, we find an effect of register (private register favours post-verbal placement). The effects of flagging, and register, have been noted in the literature, but this is the first time that we are able to disentangle the role-specific effects.

One way of looking at our results is to consider the syntax of spoken Persian as defined in terms of two opposing role-specific rules (direct objects are pre-verbal, Goals are post-verbal), with other roles being pulled in opposing directions, subject to a range of distinct contextual and register-related factors which are only partially captured in the current model.

Table 12: Logistic regression model for Goals

model coefficients		e^β	β	SE	z-val.	p-val.
(intercept)		7.78	2.05	0.64	3.21	< 0.01
register	= private	0.84	-0.170	0.59	-0.29	0.772
form	= pronominal	0.42	-0.87	0.92	-0.95	0.342
flagging	= preposition	0.63	-0.47	0.36	-1.31	0.192
humanness	= human	0.66	-0.42	0.80	-0.52	0.601
weight	per character	1.00	0.00	0.04	0.01	0.992
deviance residuals						

min.	lower	median	upper	max.
-2.09	0.53	0.53	0.66	1.13

model evaluation			
observations	266		(222 post-verbal)
null deviance	238.62		on 265 degrees of freedom
resid. deviance	233.10		on 260 degrees of freedom

4.2 Gradient boosting models

While in the previous section we have examined each role independently for its association between positioning and various factors, in this section we answer the more general question of which factors most strongly influence post-verbal placement overall, when all the factors identified here are included in the model. To do this we utilize an iterative classification algorithm, specifically a gradient boosting machine (GBM, Friedman 2001) – a cousin of the random forest algorithm that tends to yield comparatively better results for small and skewed data sets – and, for the purposes of illustration, a single classification tree.

There are a few differences in how we arrange the predictors for these analyses compared to the regression models in the previous section. First, there is no need to maintain separate models for each of the roles, as the model will automatically select role as a classifying factor in whatever it deems most appropriate, alongside the other factors of register, form, humanness, the presence of flagging, and phrase weight. Second, since there are an unequal number of data points for each role, we calculate case weights for each role to prevent more common roles from dominating the model results. In effect this means that data points for the

Table 13: Logistic regression model for locations and sources

model coefficients		e^β	β	SE	z-val.	p-val.
(intercept)		0.033	-3.42	1.11	-3.07	< 0.01
register	= private	8.12	2.09	1.05	1.99	< 0.05
form	= pronominal	1.56	0.44	1.03	0.43	0.665
flagging	= preposition	1.06	0.06	0.43	0.13	0.897
humanness	= human	1.31	0.27	0.89	0.30	0.764
weight	per character	0.98	-0.02	0.04	-0.47	0.640

deviance residuals

min.	lower	median	upper	max.
-0.93	-0.66	-0.64	-0.24	2.37

model evaluation

observations	196	(34 post-verbal)
null deviance	180.85	on 195 degrees of freedom
resid. deviance	172.77	on 190 degrees of freedom

less common roles (Goals, locations/sources, and other obliques) are given proportionally greater weight compared to the most common role (direct objects).[3] The respective case weights for each role are listed in Table 15 further below. Third, we refactor our measure of phrase weight from a scalar variable into a categorical one with four levels (≤ 5 characters, 6–8 characters, 9–12 characters, and ≥ 13 characters long) in order to simplify interpretation of model results.[4]

Figure 1 shows a binary classification tree,[5] a visual representation of the output of a recursive partitioning algorithm. Starting from the top, each "node" in

---

[3]For example, since there are 434 data points for direct objects but only 266 for Goals, each Goal is treated as if it were 434 / 266 = 1.63 data points instead.

[4]In a gradient boosting model, the estimation of the relative influence of the predictors in the model is based on how many major splits it produces over the many thousand iterations of tree-building; as a consequence, there is an inherent bias for predictors with many levels compared to, for instance, binary predictors, as the latter can only ever be selected once for a split in each iteration of the tree. Even with this change, since the other predictors in the model are all binary, we still need to account for a bias towards overvaluing phrase weight when assessing the results.

[5]Hyperparameters for the classification tree: maximum tree depth = 6; learning rate = 0.001; minimum number of observations in nodes = 10; cross-validation folds = 10. See Table 13 for the case weights applied to different roles.

Table 14: Generalized linear regression model for other obliques

model coefficients		e^{β}	β	SE	z-val.	p-val.
(intercept)		1.53	0.43	1.74	0.24	0.807
register	= private	4.08	1.41	1.10	1.28	0.199
form	= pronominal	0.38	-0.98	0.86	-1.14	0.256
flagging	= preposition	0.02	-3.73	1.22	-3.04	< 0.01
humanness	= human	1.19	0.18	0.91	0.19	0.848
weight	per character	1.01	0.01	0.06	0.25	0.806

deviance residuals				
min.	lower	median	upper	max.
-2.13	-0.38	-0.37	-0.19	2.52

model evaluation

observations	131	(17 post-verbal)
null deviance	101.12	on 130 degrees of freedom
resid. deviance	70.69	on 125 degrees of freedom

the tree produces a "split" in the data along the values of a predictor. Which values of which predictor are selected by the algorithm at which split in the tree is determined by how cleanly they divide the data (i.e. by reducing the rate of misclassification). For instance, the first split is between roles, differentiating Goals (which are overall 83% post-verbal) from all other roles (direct objects, locations/sources). Thus the observation from other WOWA data sets of Iranian languages (e.g. Nourzaei & Haig 2024 [this volume]), that spatial Goals are indeed a special case that need to be distinguished from all other roles, is confirmed in our analysis. The left branch then splits again along roles, and so on. Overall, we can identify three groups of leaf nodes: The two leftmost nodes are almost exclusively pre-verbal, containing, respectively, all direct objects, and locations/sources and other obliques in public speech. The two leaf nodes in the center contain data points that are around 20% post-verbal; they include locations/sources and other obliques in private speech that are either flagged with a preposition, or not flagged but non-human. Those that meet the latter set of criteria but are human instead are conversely 80% post-verbal; however, there are only 10 such cases in the corpus. The rightmost leaf node, as already noted above, contains all Goals and is 83% post-verbal. It is important to note that while

single tree models are nicely illustrative and (largely) intuitive to interpret, their predictions are not robust. Small changes to the data or the hyperparameters of the model can effect substantial differences in the structure of the resulting tree.
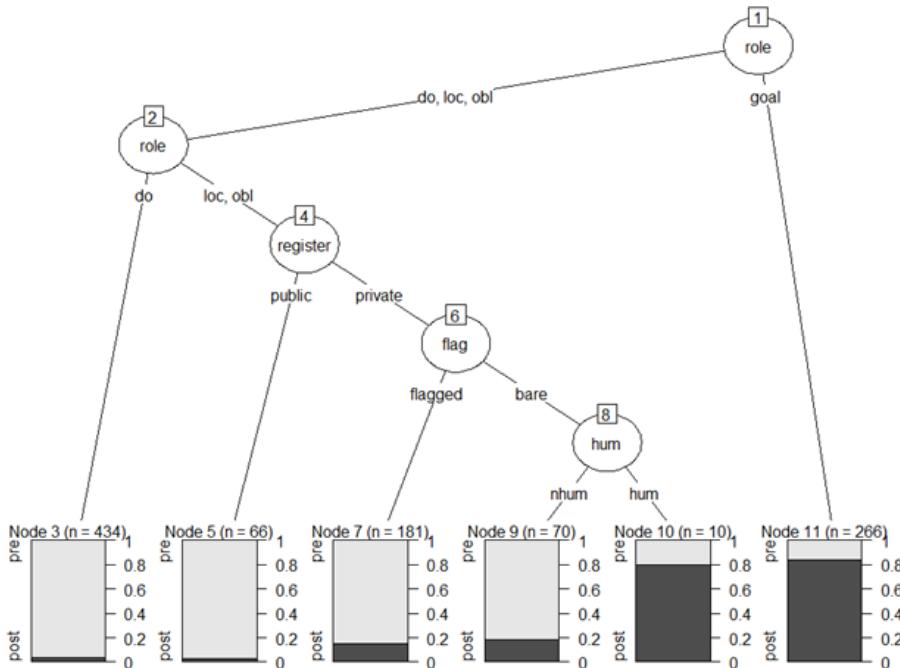
Figure 1: Binary classification tree.

These shortcomings are addressed by so-called ensemble models. Unlike classification tree algorithms, gradient boosting machines (and other methods) do not fit a single tree to the data once but rather perform a self-improving fitting process that learns as it goes, usually over thousands of iterations of trees. This greatly improves accuracy and allows each predictor to appear in a variety of contexts, thereby more thoroughly unraveling the often highly complex effects of the predictors on the response (Strobl et al. 2009: 336). The downside is that this makes the model results more difficult to interpret in their entirety, as there is no single "final" tree generated by the model.

That said, there are nevertheless many ways of summarizing their output that offer critical insight into the relationships between the model parameters. One such way is by looking at the relative importance of each predictor, which is determined by how often a particular predictor was selected for a "split" across

the many thousand iterations of trees generated by the model. The results of this analysis can be found in Figure 2, which provides an answer to the question of the relative importance of different factors in determining whether a constituent is placed pre- or post-verbally. Unsurprisingly, the semantic role of the constituent is given predominant importance, a consequence of the practically diametrically opposed profiles of Goals (chiefly post-verbal) and all other roles (largely pre-verbal, albeit to different degrees). Likewise unsurprising is the relative lack of importance of the other predictors in the model, a reflection of the results of the regression models in the previous section. In the grand scheme of things, the effect of register for placement of locations/sources that we had noted in the previous section fails to materialize as particularly influential, and the association of flagging for other oblique roles is only marginally more so. How much of the importance of phrase weight is due to biases in the structure of the predictors (see footnote above) is difficult to say, but given the lack of any sort of association in the regression models, it is unlikely to play much of a role.
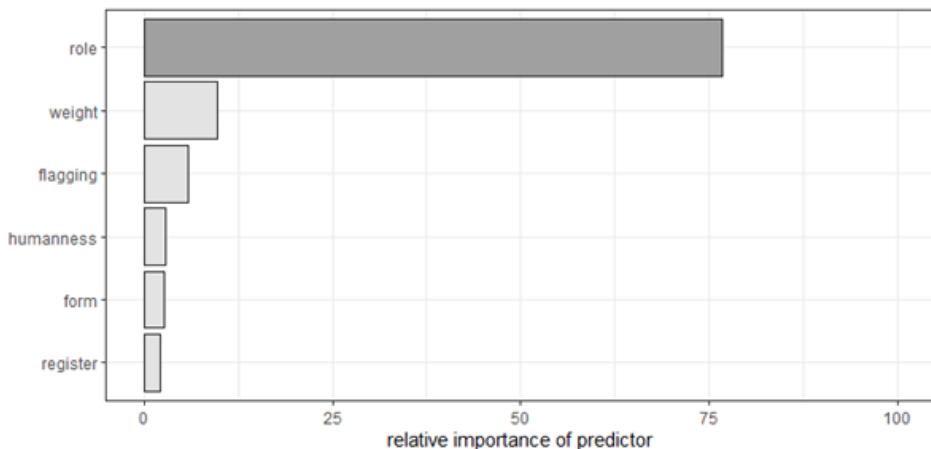
Figure 2: Relative importance of the predictors in the gradient boosting model.

   In sum, both models unsurprisingly confirm the overwhelming impact of role as the single most important predictive factor. Specifically, in contemporary spoken Persian, noun phrases bearing the semantic role of Goals of motion or caused motion are placed post-verbally with a very high probability, regardless of other factors. The other factors examined here play only a marginal role, with next most important predictor in this model being weight. The remaining predictors including humanness, form, and register turn out to have little impact overall, or their influence on word order are at best confined to specific contexts.

Table 15: Parameters for the gradient boosting model

response	position	(pre-verbal, post-verbal)
predictors	role	(direct object, Goal, location/sources, other oblique)
	register	(public, private)
	form	(nominal, pronominal)
	flagging	(none, flagged)
	humanness	(non-human, human)
	weight	(≤ 5, 6–8, 9–12, ≥ 13 characters)
error distribution function		Bernoulli (for binary response variables)
observations		1027 (290 post-verbal)
model hyperparameters		
number of trees	10000	
learning rate	0.001	
interaction depth	7	
min. obs. in nodes	25	
cross-validation folds	10	
case weights (balancing out differences in the number of observations across roles)		
direct objects	× 1.00	(434 obs.)
Goals	× 1.63	(266 obs.)
locations/sources	× 2.21	(196 obs.)
other obliques	× 3.31	(131 obs.)

5 Comparing Frommer (1981) and the HamBam data: The role of register

Having presented Frommer's (1981) data and our data from the HamBam corpus (Haig & Rasekh-Mahand 2022), we now undertake a more detailed comparison between the two, and address the question of register differentiation. First of all, we need to specify the nature of the register levels in the two data sets, and address the issue of comparability. As noted above in Section 2, Frommer's (1981) data includes a mix of spoken and written corpora, and differing grades of formality within each. For Frommer's (1981) spoken data, two levels of formality were included: casual conversational data in a domestic setting, and spoken language from the radio broadcasts of Radio Payām. We refer to these two registers as "private" and "public" respectively. The radio broadcasts included a mix of music, news "read in formal Persian", pre-recorded commercials, but also spontaneous "banter between co-hosts, and often live telephone conversations between the hosts and the listeners" (Frommer 1981: 74). Frommer included only

those sections of the recordings which he considered "to be the most relaxed and spontaneous, and bore the phonological and morphological hallmarks of colloquial style". The main difference between private and public spoken registers, as defined here, is that the former is exclusively between familiar persons in a private setting, while the latter involves a mix of familiar and unfamiliar interlocuters, produced with the knowledge that the language is publicly broadcast. Both, however, involve spoken, and largely spontaneous language.

We apply a similar distinction between private and public to the HamBam corpus. Recordings are characterized as "private" when they stem from interactions in private settings, between familiar interlocutors (kin or close friends). Recordings characterized as "public" are from publicly available sources such as radio and podcasts, often involving more academic and abstract subject matter, while still remaining quite spontaneous. Unfortunately, the amount of "public" register in HamBam texts is not very high, so this aspect of the comparison is tentative. Table 16 provides an overview of the HamBam data that feed into the comparison.

Table 16: Frequency of post-predicate elements in private and public genres of HamBam corpus

	total	public	% Po	private	% Po
Total tokens	3219	574	17.8%	2645	82.2%
Number of analyzed tokens	1624	273	16.8%	1351	83.2%
Number of non-classified tokens	1595	301	18.9%	1294	81.1%
Rate of post-predicate elements (all roles)	413	51	18.7%	362	26.8%

Note that the basic unit used for quantitative analysis in Frommer (1981) was the clause, whereas in HamBam, the basic unit is a referential, non-subject constituent. This makes a global comparison of rates of post-verbality difficult: Frommer calculates the proportion of clauses containing any post-verbal material among the totality of clauses in the corpus; our measure would be the proportion of all relevant constituents in the corpus that occur post-verbally. In fact, our metric is likely to make the overall value higher than Frommer's (1981) (because we do not count, for example, clauses lacking a relevant non-subject constituent).

A role-specific comparison is more reliable, at least for those roles which are defined in a comparable way in both studies. Table 17 summarizes the findings for

direct objects and Goals, in both corpora, distinguishing the two spoken registers public vs. private.[6] Post-verbal frequencies are grey-shaded.

Table 17: Comparing post-verbal frequencies for selected roles across two genres and two time periods

| | Frommer (1981) | | | | HamBam (2022) | | | |
| | public | | private | | public | | private | |
	N	%Po	N	%Po	N	%Po	N	%Po
Direct objects (all)	337	0.9	646	4.2	80	3.8	356	4.9
Direct objects (+*rā*)	178	1.1	224	9.4	47	4	193	5.1
Direct objects (bare)	159	0.6	422	1.4	33	3	163	2.5
Goals (all)	38	39.5	203	82.8	27	85	239	83
Goals (PP)	24	16.7	51	73.9	9	88.9	94	78.7
Goals (bare)	14	78.6	134	87.3	18	83.3	143	86.7
Other PP (not Goals)	622	13	519	17.9	98	14.3	369	20

Considering first the private register, it is evident that little has changed between the late 1970's and 2020's: The frequency of post-predicate elements in the selected roles has remained more or less the same. There is nevertheless one important difference between the late 1970's and the 2020's: In Frommer's (1981) data, private and public spoken language differ, for all categories, and in the same direction (frequencies of post-verbal placement increase between public and private), most notably for Goals. In the recent HamBam data, on the other hand, the differences between public and private are negligible, and even go in the unexpected direction for some roles (for example in public speech, Goals are overall slightly more frequently postposed than in private speech). In other words, in today's Persian there is almost no difference between private and public speech regarding the analyzed parameters. Figure 3 visualizes the difference between the two time periods with regard to register.

One way of interpreting these findings is in terms of "levelling up" over the the last 40 years. Apparently, 40 years ago the speech of the private domain was

---

[6]Data sources in Frommer (1981): Direct objects all: p.143, Table 11; with and without =rā: p. 143, Table 12; Goals (all): p. 130–131, Table 5; other PP (not Goals): p.160, Table 21. For "other PP (not Goals)" in HamBam we included all tokens flagged with <prep> in Izadi (2022), excluding Goals and direct objects. This may not be fully identical with Frommer's (1981) category of other PP's, so at this point, the comparison between the two data sets should be treated with caution.
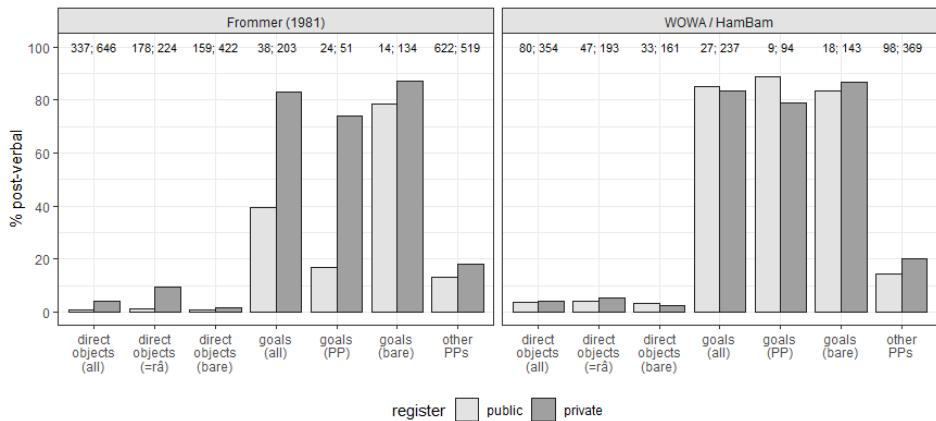
Figure 3: Register differences per role in spoken Persian 1981 and the 2020's respectively

significantly different from the public domain with regard to post-verbal syntax. But today, the difference has largely disappeared, and public speech is now essentially identical to that of the old private domain. While levelling up is generally discussed in the context of dialects (regional variants, or variation based on socio-econominc status, Dillard 1972: 200, Trudgill 1986: 98–99), the relevant concept here would be "levelling up across registers", where "register" refers to "conventionalized and recurrent" intra-speaker variation in the way speakers adapt their utterances according the context, and in particular on the perceived degree of formality (Pescuma et al. 2023: 2). With respect to Persian, our data suggest that at the time of Frommer's (1981) research in the late 1970's, speakers of Persian adapted their speech along the parameter of post-verbal placement of Goals, distinguishing between public and private domains. In the 2020's, however, it appears that the norms that were previously operative for the private domain have since spread to encompass spoken language in the public domain, i.e. that speakers no longer feel the necessity to adapt their speech in this regard (though other features of speech, such as lexical choice, phonology etc. continue to demarcate public and private speech situations).

These findings, though still tentative, open up a range of novel perspectives for understanding the dynamics of language change, particularly syntactic change. Thus while our data suggest that over the last 40 years, nothing has changed in the extreme values defined by the least formal register, by investigating different registers we are able to demonstrate that the distribution of this speech variant across different contexts has changed, namely in the form of levelling from below.

This aligns with Labov's (2001: 437) observations that change in features below the level of consciousness (which we believe holds for the syntactic phenomena under investigation here) initially "develop in spontaneous speech at the most informal level."

Returning to the question at the outset of this section then, we can tentatively conclude that private, informal speech has not changed noticeably in the last 40 years. For this register, our findings confirm Frommer's (1981) findings, suggesting a rather stable linguistic variable. Where we have identified a difference is that 40 years ago, there existed a more formal kind of "public" spoken language, distinguished from private speech by lower levels of post-verbal Goals. In our more recent data, this register appears to have merged with that of private speech.

6 Summary

Building on the pioneering work of Frommer (1981), this chapter is the most comprehensive and accountable analysis of post-predicate elements in spoken Persian currently available. We base our findings on a purpose-built and fully-accessible digital corpus of spoken colloquial Persian, the Hamedan-Bamberg Corpus of Contemporary Spoken Persian (HamBam, Haig & Rasekh-Mahand 2022), which we have adapted to the WOWA coding conventions (Izadi 2022). While the corpus size is modest in comparison to the written language corpora that underpin most contemporary corpus-based research on Persian (Faghiri et al. 2018, Faghiri & Samvelian 2020, among others), our data identify systematic differences between spoken and written Persian syntax; we conclude that generalizations regarding Persian syntax *per se* need to be tested for both modes of language production.

In the context of the present volume, we note that spoken Persian exhibits traits that are shared with the spoken Iranian languages of Western Asia (e.g. Balochi, see Nourzaei & Haig 2024 [this volume]; Gorani and Kurdish, see Mohammadirad 2024 [this volume]), most notably the strong tendency to place Goals post-verbally. In this sense, the spoken Persian investigated here is more typical for Western Iranian languages than the standard written variety of Persian, which is quite strictly verb-final (for early classical Persian, see Parizadeh & Rasekh-Mahand 2024 [this volume]). We also consider whether spoken Persian has undergone any changes over the last 40–50 years, through a comparison of our findings with those of Frommer (1981). As mentioned, there are certain difficulties with comparing the metrics used in both studies; furthermore, Frommer's (1981) original data are not available to us for verification. But for those

measures which can be reliably compared, we find no difference in frequencies of post-verbal Goals or direct objects, in the least formal sections of the samples at least, which leads us to conclude that the >80% levels of post-verbal Goals is a fairly stable variable in spoken Persian (it is also the value identified in another corpus of spoken Persian, Adibifar 2019). We do, however, find a difference in the way that post-verbal placement of Goals is mediated according to register. In the older sample, the public register exhibits lower levels of post-verbal Goals than the private register. This finding is consistent with the general consensus that standard written Persian, the pole of maximal formality, is a verb-final language, i.e. with negligible rates of post-verbal elements. From this perspective, the more formal registers of spoken Persian would be expected to be nearer to the extreme level of formality found in formal written Persian, in keeping with Frommer's (1981) conclusion regarding formality effects on word order. In our contemporary spoken data, however, we found no significant effects of register. This suggests that today's spoken language has extended what was the informal, private register, to more public settings. We assume, however, that the written language remains overwhelmingly verb final, though we have not investigated this systematically here. Spoken language may thus be prone to relatively rapid "change", but it is not the structures themselves that change; rather, it is the social indexing attached to the already available structures.

In Section 4 we conducted two different kinds of multi-variate analysis on the HamBam data in order to provide a statistically more rigorous answer to the question of what drives post-verbal placement in spoken Persian. The results confirm the effect of role, most specifically Goals versus the rest, which swamps most other factors. The logistic regression model identified effects of register and flagging, but only for specific roles; in the boosted decision trees, these effects turn out as marginal. Thus overall we find little evidence for a consistent effect of weight, humanness, or form. However, we note that our analysis is relatively coarse-grained, and a more detailed examination of individual contexts may well uncover additional predictors that were missed in our model.

Abbreviations

1	first person	COP	copula
2	second person	DEF	definite
3	third person	DEIC	deictic
ABL	ablative	DEM	demonstrative
ADDR	addressee	DRCT	directional
BEN	beneficiary	EXIST	existential

EZ	ezafe	PRS	present
FUT	future	PST	past
IND	indicative	PTCPL	participle
INDEF	indefinite	RA	object-marking clitic *=rā*
INSTR	instrumental	REL	relative
LOC	locative	SBJV	subjunctive
N	total number of tokens	SG	singular
NEG	negator	V	verb
PL	plural	VX	verb > any constituent
PO	post-posed	WOWA	= Haig et al. (2022)
PP	prepositional phrase		

References

Adibifar, Shirin. 2019. Multi-CAST Persian translated texts. In Geoffrey Haig & Stefan Schnell (eds.), *Multi-CAST: Multilingual corpus of annotated spoken texts.* Bamberg: Bamberg University Press. multicast.aspra.uni-bamberg.de/ (10 July, 2023).

Asadpour, Hiwa. 2022. Word order in Mukri Kurdish - The case of incorporated targets. In Hiwa Asadpour & Thomas Jügel (eds.), *Word order variation: Semitic, Turkic and Indo-European languages in contact*, 63–87. Berlin: De Gruyter Mouton.

Dillard, J. L. 1972. *Black English: Its history and usage in United States.* New York: Random House.

Dryer, Matthew S. 1992. The Greenbergian word order correlations. *Language* 68. 81–138.

Faghiri, Pegah & Pollet Samvelian. 2020. Word order preferences and the effect of phrasal length in SOV languages: Evidence from sentence production in Persian. *Glossa: A journal of general linguistics* 5(1). DOI: 10.5334/GJGL.1078.

Faghiri, Pegah, Pollet Samvelian & Barbara Hemforth. 2014. Accessibility and word order: The case of ditransitive constructions in Persian. *Proceedings of the 21st International Conference on Head-Driven Phrase Structure Grammar.* 217–237. DOI: 10.21248/hpsg.2014.12.

Faghiri, Pegah, Pollet Samvelian & Barbara Hemforth. 2018. Is there a canonical order in Persian ditransitive constructions? Corpus based and experimental studies. In Agnes Korn & Andrej Malchukov (eds.), *Ditransitive constructions in a cross-linguistic perspective*, 165–185. Wiesbaden: Reichert.

Friedman, Jerome H. 2001. Greedy function approximation: A gradient boosting machine. *Annals of Statistics* 29(5). 1189–1232. DOI: 10.1214/aos/1013203451.

Frommer, Paul. 1981. *Post-verbal phenomena in colloquial Persian syntax.* Los Angeles: University of Southern California. (Doctoral dissertation).

Haig, Geoffrey. 2017. *Post-predicate arguments in Iranian languages. Keynote lecture, seventh international conference on Iranian languages (ICIL7), Lomonosov Moscow State University, 28-30th August 2017.*

Haig, Geoffrey. 2022. Post-predicate constituents in Kurdish. In Yaron Matras, Geoffrey Haig & Ergin Öpengin (eds.), *Structural and typological variation in the dialects of Kurdish*, 335–377. Cham: Springer International Publishing. DOI: 10.1007/978-3-030-78837-7_8.

Haig, Geoffrey & Mohammad Rasekh-Mahand. 2019. *Post-predicate elements in Iranian and neigbouring languages: Inheritance, contact, and information structure. Position paper, for the project on post-predicate constituents in Iranian and neighbouring languages.* (10 July, 2023).

Haig, Geoffrey & Mohammad Rasekh-Mahand. 2022. *HamBam: The Hamedan-Bamberg corpus of contemporary spoken Persian.* multicast.aspra.uni-bamberg.de/resources/hambam/ (10 July, 2023).

Haig, Geoffrey, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils Schiborr. 2024. Post-predicate elements in the Western Asian Transition Zone: Data, theory, and methods. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 3–54. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266331.

Haig, Geoffrey, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr. 2022. *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

Haiman, John. 1980. The iconicity of grammar: Isomorphism and motivation. *Language* 56(3). 515–540. DOI: 10.2307/414448.

Hawkins, John A. 2008. An asymmetry between VO and OV languages: The ordering of obliques. In Greville Corbett & Michael Noonan (eds.), *Case and grammatical relations: Studies in honor of Bernard Comrie*, 167–190. Amsterdam: Benjamins. DOI: doi.org/10.1075/tsl.81.08ana.

Izadi, Elham. 2022. Persian (New). In Geoffrey Haig, Donald Stilo, Nils N. Schiborr & Mahîr Dogan (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

Jahani, Carina. 2018. Post-verbal arguments in Balochi. Paper presented at *the International Conference on Anatolia-Caucasus-Iran: Ethnic and linguistic contacts (ACIC)*, Yerevan University, 10–12 May.

Korn, Agnes. 2022. Targets and other post-verbal arguments in Southern Balochi: A multidimensional cline. In Hiwa Asadpour & Thomas Jügel (eds.), *Word order variation – Semitic, Turkic and Indo-European languages in contact*, 89–128. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110790368-005.

Labov, William. 2001. *Principles of linguistic change, vol. 2: Social factors*. Malden, MA: Blackwell Publishers.

Lazard, Gilbert. 1957. *Grammaire du Persan contemporain*. Paris: Klincksieck.

Mohammadirad, Masoud. 2024. Zagros region: The Kurdish-Gorani continuum. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 245–279. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266347.

Nourzaei, Maryam & Geoffrey Haig. 2024. Balochi: A cross-dialect investigation of post-verbal elements. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 121–154. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266337.

Parizadeh, Mehdi & Mohammad Rasekh-Mahand. 2024. Post-predicate elements in Early New Persian (10–13th Century CE). In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 231–244. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266345.

Pescuma, Valentina N., Dina Serova, Julia Lukassek, Antje Sauermann, Roland Schäfer, Aria Adli, Felix Bildhauer, Markus Egg, Kristina Hülk, Aine Ito, Stefanie Jannedy, Valia Kordoni, Milena Kuehnast, Silvia Kutscher, Robert Lange, Nico Lehmann, Mingya Liu, Beate Lütke, Katja Maquate, Christine Mooshammer, Vahid Mortezapour, Stefan Müller, Muriel Norde, Elizabeth Pankratz, Angela G. Patarroyo, Ana-M. Pleșca, Camilo R. Ronderos, Stephanie Rotter, Uli Sauerland, Gohar Schnelle, Britta Schulte, Gediminas Schüppenhauer, Bianka M. Sell, Stephanie Solt, Megumi Terada, Dimitra Tsiapou, Elisabeth Verhoeven, Melanie Weirich, Heike Wiese, Kathy Zaruba, Lars E. Zeige, Anke Lüdeling & Pia Knoeferle. 2023. Situating language register across the ages, languages, modalities, and cultural aspects: Evidence from complementary methods. *Frontiers in Psychology* 13. 964658. DOI: 10.3389/fpsyg.2022.964658.

Skopeteas, Stavros. 2024. Post-predicate prosody in OV languages. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 87–117. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266335.

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

Strobl, Carolin, James Malley & Gerhard Tutz. 2009. An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods* 14(4). 323–348. DOI: 10.1037/a0016973.

Trudgill, Peter. 1986. *Dialects in contact.* Oxford: Blackwell.

Chapter 8

Post-predicate elements in Early New Persian (10–13th Century CE)

Mehdi Parizadeh[a] & Mohammad Rasekh-Mahand[a]

[a]Bu-Ali Sina University, Hamedan

In this chapter, we investigate post-predicate elements in written Early New Persian texts from four different sources, covering the 10th to 13th centuries. The analysis shows that post-predicate elements are overall far less frequent than in contemporary spoken Persian corpora, and that the effect of semantic role is negligible. In the written texts investigated here, post-verbal elements do not form a semantically homogenous group, and goals are not more prone to postposing than other roles. However, we do find an overall effect of weight, with heavier constituents more likely to be postposed, as well as an effect of register.

1 Introduction

The history of Persian language is normally divided into three main eras: Old Persian (OP) (6th to 4th century B.C.E), Middle Persian (MP) (3rd to 7th century C.E.), and New Persian (NP) (8th to present time). The New Persian era is subdivided differently by different scholars (see Paul 2019: 572–576), but normally the first four centuries are referred to as Early New Persian, which has some peculiar morpho-syntactic features. Paul (2019: 576) gives the following subdivisions: Early New Persian (8th-12th centuries), Standard New Persian (13th - 19th centuries) and Modern (High New) Persian (19th century-present). In this study, we analyze the post-predicate elements from a selection of texts written in Early New Persian. Since there are no authentic written materials from the first two centuries, our corpus covers materials from 10th to 13th centuries. Post-predicate elements are reportedly frequent in contemporary spoken Persian

(Frommer 1981, Rasekh-Mahand et al. 2024 [this volume]). Lazard (1963) is among the first studies which dealt with constituent order in ENP. This study focuses on written materials from the first records of New Persian, in search of similarities and differences between them. A further important source for ENP are Early Judaeo-Persian texts, written mainly from the 10th–12th centuries CE. The Early Judaeo-Persian corpus contains autographs of private letters that represent an authentic colloquial style that is hardly found in ENP literary texts, their analysis could broaden our understanding of word order in ENP (Paul 2013), but lie beyond the scope of the current study. Persian boasts a rich legacy of written texts, affording scholars the opportunity to scrutinize the attributes of this language over the course of centuries. Although various aspects of the language have been examined from a historical standpoint, no prior research has addressed the post-predicate phenomena. By shedding light on word order in Early New Persian, this study brings a much-needed historical perspective into the discussion and a point of departure for quantitative studies on historical Persian (and Iranian) syntax. It also contrasts with the majority of studies in this volume, which focus on spoken language

2 The corpus

Four books from the Early New Persian were selected: *Tārikh Tabari* (10th cent.), *Qābusnameh* (11th cent.), *Tazkerat al-Oliyā* (12th cent.) and *Fihe mā Fih* (13th cent.). Regarding the sampling methodology, approximately 2000 words (five pages) were selected from each book and included in the sample. It is noteworthy that all the selected pages were the first pages of their respective books, with the exception of *Tārikh Tabari* (10th cent.) which was limited to pages 414–419. These samples yielded in total 2228 tokens of non-subject constituents for the analysis. Further details about these texts are briefly introduced in the subsequent sections.

Tārikh Tabari (10th cent.) is written by Mohammad Ibn Jarir al-Tabari, who was a Persian historian and Islamic scholar from Tabarestan, north of Iran. This book is one of the most reliable and famous references for understanding the history of religions and prophets in Iran. The book begins with the creation time and recounts stories of the prophets and kings until the time of the prophet of Islam. In the first part, it narrates the history before Islam, and in the second part, the history after Islam. This book is among the oldest texts available in Early New Persian. From the 11th century, we have used *Qābusnameh*, written by Onsor al-Ma'ali KeyKavus Ibn Iskandar-e Ziyāri, the ruler of parts of Tabarestan, north of Iran. It is arranged in 44 chapters besides an introduction. It is dedicated to

his son, Gilānšāh. In this book, he gives advice to his son as a guidance for his governance. He discusses issues such as military practices and social customs.

Tazkerat al-Oliyā (12th cent.) is a hagiographic collection of ninety-six Sufi stories by the Persian poet and mystic, Farīd al-Dīn Aṭṭār. Aṭṭār's only surviving prose work has 72 chapters, beginning with the life of the Sixth Shi'a Imam, and ending with the Sufi Martyr, Mansur Hallāj. The lives of the Sufis in this book are set in a more or less uniform format. Each biography starts with a set of embellished phrases, rhyming with one another and mentioning the subject's name, and alluding to his or her attributes before expounding on them through stories about their lives, and then by quotations from their sayings. We have excluded these sections from our analysis.

Fihe Mā Fih (13th cent.) is in prose composed by Molānā Jalāl al-Din Moham-mad Balkhi (Rumi). The subject of *Fihe Mā Fih* is mystical criticism and Rumi's interpretations of sacred texts. It includes notes written by his students, com-piled over the course of 30 years. The text is simple, and it contains thoughts on mysticism, religion and morality.

Table 1 shows the general information and the overall frequency of post-predicate non-subject elements in these selected texts. (TT stands for *Tārikh Tabari* (10th cent.), QA stands for *Qābusnāmeh* (11th cent.), TO stands for *Tazkerat al-Oliyā* (12th cent.), FF stands for *Fihe Mā Fih* (13th cent).

Table 1: Overview of the Early New Persian text corpus

	TT (10 c.)	QA (11 c.)	TO (12 c.)	FF (13 c.)
Sample length (words)	2945	2322	2447	2015
Total number of tokens	694	521	541	505
Number of non-classified tokens	166	81	109	100
% post-pred tokens	0%	9.7%	4.4%	3.7%

3 Post-predicate elements in different roles

The analysis reveals that the rate of post-predicate elements in a written corpus derived from Early New Persian exhibits certain peculiarities in comparison to the spoken-language corpora in WOWA. The primary observation is that the overall frequency of post-predicate elements is markedly low. The range of post-verbal elements varies between zero in *Tārikh Tabari* to approximately 10 percent

in *Qābusnāmeh*, while *Tazkerat al-Oliyā* and *Fihe Mā Fih* have an average of approximately 4 percent. Out of the total of 1819 tokens analyzed, only 89 were found to be in post-predicate position, which amounts to an average of 4.8 percent.

The second point to be noted is that the occurrence of post-predicate elements in written discourse is significantly influenced by the register, the content of the text, and the personal style of the writer. For instance, in *Tārikh Tabari* (10th century), which is a historical account of important events and figures, the writer has adopted a highly formal writing style, and consequently, no elements in post-predicate position are found. Conversely, in *Qābusnāmeh* (11th century), which is a father's advice to his son, the writer employs a more informal style due to the subject matter, and as a result, the text displays the highest frequency of post-posed elements. Thus, the variation in the frequency of post-predicate elements may be attributed to the register, as demonstrated by Frommer (1981) for different varieties of spoken and written informal Persian (Rasekh-Mahand et al. 2024 [this volume]). The number of analyzed tokens and frequency of post-predicate elements for the whole corpus is presented in Table 2:

Table 2: Frequency of post-predicate elements in Early New Persian

Total length	2261	100%
Number of analyzed tokens	1807	79.9%
Number of non-classified tokens	454	20.1%
Rate of post-predicate elements (all roles)	77	4.3%

The non-classified tokens are mainly those which do not have a verb and could not be analyzed. According to the Table 2, the overall frequency of post-posed elements in Early New Persian is determined to be about 4.8%. As no comparable study on written texts in Persian is available, a comparison with such texts could not be made. However, when compared with other studies on Persian post-predicate elements, the rate is found to be significantly lower than those observed in spoken New Persian (Frommer 1981, Rasekh-Mahand et al. 2024 [this volume]). Frommer (1981) reports 16.6% and 12.6% of post-predicate elements for informal and formal spoken Persian, respectively. Rasekh-Mahand et al. (2024 [this volume]) report 18.7% for public register and 26.8% for private register. Contemporary research on word order in formal written Persian does not actually consider post-predicate elements, focusing solely on the ordering of elements (e.g. direct and indirect objects) before the verb (Faghiri et al. 2014, Faghiri & Samvelian

2020). Thus, the general observation is that post-predicate phenomena are less frequent in writing.

In the remaining part of this section, we examine the different roles and frequency of elements in post-predicate position, and provide examples from various roles. It is worth noting that most of the roles contain a limited number of tokens, and post-verbal placement is characterized by exceptionally low frequencies. In Early New Persian, the benefactive role exhibits the highest proportion of post-predicate elements, with 10 out of 57 occurrences appearing in post-predicate position. The following are some examples:

(1) Benefactive:
Early New Persian (Parizadeh 2022: C, 259)
*tazkere-i   sāxt-am   **oliyā   rā***
biography-INDF build.PST-1SG clergies RA
'I created a biography for the clergies.'

In example (1), the benefactive is marked with *rā*. According to Rasekh-Mahand & Parizadeh (2024), it is not uncommon for the benefactive role to be marked by *rā* in addition to other roles in Early New Persian texts. Out of the ten tokens of benefactives in post-predicate position, four are marked by *rā*.

Tokens categorized as "Other" represent the most frequent type of role found in post-predicate position, with a total of 50 out of 508 occurrences. The following examples demonstrate their appearance in post-verbal position:

(2) Other:
Early New Persian (Parizadeh 2022: D, 221)
*va vey rā vasiyat na-kard-i   **be tafsil***
and 3SG RA will   NEG-do.PST-2SG to details
'And you did not bequeath him in detail.'

(3) Other:
Early New Persian (Parizadeh 2022: B, 436)
*darviš-i   miy-ām-ad   **pāy berahne***
poor-INDF IND-come.PST-3SG feet bare
'A poor man was coming on foot.'

(4) Other:
Early New Persian (Parizadeh 2022: C, 52)
yek ketāb-e digar mi-bāyest   jodāgāne
one book-EZ other IND-must.3SG separately
'There must be another book separately.'

Out of the 57 tokens classified as ablatives, only three appear in post-verbal position. Similarly, among the 42 tokens of stimulus, only two tokens are found in post-verbal position. Out of 28 instances of instruments, only one has appeared in post-predicate position:

(5) Ablative:
Early New Persian (Parizadeh 2022: B, 346)
*valikan qaraz dar ruze   mehr-i   ast   **az   xodāvand-e molk***
but   aim   in   fasting love-INDF be.PRS.3SG from God-EZ   world
bar molk-e   xiš
to   world-EZ self
'The purpose of fasting is the love that God has for his created world.'

(6) Stimulus:
Early New Persian (Parizadeh 2022: C, 3)
*va jamāʔati az   dust-ān   rā raqbati   tamām mi-did-am   **be***
and people   from friends-PL RA interested much   IND-see.PST-1SG to
soxan-e in   qom
word-EZ this group
'And I saw some friends were very interested in the words of this group.'

(7) Stimulus:
Early New Persian (Parizadeh 2022: C, 4)
*va ma-rā niz meyl-i   azim bud   **be motāleʔe-ye ahvāl va***
and 1SG-RA too desire-INDF great be.PST.3SG to study-EZ   vita   and
soxan-e   išān
utterance-EZ 3PL
'I was very interested in studying their lives and sayings.'

(8) Instrument:
Early New Persian (Parizadeh 2022: D, 110)
*ke mi-bin-am   **be   češm-e sar***
that IND-see.PRS-1SG with eye-EZ   head
'That I see through the eyes in my head.'

The number of locatives in out corpus is 93, from which 2 tokens appear post-verbally:

(9)   Locative:
      Early New Persian (Parizadeh 2022: D, 324)
      surat-hā-ye xub namāy-ad **dar šekam-e** *ān surat-hā-ye bad*
      form-PL-EZ good show.PRS-3SG in abdomen-EZ that form-PL-EZ bad
      'He shows good forms inside bad forms.'

(10) Locative:
      Early New Persian (Parizadeh 2022: C, 297)
      ke in če dard bud-e *ast* **dar jān-hā-ye išān**
      that this what pain be.PST-PTCPL is in heart-PL-EZ their
      'That which was a pain in their hearts.'

In all other Iranian languages examined in the WOWA data set, the highest rate of post-verbal elements is associated with goals of motion and caused motion verbs. However, in Early New Persian texts, only one instance out of 59 was discovered in post-verbal position, and there is no indication of a tendency toward goal-last placement as observed in other languages.

(11) Goal:
      Early New Persian (Parizadeh 2022: B, 223)
      va hame-ye peyqāmbar-ān rā be rāstgui dān-ad az
      and all-EZ prophet-PL RA to truthfulness know.PRS-3SG from
      ādam tā **peyqāmbar-e mā Mohammad**
      Adam till prophet-EZ 1PL mohammad
      'And he considers all the prophets from Adam to our Prophet Mohammad to be truthful.'

Out of the 67 tokens of complements of 'become', only one is found in post-predicate position. Complements of copular verbs typically precede the verb, but among the 304 instances, three have been observed in post-verbal position.

(12) Complements of 'become':
      Early New Persian (Parizadeh 2022: B, 69)
      yā be tarkib az do bov-ad **čon jesm**
      or in combination from two become.PRS-3SG like body
      'And in combination it becomes one from two things like body.'

(13) Copula complements:
      Early New Persian (Parizadeh 2022: B, 2)
      *ke hiččiz   ni-st   **az   budani va nābudani***
      so nothing NEG-be.PRS.3SG from being   and non-being
      'There is nothing either being or non-being.'

(14) Copula complements:
      Early New Persian (Parizadeh 2022: B, 15)
      *va mesāl-e   šenāxtan   čon manquš ast **va šenāsande čon naqāš***
      and example-EZ recognition like paint   is   and recognizer like painter
      'And the example of recognition to recognizer is like paint to painter.'

In Early New Persian, direct objects typically appear before the verb. In our dataset, only four tokens out of 464 were observed in post-verbal position:

(15) Direct object:
      Early New Persian (Parizadeh 2022: C, 115)
      *tā   be-dān-i   **fazl-e   išān va eflās-e   xod***
      until SBJV-know.PRS-2SG privilege-EZ 3PL   and misery-EZ self
      'So that you understand their privilege and your own misery.'

Thus far, we have examined the placement of various constituents in relation to the verb in Early New Persian. In terms of preverbal arguments, benefactives and other roles exhibit a relatively high frequency in post-predicate position, accounting for 60 out of 77 post-verbal tokens. On the other hand, some roles such as addressee, comitative, recipient, and possessive are not attested in post-predicate position. Other roles, comprising only 16 tokens, are rarely observed in post-predicate position. Table 3 shows the total number of clauses in relation to post and preverbal positions for each verb argument.

Figure 1 illustrates the proportions of post-predicate placement for different constituents in Early New Persian.

Table 4 presents the frequency of post-predicate elements in each century, with the exception of *Tārikh Tabari* (10th century), which did not contain any post-predicate elements and is therefore not included in this table.

Table 4 reveals that the majority of post-predicate elements occur in *Qābus-nāmeh* (11th century), which is known for its informal register in comparison to other texts.

Table 3: Early New Persian

	Total number of clauses	preverbal	post-verbal	preverbal	post-verbal
Benefactive	57	47	10	82.5%	17.5%
Other	508	458	50	91.2%	9.8%
Ablative	57	54	3	94.7%	5.3%
Stimulus	42	40	2	95.2%	4.8%
Instrumental	28	27	1	96.4%	3.6%
Locative	93	91	2	97.9%	2.1%
Goal	59	58	1	98.3%	1.7%
'become' complement	67	66	1	98.5%	1.5%
Copular complement	304	301	3	99%	1%
Direct object	464	460	4	99.2%	0.8%
Addressee	46	46	0	100%	0%
Comitative	17	17	0	100%	0%
Goal (caused motion)	14	14	0	100%	0%
Recipient	18	18	0	100%	0%
Recipient + Benefactive	16	16	0	100%	0%
Possessive	17	17	0	100%	0%
Total	1807	1730	77	95.7%	4.3%
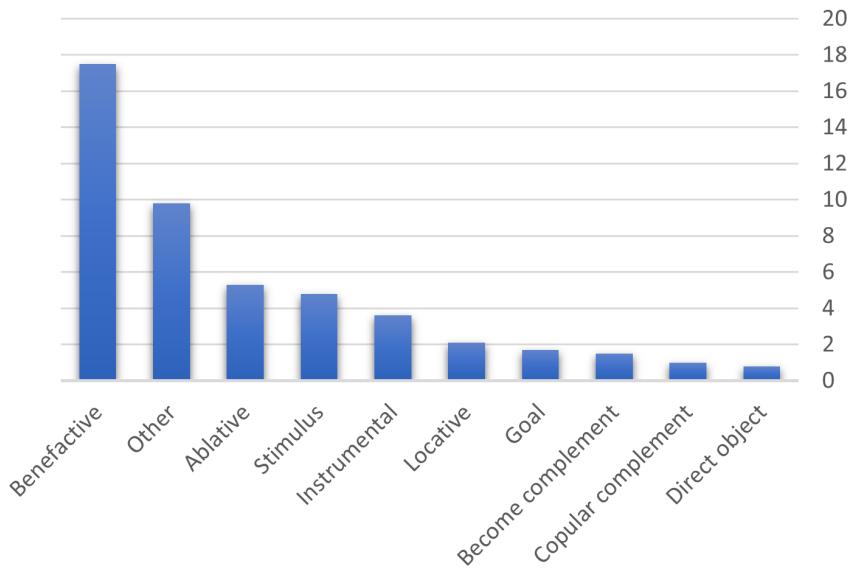
Figure 1: Post-predicate placement of different constituents in Early New Persian

Table 4: Post-predicate elements in three centuries

	QA. 11th cent.	TO. 12th cent.	FF. 13th cent.
benefactive	5	4	1
other	31	10	9
ablative	2	1	0
stimulus	0	2	0
instrument	1	0	0
locative	1	1	0
Goal	1	0	0
'become' complements	1	0	0
copula complements	3	0	0
direct object	0	1	3
Total	45	19	13

4 Heaviness

The impact of heaviness or weight on word order has been extensively discussed in the field of linguistics, with scholars such as Behaghel (1909), Quirk et al. (1972), Hawkins (1995), Wasow (1997), and Arnold et al. (2000) among others exploring this phenomenon. Some argue that in VO languages, short constituents tend to precede heavy ones (Wasow 1997, Stallings et al. 1998, Hawkins 1990, 1995), while others maintain that the reverse order holds (Yamashita & Chang 2001). Several studies have examined the effect of weight on word order in Persian (e.g., Rasekh-Mahand et al. 2016, Faghiri & Samvelian 2014, 2020, Faghiri et al. 2014, 2018), providing various analyses and occasionally conflicting results (for a detailed review, see Rasekh-Mahand et al. 2024 [this volume]). We investigated the impact of weight on post-predicate placement in our dataset, adopting the basic classification of constituent weight applied in WOWA, which recognizes four classes: w1, consisting of one phonological word; w2, consisting of two phonological words; w3, consisting of three phonological words; and w4, consisting of four or more words. Table 5 demonstrates that as the weight of the token increases, the likelihood of appearing in post-predicate position also increases.

The transition between groups w1 and w2, as well as between w2 and w4, exhibits a noticeable jump in post-predicate placement probability, as illustrated in Figure 2. While the difference between groups w3 and w4 is not as substantial, it is still discernible.

Table 5: Percentages of post-verbal placement according to weight (across all constituent types)

Weight	W1	W2	W3	W ≥ 4
Post-predicate	27	28	13	9
Total	1100	471	143	93
Percent	2.5%	5.9%	9%	9.7%
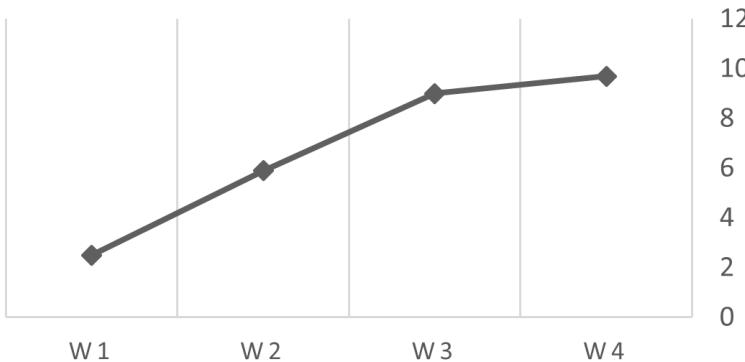
Figure 2: Weight effects on post-predicate elements in Early New Persian

Our findings indicate that weight is a factor for predicting post-predicate placement in written Early New Persian. This is consistent with much of the corpus-based literature (Hawkins 1995, Stallings et al. 1998, Arnold et al. 2000), and is particularly intriguing in the context of this volume, where weight has not yielded consistent or significant results in some of the other languages studied (see Rasekh-Mahand et al. 2024 [this volume]). The most plausible explanation for this discrepancy is the fact that the Early New Persian corpus is the only written corpus included in the WOWA collection; see Schnell & Schiborr (2022) for empirical evidence for the differences between spoken and written corpora in this respect). However, it may also be connected to the content of the texts; this remains to be investigated.

5 Summary

In this chapter, we examined post-predicate elements in written Early New Persian texts. Our analysis revealed that there are relatively few post-predicate el-

ements in these texts compared to related spoken materials, with an average of 4.3%. However, the post-predicate elements that do occur exhibit significant syntactic and semantic diversity, encompassing a range of distinct functions. The most common roles are benefactives and other items that are not easily classified using the WOWA tagging set (coded as "other"). Notably, the "goals last" effect commonly observed in spoken-language corpora studied in this volume is not present in written Early New Persian. Nevertheless, our findings suggest that weight does play a role in post-verbal phenomena in these written texts, with longer constituents being more likely to appear in post-predicate position.

The general paucity of post-verbal elements when compared to contemporary spoken Persian (Rasekh-Mahand et al. 2024 [this volume]), could be ascribed to at least three different causes: a difference in medium (spoken versus written), a difference in chronological stage of the language, or a difference in register, or some combination thereof. For obvious reasons, we have no reliable record of the spoken language in the ENP stage, so it is impossible to say whether our written texts faithfully reflect the language as it was spoken at the time. However, we do have both spoken and written texts for contemporary Persian, and initial findings suggest that there is a considerable difference between them with regard to post-verbal elements. Overall, it can be noted that less formal registers favour greater frequency of post-verbal elements, and spoken language is overall much more likely to exhibit high frequencies of post-verbal elements (Rasekh-Mahand et al. 2024 [this volume]). Our data also exhibit a slight effect of register (the least formal text has the highest overall rate of post-verbal elements), so we are inclined to consider the register and medium effects as persistent characteristics of the Persian culture of literacy over the last 1000 years. In other words, we assume that the spoken language of the ENP period was probably significantly different with respect to post-verbal elements, though the magnitude of the difference is impossible to estimate. We therefore urge caution in interpreting our results as baselines for "the" Persian language; rather we assume that our results reflect quite specific characteristics of written language, which do not necessarily faithfully reflect the spoken language of the period.

Abbreviations

1	first person	IND	indicative
2	second person	NEG	negator
3	third person	PL	plural
EZ	ezafe	PRS	present

PST	past	SG	singular
PTCPL	participle	V	verb
RA	object-marking clitic *=rā*	WOWA	= Haig et al. (2022)
SBJV	subjunctive		

References

Arnold, Jennifer E., Anthony Losongco, Thomas Wasow & Ryan Ginstrom. 2000. Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language* 76(1). 28–55. DOI: 10.1353/lan.2000.0045.

Behaghel, Otto. 1909. Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern. *Indogermanische Forschungen* 25. 110–142.

Faghiri, Pegah & Pollet Samvelian. 2014. Constituent ordering in Persian and the weight factor. In Christopher Piñón (ed.), *Empirical issues in syntax and semantics* (Empirical Issues in Syntax and Semantics 10), 215–232. Paris: CNRS.

Faghiri, Pegah & Pollet Samvelian. 2020. Word order preferences and the effect of phrasal length in SOV languages: Evidence from sentence production in Persian. *Glossa: A journal of general linguistics* 5(1). DOI: 10.5334/GJGL.1078.

Faghiri, Pegah, Pollet Samvelian & Barbara Hemforth. 2014. Accessibility and word order: The case of ditransitive constructions in Persian. *Proceedings of the 21st International Conference on Head-Driven Phrase Structure Grammar*. 217–237. DOI: 10.21248/hpsg.2014.12.

Faghiri, Pegah, Pollet Samvelian & Barbara Hemforth. 2018. Is there a canonical order in Persian ditransitive constructions? Corpus based and experimental studies. In Agnes Korn & Andrej Malchukov (eds.), *Ditransitive constructions in a cross-linguistic perspective*, 165–185. Wiesbaden: Reichert.

Frommer, Paul. 1981. *Post-verbal phenomena in colloquial Persian syntax*. Los Angeles: University of Southern California. (Doctoral dissertation).

Haig, Geoffrey, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr. 2022. *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation*. Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

Hawkins, John A. 1990. A parsing theory of word order universals. *Linguistic Inquiry* 21(2). 223–261.

Hawkins, John A. 1995. *A performance theory of order and constituency*. Cambridge: Cambridge University Press. DOI: 10.1017/cbo9780511554285.

Lazard, Gilbert. 1963. *La langue des plus anciens monuments de la prose Persane*. Paris: Klincksieck.

Parizadeh, Mehdi. 2022. Persian (Early new). In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation*. Bamberg: University of Bamberg. multicast.aspra.uni-bamberg. de/resources/wowa/.

Paul, Ludwig. 2013. *A grammar of early Judaeo-Persian*. Wiesbaden: Reichert.

Paul, Ludwig. 2019. Persian. In Geoffrey Haig & Geoffrey Khan (eds.), *The languages and linguistics of Western Asia: An areal perspective*, 569–624. Berlin: De Gruyter Mouton.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1972. *A grammar of contemporary English*. London: Longman.

Rasekh-Mahand, Mohammad, Mojtaba Alizadeh-Sahraie & Raheleh Izadifar. 2016. A corpus-based analysis of relative clause extraposition in Persian. *Ampersand* 3. 21–31. DOI: 10.1016/j.amper.2016.02.001.

Rasekh-Mahand, Mohammad, Elham Izadi, Mehdi Parizadeh, Geoffrey Haig & Nils Schiborr. 2024. Post-predicate elements in modern colloquial Persian: A multifactorial analysis. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 197–230. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266343.

Rasekh-Mahand, Mohammad & Mehdi Parizadeh. 2024. Different functions of 'rā' in New Persian; A semantic map analysis. *Journal of Historical Linguistics* 14(1). 31–57. DOI: 10.1075/jhl.21056.ras.

Schnell, Stefan & Nils N. Schiborr. 2022. Crosslinguistic corpus studies in linguistic typology. *Annual Review of Linguistics* 8(1). 171–191. DOI: 10.1146/annurev-linguistics-031120-104629.

Stallings, Lynne M., Maryellen C. MacDonald & Padraig G. O'Seaghdha. 1998. Phrasal ordering constraints in sentence production: Phrase length and verb disposition in heavy-NP shift. *Journal of Memory and Language* 39(3). 392–417. DOI: 10.1006/jmla.1998.2586.

Wasow, Thomas. 1997. Remarks on grammatical weight. *Language Variation and Change* 9(1). 81–105. DOI: 10.1017/S0954394500001800.

Yamashita, Hiroko & Franklin Chang. 2001. "Long before short" preference in the production of a head-final language. *Cognition* 81(2). B45–B55. DOI: 10.1016/S0010-0277(01)00121-4.

Chapter 9

Zagros region: The Kurdish-Gorani continuum

⓪ Masoud Mohammadirad
University of Cambridge

This chapter investigates the word order configuration of three Kurdic dialects confined within the Zagros mountains of western Iran: Gorani Gawraǰu; Central Kurdish Sanandaj; Southern Kurdish Bijar. These dialects share the commonality of having OV and Verb-Goal in their constituent ordering. However, the Gorani epicentre is characterised by extending post-predicate placement to Addressees, light-verb complements, and Locational complements in copular clauses. It will be argued that Central Kurdish dialects with post-predicate tendency for the above-mentioned constituents reflect a Gorani substrate.

1 Introduction

This chapter investigates the word order profiles of three Kurdic varieties spoken in the mountainous Zagros regions of Western Iran. The term Kurdic here refers both to the varieties that are "linguistically" considered Kurdish, i.e. Central Kurdish, Southern Kurdish, and Northern Kurdish and to the closely related but genetically more divergent languages Gorani and Zazaki that are considered Kurdish in the broader ethnic and socio-cultural sense of the term. The Kurdish languages belong to the North-western branch of Iranian languages, along with Taleshi, Tati, Mazandarani, Balochi and others.

The Kurdic varieties investigated in this study include the following: Gorani Gawraǰu, Central Kurdish of Sanandaj region, and Southern Kurdish Bijar. These dialects are representative of Kurdic languages in Western Iran—and at least in the case of the Central Kurdish dialect spoken around Sanandaj — previous scholarship has identified a greater degree of influence from Gorani than in other

varieties of Kurdish (see below). This chapter focuses on word order and considers the possible effects of Gorani on this domain of syntax. To better assess the extent of Gorani influence, we will also consider three additional varieties as control languages: Hawrami Takht, a representative of a more conservative variety of Gorani, and two varieties of Central Kurdish, Mukri and Bingird, as representatives of Kurdish varieties outside immediate Gorani influence. Figure 1 illustrates the localities in which these varieties are spoken.
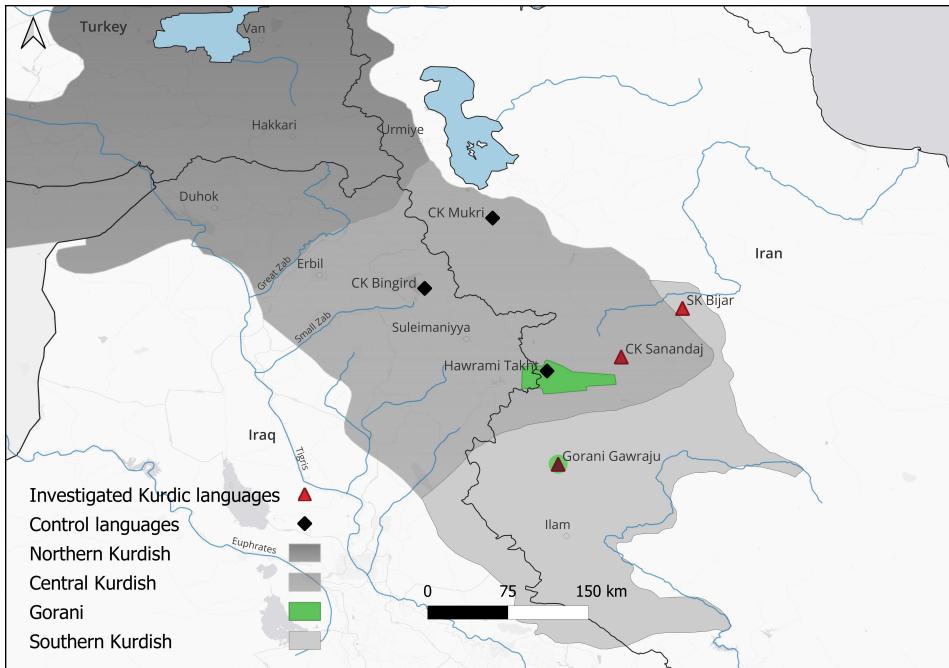
Figure 1: Investigated Kurdic dialects and control languages

The investigated dialects are situated at the intersection of Southern Kurdish and Central Kurdish-speaking areas. This intersection zone was once the Gorani heartland, which has since contracted to the mountainous Hawraman region and a few pockets across the region. There are accounts of language shift from Gorani to Kurdish in the Sanandaj region (see §4). This motivates studying the doculects in this region to see what the effects of such an assumed shift are on the word order profile of individual Central Kurdish dialects and whether it has any bearing on the classification of Central Kurdish dialects. There are now studies tackling this second aspect: Mohammadirad (In press) is a study of the organisation of inflectional paradigm in the periphery of the verb within Central Kurdish dialects. The author shows that the Southern Central Kurdish dialects adopt the

Gorani pattern of (bound) argument ordering in the verbal periphery, whereas the Northern Central Kurdish dialects opt for the opposing ordering. Mohammadirad (2024) is a case study of the Gorani substrate features in the Central Kurdish dialect of the Sanandaj area across the whole grammar. The choice of the three doculects for this study is also motivated by the fact that Kurdic vernaculars in this region have remained unstudied with respect to their word order properties (see §2).

The Gorani substrate hypothesis will be explored further in §4. It will be seen that the Gorani substrate can explain some of the differences in word order configuration in this region, which has contributed to a north/south continuum in Central Kurdish word order. The northern half concerns the Mukri and Erbil dialects (to which Southern Kurdish Bijar matches in word order). In contrast, the southern half is relevant for the Gorani zone of influence and contains the Central Kurdish Sanandaj (and possibly neighbouring Southern Kurdish varieties). Nonetheless, it should be noted that the number of data points, i.e., doculects, for this study is low, leading to the tentative nature of conclusions.

The Kurdish dialect of Bijar, also labelled Garrousi, lies at the northernmost edge of the Southern Kurdish speech zone. It is adjacent to Central Kurdish Mukri dialects to the northwest, Sanandaj-type Central Kurdish dialects to the southwest, and Azeri Turkish dialects to the east. Earlier works on the Southern Kurdish dialect of Bijar include small grammatical notes in De Morgan (1904) and Fattah (2000), and a vocabulary list in Querry (1896). It is assumed that the Southern Kurdish Bijar is the result of migration from around Ilam, situated further to the south in Zagros mountains (Fattah 2000: 18).

Gawraǰu is a small village in the north of Kermanshah, western Iran, since destroyed due to a dam building in the region. Its Gorani dialect has been described in Mahmoudveysi et al. (2012), and especially in Bailey (2018). The dialect is at the periphery of Gorani dialects and has been Kurdicised to a great degree; for instance, nominal case and gender marking are lost in Gorani Gawraǰu, contrary to the more conservative Gorani dialects.

The Kurdish dialect of the Sanandaj area is one of the southernmost dialects of Central Kurdish. The dialect is generally referred to as Ardalani when referring to the vernacular of Sanandaj itself and Laylakhi when referring to the vernaculars to the east of the Sanandaj region. Earlier work on Central Kurdish Sanandaj is restricted to De Morgan's (1904) grammatical notes. More recently, the dialect has been described in more detail within the language contact setting in Sanandaj (Khan & Mohammadirad 2024).

The data for all three dialects are from unscripted spoken narratives that were collected by the author (in the case of Kurdish Sanandaj and Kurdish Bijar) or

imported from existing studies in the case of Gorani of Gawraǰu. The recordings were transcribed, translated, and analysed using the WOWA framework (Haig et al. 2024 [this volume]); see Table 1 for an overview of sources and corpus size.

Table 1: Datasets from the WOWA corpus discussed in this chapter

Doculect	Speakers	Total tokens	Analysed tokens	Source
CK Sanandaj	8	1199	1180	Mohammadirad 2022b
SK Bjiar	3	1187	1150	Mohammadirad 2022c
G Gawraǰū	3	1325	1015	Mohammadirad 2022a based on Mahmoudveysi et al. 2012; Bailey 2018

The chapter pursues two main objectives: (i) an overview of post-predicate syntax in a sample of three doculects from the Zagros region. (ii) arguing for a Gorani substrate as an explanation for some of the differences found in this region, which has contributed to a north/south continuum in Central Kurdish word order. Applying a gradient corpus-based approach, this paper showcases the effect of language shift through Gorani substrate in contributing to the formation of a north/south continuum in the word order profile of the Central Kurdish dialects. More precisely, I suggest that the shift from Gorani to Kurdish and/or high bilingualism in Gorani in the Sanandaj region has led the Central Kurdish dialect of Sanandaj to be strikingly different in some of the post-predicate syntax from the Central Kurdish dialects in the North, e.g. Mukri. It should be noted, though, that the status of Gorani Gawraǰu is hard to reconcile with point (ii). As will be seen in §4, it has been Kurdicised to a large degree and cannot be regarded as an assumed original state of a Gorani dialect once spoken in the region.

This chapter is structured as follows. §2 gives an overview of the literature on post-predicate arguments across Kurdish. §3 classifies major word order parameters across the three investigated Kurdic dialects. §4 outlines areal issues in the configuration of certain word order profiles. §5 is the conclusion.

2 Previous scholarship on post-predicate arguments in Kurdish

Haig (2015) is the first study on the ordering of "Goals" across Kurdish dialects in the context of contact with the north-eastern branch of Neo-Aramaic dialects,

commonly called NENA. In Haig's earlier (2015) terminology, "Goal" was used as a cover term encompassing human and non-human arguments of verbs of movement, recipients of verbs of transfer, and Addressees of verbs of speech. These are all constituents that share "endpoint semantics."[1] Haig concludes that all Kurdish varieties share the commonality of post-predicate realisation of goals of verbs of movement and recipients of the verb 'give,' a head-initial trait which may be linked to an earlier imprint of Aramaic on Kurdish languages in their formative stages.

The main distinction between Kurdish dialects, according to Haig, is in finer-grained differences regarding the placement of Addressees (see also Haig 2017) and beneficiaries. Haig makes the important observation that post-predicate Goals are more prevalent across Kurdish dialects where there is a greater intensity of NENA speaking communities.[2] Thus, within Northern Kurdish, the dialects to the south-east of Anatolia/northern Iraq, which have been in close contact with vernaculars of NENA for centuries, show more post-predicate realisation than Zazaki and other Kurdish varieties in Central Anatolia, which show contact influence from Turkish and Armenian, and have the basic pattern of pre-predicate positioning for Addressees of verbs of speech. More recent research on larger samples from Kurdish and neighbouring languages has confirmed the preferred post-verbal placement of goals of verbs of motion and recipients, but some of the details regarding other roles have been revised in the light of additional data, see Haig (2022), Haig et al. 2024 [this volume], and Section 3 below.

A relevant publication is Asadpour's (2022a) dissertation on word order variation within Kurdish languages in north-western Iran in contact with non-Iranian languages Neo-Aramaic, and Turkic. Asadpour (2022b) is another work on word order variation in Kurdish. His paper focuses on "incorporated targets" in the Mukri variety of Central Kurdish. By incorporated targets, the author means Goal-like arguments that appear between the constituting elements of a complex predicate, namely the non-verbal element and the light verb, for example, *min* '1sg' in *wiłām=ī min=ī dā-w=a* [answer=EZ 1SG=3SG:A give.PST-PTCP=PERF] 'He gave an answer to me.' The author concludes that the variant ordering of incorporated targets in Mukri is accounted for by factors such as animacy and length, which favour the preverbal position.

---

[1]In more recent work, the inclusive use of the term 'Goal' has been abandoned, and 'Goal' is reserved exclusively for endpoint arguments of verbs of movement (see Haig et al. 2024 [this volume]).

[2]This is now statistically demonstrated in a recent study (Haig et al. In press): The Northern Kurdish dialect spoken in Ankara shows less tendency for post-predicative realisation of Goals than Kurdish varieties in the immediate contact zone with NENA dialects.

In short, previous scholarship has mainly taken a broader perspective on the effects of language contact on the word order profile of Kurdish. Central Kurdish dialects, in general, and especially the Gorani zone of influence corresponding to the south of Central Kurdish speech zone, have remained understudied with regard to their word order properties. Given this background, this chapter attempts to provide a more focused case study of language contact and word order within Kurdish, in particular, the southeasterly periphery.

3 Word order parameters

Barring a few features, e.g., OV order, the word order parameters across the investigated dialects are predominantly head-initial.

3.1 Adjective/noun

In the investigated dialects, attribution in the noun phrase is formed by placing the adjective following the head noun through a sort of head-marking formative called ezafe, which has different forms depending on the status of the head noun being indefinite (1.a-1.b)[3], or definite (1.c)[4]

(1)   a.   Gorani Gawraǰu (Mohammadirad 2022a: E, 0770)
            žan=e   ǰwān-ēk
            woman=EZ young-INDF
            'a young woman'

      b.   Southern Kurdish Bijar (Mohammadirad 2022c: B, 0166)
            sag=e   sīya=y   zil-aka-y
            dog=EZ black=EZ big-SPEC-INDF
            'a big black dog'

      c.   Central Kurdish Sanandaj (Mohammadirad 2022b: A, 0065)
            kanīšk=a gawra-(a)ka
            girl=EZ   old-DEF
            'the older girl'

---

[3]Alternatively, the adjective can be linked to the head noun through juxtaposition, e.g., Southern Kurdish Bijar *usāswār xās-ē* 'a good horse-riding master' (Mohammadirad 2022c: D, 0285).

[4]The ezafe particle *=a*, generally dubbed "compound ezafe," is a feature of all Central Kurdish, Southern Kurdish, and Gorani dialects but absent in the majority of Northern Kurdish dialects (cf. MacKenzie 1962: 83).

While the default pattern remains head-initial in adjective-noun constructions, these dialects show minor traces of head-final syntax in a few closely-knit compound NPs. Here, the ezafe linker appears on the adjective and has the form of *=a* similar to (1c):

(2) Gorani Gawraǰu, Central Kurdish Sanandaj, and Southern Kurdish Bijar
    juwān=a   žin
    beautiful=EZ woman
    'beautiful woman'

3.2 Possessor/possessed

In the investigated dialects, the structure of possessive constructions is possessed first, possessor second. The unmarked pattern for linking the possessor to the possessed noun is through simple juxtaposition in Gorani Gawraǰu and Central Kurdish Sanandaj, but via ezafe linker in Southern Kurdish Bijar:

(3) a. Gorani Gawraǰu (Mohammadirad 2022a: A, 0175)
    dim pišī-aka
    tail cat-DEF
    'the tail of the cat'

    b. Central Kurdish Sanandaj (Mohammadirad 2022b: G, 0714)
    kanīšk   pāwšā
    daughter king
    'the king's daughter'

    c. Southern Kurdish Bijar (Mohammadirad 2022c: D, 0318)
    zārū=ē   pādišā=y maymün-ān
    child=EZ king=EZ monkey-PL
    'the child of the king of monkeys'

3.3 Demonstrative/noun

In all the three dialects, the demonstrative attributives are discontinuous: they consist of the demonstrative attributes, sensitive to distance distinction, to the left of the head noun, and the invariable deictic form *=a* which attaches to the rightmost boundary of the NP, see Table 2 for Gorani forms. Thus, the order of demonstrative plus the head noun cannot be readily classified as fitting into either head-initial or head-final syntax.

Table 2: Demonstrative attributes in Gorani Gawraǰu (Bailey 2018: 169, simplified)

	Proximal	Distal
SG/PL	*ī …… a*	*ā …… a*

(4) a. Gorani Gawraǰu (Mohammadirad 2022a: B, 0334)
   ī   bizin=a
   DEM.PROX goat=DEIC
   'this goat'

   b. Gorani Gawraǰu (Bailey 2018: 170)
   ā   kār-ān=a
   DEM.DIST task-PL=DEIC
   'those tasks'

3.4 Numeral/noun

The investigated varieties share the commonality of ordering the numerals before head nouns. Morphologically, the head noun does not show number agreement with numerals above 'one.' Examples:

(5) a. Gorani Gawraǰu (Mohammadirad 2022a: B, 0237)
   dü   wačka
   two offspring
   'two offsprings'

   b. Central Kurdish Sanandaj (Mohammadirad 2022b: A, 0001)
   sē   kuř
   three son
   'three sons'

   c. Southern Kurdish Bijar (Mohammadirad 2022c: H, 1148)
   haš   kīsa
   eight sack
   'eight sacks'

Taken together, in the investigated Kurdic languages, the structure of NP is DEM NUM N ADJ, e.g. Central Kurdish Sanandaj. *aw sē kanīšk zarīf=a* [DEM.DIST three girl beautiful=DEIC] 'those three beautiful girls.'

3.5 Adpositions

A mixed adpositional typology is a common feature of most Kurdish varieties. This is a reflection of their geographical distribution between OV languages, e.g. Armenian, Turkic, Caucasian, and VO languages, e.g. Arabic, Aramaic (cf. Stilo 2009: 6–7). The Kurdic dialects investigated here are no exception, though the levels of postpositionality differ significantly in these doculects (see §3.20). Therefore, prepositions, postpositions, and circumpositions occur in these dialects. It is thus not straightforward to categorize these dialects easily into head-initial vs. head-final types.

(6) a. Central Kurdish Sanandaj (Mohammadirad 2022b: B, 0261)
        bo šār
        to town
        'to the town'

    b. Central Kurdish Sanandaj (Mohammadirad 2022b: A, 0010)
        la kēf   řaš=ā
        in mountain black=POST
        'at the black mountain'

    c. Central Kurdish Sanandaj (Mohammadirad 2022b: K, 1160)
        das   a-nē-t=a   žin-ak(a)=aw
        hand IND-put.PRS-3SG:A=DRCT woman-DEF=POST
        'She nudged the woman.'

3.6 Auxiliary/main verb

Given the breadth of the subject, this section only deals with auxiliary verbs in progressive constructions,[5] which precede the main verb, in contrast to the tendency in OV languages (Dryer 1992: 100).

(7) a. Central Kurdish Sanandaj (Mohammadirad 2022b: D, 0449)
        xarīk=a   jift   a-kā
        AUX=COP.3SG:S plough IND-do.PRS.3SG:A
        'He is ploughing.'

    b. Gorani Gawraǰu (Mahmoudveysi et al. 2012: 43)
        tū hē=t   kār ma-kar-ī
        2SG PTCL=2SG:A work IND-do.PRS-2SG:A
        'You are working.'

---

[5]The more archaic auxiliaries, such as various forms of the verb 'to be', are now univerbated with the lexical verb, and cannot be readily analysed as head-final auxiliaries.

3.7 Complement clause/matrix verb

The complement clause follows the matrix verb in the investigated dialects. The complementation strategy is generally asyndetic without any connective particles.

(8)   a.   Southern Kurdish Bijar (Mohammadirad 2022c: F, 0848)
        na-zānist   ča   kirdī=ya
        NEG-know.PST.3SG:A what do.PST.PTCP.3SG:A=PERF
        'He didn't know what he had done.'

     b.   Central Kurdish Sanandaj (Mohammadirad 2022b: D, 0444)
        wā   a-zān-ē   a=y-xwā
        DEIC IND-know.PRS-3SG:A IND=3SG:O-eat.PRS.3SG:A
        'He thought it (the wolf) would eat him.'

     c.   Gorani Gawraǰu (Mohammadirad 2022a: C, 0396)
        ma-wīn-ē   hüč   nīya   b-war-ē
        IND-see.PRS-3SG:A nothing NEG.COP.3SG:S SBJV-eat.PRS.3SG:A
        'He sees that there is absolutely nothing he may eat.'

3.8 Position of complementizer within the complement clause

As remarked, complement clauses are not usually introduced by complementizers. However, the corpus data shows that young, educated speakers tend to use the complementizer *ki, ka* in the initial position of some complement clauses. Kurdic varieties thus pattern with VO languages in this regard, as do all other Western Asian varieties of Iranian.

(9)   a.   Southern Kurdish Bijar (Mohammadirad 2022c: H, 1062)
        zānist   ki   bār-a=y   sangīn=a
        know.PST.3SG:A COMP load-DEF=3SG:POS heavy=COP.3SG:S
        'He (the man) knew that its (the donkey's) load was heavy.'

     b.   Central Kurdish Sanandaj (Mohammadirad 2022b: B, 0278)
        na=y-hēšt   ka   hamro-akān bi-diz-ē
        NEG=3SG:A-let.PST COMP pear-DEF.PL SBJV-steal.PRS.3SG:A
        'He didn't let him steal the pears.'

     c.   Gorani Gawraǰu (Bailey 2018: 424)
        ni-m-wāz-ē   ka   bi-zān-ē
        NEG-IND-want.PRS.3SG:A COMP SBJV-know.PRS.3SG:A
        'He doesn't want to know.'

3.9 Nominal direct object/verb

Kurdish languages are generally claimed to have basic OV word order (see McCarus 2009: 613 and Öpengin 2016: 51 for Central Kurdish; Fattah 2000: 672 for Southern Kurdish; Mahmoudveysi & Bailey 2013: 72 for Gorani Zarda). Table 3 summarizes the OV ratio of nominal direct objects for each doculect.

Table 3: Frequencies of post-verbal (Po) nominal direct objects

	G Gawraǰu			CK Sanandaj			SK Bijar		
	n	n Po	Po	*n*	n Po	Po	*n*	n Po	Po
Direct object	285	13	5%	316	3	1%	298	7	2%

It can be seen that Kurdic varieties are predominantly OV, in line with the claims made in the literature. Nonetheless, it is noteworthy that Gorani Gawraǰu allows nominal direct objects to be placed after the verb at a relatively higher rate than Central Kurdish and Southern Kurdish. This might be a reflection of earlier contact between Gorani vernaculars and Semitic languages, such as Arabic and Aramaic. Note, however, that the absolute numbers of postverbal objects in Gorani Gawraǰu are very small (13). One might even consider a contact effect from Persian here. Post-verbal direct objects in Gorani are overwhelmingly definite (11 out of a total of 13 post-verbal objects) and could be reconciled with some notion of afterthought.). In addition, they are usually arguments of the verb 'lift, grasp,' which is then followed by the directional particle.[6]

(10) a. Gorani Gawraǰu (Mohammadirad 2022a: D, 0677)
       ānī ma-nam=ya   qułang
       3SG IND-lift.PRS.3SG:A=DRCT pickaxe
       'He (Farhad) grasps (lit. lift) the pickaxe.'

    b. Gorani Gawraǰu (Mohammadirad 2022a: D, 0685)
       tamāšā=m   xās na-kard-ē   ʕask-akān
       looking=1SG:A well NEG-do.PST-PTCP picture-DEF.PL
       'I have not looked very well at the pictures.'

    c. Gorani Gawraǰu (Mohammadirad 2022a: D, 0566)
       ma-wīn-ē   ī   dawrīš-a
       IND-see.PRS-3SG:A DEM.PROX dervish-DEIC
       'She sees this dervish.'

The role of definiteness in licensing post-verbal objects in otherwise OV languages has been noted in several contributions to this volume; for instance, in Persian, see Rasekh-Mahand et al. (2024 [this volume]) and Parizadeh & Rasekh-Mahand (2024 [this volume]). Given the overall low frequency of post-verbal objects, it is hard to make a generalization about the effect of humanness vs. non-humanness in the VO configuration.

3.10 Pronominal direct object/verb

Similarly, OV is the preferred order for free pronoun direct objects; see examples in (11a-11b). Note that the number of tokens does not exceed 10 in each dataset, which precludes a premature conclusion on the word order configuration of free pronouns. The main reason for the low frequency of free pronouns as direct objects seems to be that a direct object is often indexed by a bound pronoun in these languages, which is in complementary distribution with free pronouns.

(11) a. Gorani Gawraǰu (Mohammadirad 2022a: E, 0920)
          tu   min=it   kušt
          2SG 1SG=2SG:A kill.PST
          'You killed me.'

      b. Southern Kurdish Bijar (Mohammadirad 2022c: D, 0457)
          wāna san-ī
          3PL   buy.PRS-3SG:A
          'He buys them.'

3.11 Copula complements

Complements of copular verbs show a high propensity to be placed pre-verbally. They exhibit less than 10% of post-verbal realisation across three datasets (with Gorani having the highest rate of post-predicate copula complements due to

---

[6]Likewise, a special VO syntax is associated with the verb *rāhištin* 'to lift' in Kurmanji dialects. However, the issue seems to be more complicated in Gorani. In the closely related Hawrami varieties, the verb *namāy* 'to lift' is intransitive, but its non-subject argument always appears post-verbally. On the other hand, assuming that the directional particle in (10a) flags a direct object, then this could be linked to an Aramaic influence, as noted by Don Stilo (p.c). In Khan's (2009) description of the Neo-Aramaic dialect of Sanandaj, it is mentioned that a minor strategy for marking direct objects is to use an allative strategy. However, unlike Gorani, the direct object precedes the verb. This could mean that the allative strategy for direct objects in Gorani and Kurdish is a secondary pattern borrowed from NENA but used for very specific/restricted contexts.

semi-poetic language in one of the tales). At any rate, copula complements align with direct objects in having predominantly pre-predicate placement, as noted for all other OV languages in the WOWA data set.

(12)   a. Central Kurdish Sanandaj (Mohammadirad 2022b: B, 0543)
        fra   wiryā a-w-ē
        very clever IND-be.PRS-3SG:S
        'She was very clever.'

   b. Gorani Gawraǰu (Mohammadirad 2022a: C, 0524)
        ēma řafīq bīs-yām̄
        1PL   friend be.PST-1PL:S
        'We were friends.'

   c. Southern Kurdish Bijar (Mohammadirad 2022c: D, 0278)
        žin-a=y   dugīyān d-ū
        wife-DEF=3SG:POS pregnant IND-be.PRS.3SG:S
        'His wife was pregnant.'

3.12 Goal/verb

Haig (2015, 2022) suggests that within Kurdish goals of verbs of movement (e.g., 'go,' 'come') and goals of verbs of caused motion (e.g. 'put,' 'take') have the highest propensity to occur in the post-predicate position among endpoint constituents. Table 4 exhibits the linear position of Goals and Caused goals relative to the verb.

Table 4: Frequencies of post-verbal (Po) Goals in three Kurdic doculects

	G Gawraǰu			CK Sanandaj			SK Bijar		
	n	n Po	Po	*n*	n Po	Po	*n*	n Po	Po
Goals (simple)	158	148	94%	183	162	88%	181	172	95%
Goals (caused)	105	103	98%	150	144	96%	121	117	97%
All Goals	263	251	96%	333	306	92%	302	289	96%

As can be seen from Table 4, overall Goals show more than 90% post-predicate realisation. This has been claimed to reflect the convergence of these Kurdic dialects with Semitic languages (cf. Haig et al. In press). It is notable that Goals slightly lag behind Caused goals in post-predicate realisation. Pre-verbal Goals

in these varieties occur often with some notion of "refined motion," which often expresses atelicity, see (13a), whereas caused goals express more clearly an endpoint to the action of the verb, thus post-verbal, cf. (13b).

(13) a. Southern Kurdish Bijar (Mohammadirad 2022c: H, 1128)
      waraw samt=i   bāzār du-r̆ī-yā-n
      to   direction=EZ bazaar IPFV-go.PST-IPFV-3PL:S
      'They were going in the direction of bazaar.'

    b. Southern Kurdish Bijar (Mohammadirad 2022c: D, 0476)
      siqān-a nīyā   war saw=aw
      bone-DEF put.PST.3SG:A front dog=POST
      'She put the bone in front of the dog.'

3.13 Complements of 'become'

Within the investigated Kurdic dialects, the inchoative verb 'become' has a special syntax. It implies a change of state, e.g. 'He became a king.' The verb 'become' has an identical morphology as 'be,' but unlike the latter, the complement of 'become' is usually realised post-predicatively, if it is a NP. Adjective complements of 'become,' on the other hand, are generally pre-verbal. The examples in (14) illustrate the placement of nominal complements. It is notable that the nominal complement of inchoative 'become' is flagged by the preposition *ba*, cf. (14b), which can be and is often cliticised to the verb and glossed as DRCT (directional particle).

(14) a. Gorani Gawraǰu (Mohammadirad 2022a: E, 0786)
      ma-sūz-ē   ma-w-u   xuł
      IND-burn.PRS-3SG:S IND-be.PRS-3SG:S ash
      'It (the wood) has burned up (and) turned to ashes.'

    b. Central Kurdish Sanandaj (Mohammadirad 2022b: C, 0430)
      bū   ba   šaw
      become.PST.3SG:S into night
      'It turned night.'

    c. Southern Kurdish Bijar (Mohammadirad 2022c: B, 112)
      tā   d-ū=a   nīmarū
      until IND-become.PRS.3SG:S=DRCT noon
      'Until it became noon.'

The reversed ordering for adjectival complements is exhibited in (15):

(15) a. Gorani Gawraǰu (Mohammadirad 2022a: B, 0220)
     āwis   ma-w-u   bizin-aka
     pregnant IND-be.PRS-3SG:S goat-DEF
     'The goat becomes pregnant.'

   b. Central Kurdish Sanandaj (Mohammadirad 2022b: C, 0370)
     dang=ī   nāzik=aw   bū
     voice=3SG:POS soft=COMPL be.PST.3SG:S
     'His voice became soft.'

   c. Southern Kurdish Bijar (Mohammadirad 2022c: D, 405)
     mas   d-ū
     drunk IND-be.PRS.3SG:S
     'She becomes drunk.'

Table 5 summarizes the linear positioning of complements of 'become':

Table 5: Frequencies of post-verbal (Po) complements of 'become' in three Kurdic doculects

	G Gawraǰu			CK Sanandaj			SK Bijar		
	n	n Po	Po	*n*	n Po	Po	*n*	n Po	Po
N complement	14	12	86%	22	22	100%	14	13	93%
Adj complement	20	2	10%	12	1	8%	17	3	18%
All complements	34	14	41%	34	23	68%	31	16	51%

As can be seen in Table 5, it is clear that adjectival complements of inchoative 'become' show strikingly less tendency than nominal complements to be realised post-predicatively. The reason lies perhaps in the fact that the adjective is treated as the non-verbal complement of 'become'; hence, the combination Adjective + become acts more like a complex predicate, in which the placement of adjectives is fixed preverbally. In contrast, the nominal complements are framed into a prepositional phrase and are treated like a Goal of verbs of movement.

3.14 Recipient/verb

Recipients of verbs of 'giving' are next in line in the likelihood to appear in the post-predicate position. An important observation is that in all three varieties, nominal and/or free pronouns recipients are overwhelmingly post-predicate; see Table 6.

Table 6: Frequencies of post-verbal (Po) recipients in three Kurdic doculects

	G Gawraǰu			CK Sanandaj			SK Bijar		
	n	n Po	Po	*n*	n Po	Po	*n*	n Po	Po
Non-bound recipients	13	13	100%	9	9	100%	9	9	100%
Bound recipients	31	31	100%	26	12	46%	23	0	0%
All recipients	44	44	100%	35	21	60%	32	9	28%

In the investigated dialects, pronominal recipients often occur as clitic pronouns, prosodically dependent on some other item. The dialects differ significantly in the positioning of these bound recipients; see Table 6. They also differ in the syntax of these formatives. In Southern Kurdish Bijar and Gorani Gawraǰu, bound recipients remain attached to their governing preposition but never occur on the verb. However, in Central Kurdish Sanandaj, bound pronouns have a special syntax of their own in the present tense such that they are realized on the constituent preceding their governing head (see Mohammadirad 2020; Öpengin & Mohammadirad 2022 for the grammar of bound pronouns within Kurdish).

(16) a. Southern Kurdish Bijar (Mohammadirad 2022c: A, 0069)
   mirw-aga xā **wa=y** *da-y*
   hen-DEF   egg to=3SG:R give.PRS-3SG:A
   'The hen gives him egg.'

   b. Central Kurdish Sanandaj (Mohammadirad 2022b: I, 0990)
   poł=o   *pē a-wa-m*
   money=2SG:R to IND-give.PRS-1SG:A
   'I will give you money.'

The following examples illustrate the post-verbal positioning of bound recipients:

(17) a. Gorani Gawraǰu (Mohammadirad 2022a: B, 0233)
   šīr-aka=š   *ma-t-ī=ya*   **wan=šān**
   milk-DEF=3SG:POS IND-give.PRS-3SG:A=DRCT to=3PL:R
   'She gives them her milk.'

b. Central Kurdish Sanandaj (Mohammadirad 2022b: B, 0323)
   *sē   dāna hamro=y   dā   **pē=yān***
   three CLF   pear=3SG:A give.PST to=3PL:R
   'He gave them three pears.'

According to Table 6, bound pronouns have different word order preferences than nominals, except in Gorani Gawraǰu. The crucial point here is that in Gorani Gawraǰu, the adposition itself occurs after the verb (and therefore the non-mobile clitic with it). In contrast, in Southern Kurdish Bijar, the adposition and the non-mobile clitic appear pre-verbally. Likewise, in Central Kurdish Sanandaj, the adposition remains preverbal, at least quite frequently. The different word order preferences of nominal and bound recipients are indeed worthy of further research, especially that Hawkins's (2008) typology of "obliques" only accounts for the word order constellation of nominal constituents.

3.15 Addressee/verb

Discussing the word order preferences of Addressees across Northern Kurdish, Haig (2022) makes several important observations: (i) there is a correlation between the flagging of the Addressee argument and its position relative to the verb, such that post-predicate Addressees are not flagged via postpositions and/or circumpositions; (ii) Addressees of verbs which have telic aspectual meanings, i.e. 'say/tell' are more expected to occur post-predicatively than Addressees of a verb of speech which indicates non-telic aspectual meaning, e.g. 'speak,' the reason being that the former is associated with an endpoint activity whereas the latter is not. Examples of the positionality of Addressee arguments are presented below, see Table 7 for percentages.

(18) a. Central Kurdish Sanandaj (Mohammadirad 2022b: G, 0660)
         mard wit=ī=ya   žin-aka=y
         PN   say.PST=3SG:A=DRCT wife-DEF=3SG:POS
         'Mard said to his wife.'

      b. Gorani Gawraǰu (Mohammadirad 2022a: A, 0079)
         m-wā=ya   dāyka=y   čīman
         IND-say.PRS.3SG:A=DRCT mother=EZ PN
         'She says to the mother of Čiman.'

      c. Southern Kurdish Bijar (Mohammadirad 2022c: A, 0004)
         wa pišī-ya   īš-ī
         to cat-DEF say.PRS-3SG:A
         'She says to the cat.'

Table 7: Frequencies of post-verbal (Po) nominal Addressees of all verbs vs. nominal Addressees of 'say/tell'

	G Gawraǰu			CK Sanandaj			SK Bijar		
	n	n Po	Po	*n*	n Po	Po	*n*	n Po	Po
Addressees of 'say/tell'	13	6	46%	13	13	92%	23	0	0%
Addressees (total)	25	10	40%	22	16	72%	52	2	4%

It is clear that Addressees in Central Kurdish Sanandaj and Gorani Gawraǰu are realised far more frequently rightward after the predicate than Addressees in the Southern Kurdish Bijar. Note, however, that the rate of post-verbal Addressees in the latter remains near zero (only two tokens out of 52 have postverbal ordering). Thus, the preverbal position of Addressees in Southern Kurdish Bijar is independent of the type of verb; it is a structural position.

It is also notable that Addressees of 'say/tell' statistically yield even more post-predicate tendency, reflecting that Addressees of 'say/tell' are more clearly associated with the notion of endpoint than verbs like 'speak' (cf. Haig 2022: 359).

As for the flagging strategy, prepositional flagging remains the primary mode of expressing Addressees across the dialects investigated here. Circumpositions rarely occur, and if they do, the pre-verbal position is the only option, in line with the tendency reported for Northern Kurdish in Haig (2022).

3.16 Place constituents

Place constituents here refer to arguments which denote 'static location' in clauses like 'He works at a factory.'; see Table 8 for the rate of post-predicate place constituents.

(19)   a.   Southern Kurdish Bijar (Mohammadirad 2022c: A, 0079)
     la sar ay   gūl=a   mālāw-in
     on top DEM.PROX pond=DEIC bathe.PRS-3PL:S
     'They bathe at this pond.'

     b.   Central Kurdish Sanandaj (Mohammadirad 2022b: A, 0012)
     šaw la kēf   řaš=ā   na-xaf-in
     night in mountain black=POST PROH-sleep.PRS-2PL:S
     'Do not sleep the night in the black mountain!'

Table 8: Frequencies of post-verbal (Po) nominal Addressees of all verbs vs. nominal Addressees of 'say/tell'

	G Gawraǰu			CK Sanandaj			SK Bijar		
	n	n Po	Po	**n**	n Po	Po	**n**	n Po	Po
Place arguments	60	25	42%	115	20	17%	65	10	15%

    c. Gorani Gawraǰu (Mohammadirad 2022a: B, 0352)
       ča=tān   waš ka(rd) a(ž) ka=y   lālo
       what=2PL:A good do.PST in   house=EZ uncle
       'What did you prepare in your uncle's house?'

Comparing the post-predicate realisation of place constituents to Goals (see 3.12) reveals that Goals differ significantly from Place constituents in post-verbal occurrence. Nevertheless, Place constituents are more post-posed than, say, direct objects, suggesting that the notion of 'location,' whether endpoint or not, triggers extraposition in the Kurdic varieties of the Zagros region.

3.17 Place constituents of a copular verb

Locational copula construction s are clauses in which the place constituent is a complement of the copular verb, e.g. 'I am at home.' In Central Kurdish Sanandaj and Gorani Gawraǰu, the predicate is an existential copula in such constructions, which requires the place complement to appear post-verbally. This existential construction is limited to the present tense, though; in the past tense, the past base of the verb 'to be' is used as the predicate, and its complement is generally realised preverbally. Southern Kurdish Bijar, on the other hand, consistently puts the place constituent before the copula verb; see Table 9 for percentages. [7]

(20) a. Gorani Gawraǰu (Mohammadirad 2022a: G, 1306)
       zangoł=a ziřa   hā   pišt=e kanü
       bell=EZ swinging EXIST.3SG:S back=EZ flour_bin
       'The swinging bell is in the back of bin of flour.'

---

[7]The high percentage of post-verbal place constituents of present copula verbs might be an indication of Aramaic influence in the core Gorani speech zone (Don Stilo, p.c.). However, it is noticeable that in NENA Sanandaj, locative complements of copula verbs are by default realised pre-verbally (see Khan 2009).

Table 9: Frequencies of post-verbal (Po) place arguments of present tense copula constructions in three Kurdic doculects

	G Gawraǰu			CK Sanandaj			SK Bijar		
	n	n Po	Po	*n*	n Po	Po	*n*	n Po	Po
Place arguments	6	4	66%	7	7	100%	11	0	0%

> b. Central Kurdish Sanandaj (Mohammadirad 2022b: K, 1071)
> *hā   la māł-ē   tārīk=ā*
> EXIST.3SG:S in house-INDF dark=POST
> '(He) is in a dark house.'

> c. Southern Kurdish Bijar (Mohammadirad 2022c: D, 0397)
> *kēnī   la pāl daryā d-ū*
> spring in side sea   IND-be.PRS.3SG:S
> 'The spring is next to the sea.'

While the number of tokens is low and no categorial conclusions can be made, it is clear that the Kurdic varieties in the Gorani zone of influence have a clear tendency for post-verbal positioning of place arguments of copular verbs. In contrast, Southern Kurdish Bijar opts for an opposing tendency, namely preverbal placement.

3.18 Light verb complements

In the Kurdic dialects investigated here, light verb complements of certain types of complex predicates appear after the light verb, e.g., Central Kurdish Sanandaj *kaft=a řē* [fall.PST.3SG:S=DRCT road] 'He set out.' In such predicates, the light verb is obligatorily followed by the directional clitic. A variety of light verbs can be used in such constructions, e.g., 'do,' 'fall,' 'come,' 'sit,' 'grab,' etc. It is notable that several of these light verbs involve motion verbs, e.g. 'fall,' 'come,' and the complement can be considered a metaphorical Goal. The light verbs used here can sometimes have an 'inceptive' sense.

> (21) a. Central Kurdish Sanandaj (Mohammadirad 2022b: D, 484)
> *a-kaf-ēt=a   xwašī=yā*
> IND-fall.PRS-3SG:S=DRCT happiness=POST
> 'He will get rich. (Lit. He will fall into happiness)'

b. Gorani Gawraǰu (Mohammadirad 2022a: A, 0061)
   hānī-aka   m-ā=ya   qisa
   spring-DEF IND-come.PRS.3SG:S=DRCT speech
   'The water spring starts to speak.'

c. Southern Kurdish Bijar (Mohammadirad 2022c: E, 677)
   dāłik=ī   girt=ay=a   bāwiš=aw
   mother=3SG:POS grab.PST.3SG:A=3SG:O=DRCT hug=POST
   'Her mother hugged her.'

In a similar manner, in the Bahdini variety of Northern Kurdish, some light verb complements occur after the light verb, followed by the directional particle. The difference is that the postverbal placement of the complement seems to be the case only with the light verb *kirin* 'do' (cf. Haig 2022: 344).

This might suggest that we are dealing with a common Kurdish syntax. However, it is notable that within Central Kurdish, the post-verbal placement of light-verb complement is fading out towards the northern dialects. The following examples show different treatment of the prepositional light verb complements in the southern dialect of Sanandaj vs., the northern dialects of Central Kurdish.

(22) a. Central Kurdish Sanandaj
        hāt=a   jiwāw
        come.PST.3SG:S=DRCT answer
        'He started to speak.'

     b. Central Kurdish Shaqlawa (Khan et al. 2022: 194)
        ba jiwāb   hāt
        to answer come.PST.3SG:S
        'He started to speak.'

(23) a. Central Kurdish Sanandaj (Mohammadirad 2022b: F, 0644)
        kaft=a   zawī
        fall.PST:S=DRCT ground
        '[The hat] fell on the ground.'

     b. Central Kurdish Mukri (Öpengin 2016: 251)
        be ʕerz-ī   dā   kewt
        to ground-OBL POST fall.PST.3SG:S
        '[The tree trunk] fell down on the ground.'

Assuming that the directional clitic in Central Kurdish Sanandaj is the reduced form of preposition *ba/we*, then the differences between Central Kurdish Sanan-

daj and Northern Central Kurdish varieties boils down to pre-verbal vs. post-verbal treatment of a light-verb complement, which appears in the form of a prepositional phrase. While a full investigation of this division within Central Kurdish awaits future research, it can be seen that Central Kurdish dialects opt for the reversed linear positioning of the light-verb complement. It is interesting to note that in parallel constructions, postverbal positioning is also the case in Hawrami Takht , geographically neighbouring Central Kurdish Sanandaj.

(24)   a.   Takht Hawrami (Mohammadirad In review: HB.31)
           ama=we   *zuwab*
           come.PST.3SG:S=COMPL answer
           'He started to speak.'

       b.   Takht Hawrami (Mohammadirad In review: ŽM.27)
           asê=nê=m   *cîya*
           leave.PST.PTCP.PL=COP.3PL:O=1SG:A place
           'I left them behind.'

The post-predicate placement of light-verb complements in Central Kurdish Sanandaj is quite unusual within the larger context of Central Kurdish. It seems reasonable to suggest that this is a further aspect of Central Kurdish Sanandaj word order, which can be related to an assumed Gorani substrate within Central Kurdish Sanandaj, as argued in Mohammadirad (2024).

3.19 Other obliques

In addition to the constituents mentioned in the previous sections, a variety of other oblique arguments, here collectively referred to as "other obliques", can be realised post-predicatively. These include instruments, comitatives, beneficiaries, and sources (see Table 10 for figures). By way of example, the placement of beneficiaries is illustrated.

(25)   a.   Central Kurdish Sanandaj (Mohammadirad 2022b: C, 0355)
           nān   *Ø-san-ē*   *bo mināɫ-akān=ī*
           bread SBJV-buy.PRS-3SG:A for child-DEF.PL=3SG:POS
           'That she buy bread for her children.'

       b.   Gorani Gawraǰu (Mohammadirad 2022a: D, 0592)
           ī   *kūw=a*   *a(řā) tu* *bi-tāš-ū*
           DEM mountain=DEIC for   2SG SBJV-hammer.PRS-1SG:A
           'I may hammer this mountain for you.'

c. Southern Kurdish Bijar (Mohammadirad 2022c: A, 0072)
   pīnačī=ya   kawš arā=y   dürn-ī
   cobbler=DEF shoes for=3SG:R sew.PRS-3SG:A
   'The cobbler sews the shoes for him.'

Table 10: Frequencies of post-verbal (Po) instruments, comitatives, sources, and beneficiaries in three Kurdic doculects

	G Gawraǰu			CK Sanandaj			SK Bijar		
	n	n Po	Po	*n*	n Po	Po	*n*	n Po	Po
Obliques	196	76	39%	279	86	31%	238	76	32%

It can be seen from Table 10 that instruments, comitatives, sources, and beneficiaries tend less to be placed rightward to the verb than, say, recipients. The reason could be that, unlike Goals and Recipients, these oblique arguments are not directly involved with endpoint semantics, meaning that they cannot readily be interpreted as endpoint arguments in the transfer of action.

3.20 Summary of post-verbal placement of constituents

In the previous subsections, the placement of different constituents relative to the verb was investigated across three Kurdic varieties confined within Zagros mountains. It was seen that these historical OV languages exhibit remarkable drift towards head-initial syntax, contrary to the predictions of head directionality hypothesis (see Dryer 1992). The head-initial configurations are shown in Table 11.

Table 11: Head-initial configurations

Noun	Adjective
Possessed	Possessor
Matrix clause	Complement clause
Auxiliary	Main verb
Complementizer	Complement clause
Verb	Goal
Verb	Recipient

Among all arguments, direct objects and copula complements (except for a subset of place constituents in copula constructions) are the most stable in their preverbal placement. Other constituents exhibit various degrees of rightward drift, as illustrated in Figure 2.[8]
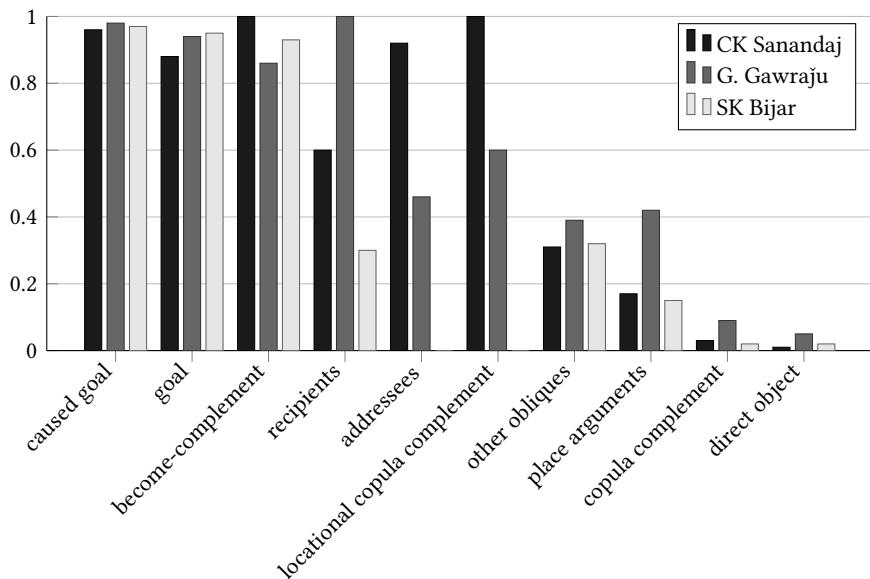
Figure 2: Post-predicate placement of different constituents across Kurdic dialects

As can be seen from Figure 2, the three dialects share the commonality of having the highest rate of post-predicate placement for caused goals and goals of verbs of movements. Similarly, they exhibit nearly the same rate of post-predicate placement for nominal complements of 'become,' which could be an indication of common Kurdish syntax (see Haig 2022: 342–343 for Northern Kurdish data).[9] Likewise, "other obliques" exhibit similar ratios of post-predicativity across these dialects. Overall, the high rate of post-predicate obliques, both end-point constituents and non-endpoint ones, suggests these varieties of Kurdish are candidates for Hawkins's (2008) OVX type.

It is notable that in two word order configurations, namely V-Addressees and V-locational copula arguments, Gorani Gawraǰu and Central Kurdish Sanandaj

---

[8]Note that the percentages given for Addressees are based on the placement of the nominal/non-bound form of these constituents. Relatedly, the ratios for locational copula complement only contain the position of such constituents in present tense copula construction s (see §3.17 for the explanations).

[9]Though see §3.13 for the distinction between adjectival vs. nominal complements of inchoative 'become.'

Table 12: Post predicate nominal phrases and their flagging type

	n. Po	prep	bare n + direct clitic on the verb	pre-nominal relational nouns[a]	bare	other (postp, case, circ)
CK Sanandaj	485	36%	21%	14%	10%	19%
G Gawraǰu	453	21%	25%	20%	21%	13%
SK Bijar	410	22%	10%	25%	28%	15%

[a]The verb may take a directional clitic when a pre-nominal relational noun flags the following noun.

prefer post-predicate realisation, whereas Southern Kurdish Bijar opts for pre-verbal ordering. In the next section, it is seen that these opposing tendencies are motivated by the geographical distribution of Kurdish dialects.

Another parameter of interest is how post-predicate nominals are flagged. Here, we should be cautious of a hasty conclusion since in all doculects some post-predicate nominals are preceded by a directional particle on the verbs, which is reconstructible as a preposition even in the same doculect. Thus, in Table 12, nominals which are preceded by a directional particle on the verb are analysed in a different column. Combining the figures for directional particles and prepo-sitional flagging, we obtain 46% and 57% prepositional flagging of post-predicate elements in Gorani Gawraǰu and Central Kurdish Sanandaj, respectively. In con-trast, the proportion is only 32% for the Southern Kurdish Bijar. This suggests that the Kurdish doculects in the immediate Gorani zone of influence, i.e., Cen-tral Kurdish Sanandaj, exhibit higher rates of prepositional flagging than those further from the realm of Gorani influence. Alternatively, the difference in post-verbal flagging seems to be largely motivated by the higher number of bare NPs in Southern Kurdish-Bijar. Presumably, they result from an original *V=DRCT NP, which becomes V NP through the loss of the directional clitic on the verb. Note that pairwise testing of the differences between the three dialects using a Fisher's Exact Test shows that the differences are all significant at $p<0.05$.

It is also conceivable to consider pre-nominal relational nouns as emerg-ing prepositions. Combining the figures for directional particles, prepositional flagging, and pre-nominal relational nouns, yields a similar picture. We obtain 71% and 66% prepositional flagging for Central Kurdish Sanandaj and Gorani Gawraǰu, respectively, whereas the proportion for Southern Kurdish Bijar is 57%.

Nonetheless, the overall picture suggests that not one special type of flagging is favoured post-predicatively. It is only in Central Kurdish Sanandaj that prepositional flagging is relatively high in the postverbal slot. This suggests that the syntax is approaching the head-initial type, with both verbs and adpositions preceding their complements.

Another parameter of interest is the overall levels of prepositionality in these Kurdic dialects, regardless of the placement of flagged nominals. The procedure for quantifying this was as follows: Taking the three Kurdic WOWA datasets, I selected the total number of tokens in the following functions: ABL(ative); ADDR(essee); BEN(efactive); COM(itative); GOAL; GOAL-C(aused); INSTR(umental); LOC(ative); REC(ipient); REC-BEN (see Nourzaei & Haig 2024 [this volume]). I then extracted those that were flagged with prepositions or pre-nominal relational nouns or by a directional particle on the verb (lumped together as "prepositional") and those that were flagged with postpositions, or circumpositions (lumped together as "postpositional").

Table 13: Overall levels of prepositionality

	n. clauses	prepositional	postpositional
CK Sanandaj	620	61%	34%
G Gawraǰu	475	78%	3%
SK Bijar	602	85%	10%

It can be seen from Table 13 that Gorani Gawraǰu and Southern Kurdish Bijar have much higher rates of prepositionality than Central Kurdish Sanandaj. Conversely, the levels of postpositionality are much higher in Central Kurdish Sanandaj compared with the other two doculects. These figures show that, barring Central Kurdish Sanandaj, mixed adpositional typology cannot be considered a feature of Southern Kurdish Bijar and especially Gorani Gawraǰu, at least in terms of token frequency in discourse. In other words, these two dialects have significantly approached the head-initial syntax with regard to adpositional typology. One way of interpreting these contradicting figures is to consider Persian influence on doculects with the lowest levels of postpositionality, considering that Modern Persian also lacks postpositions (except for the direct object marking *-rā*, which is irrelevant to our discussion). Persian influence seems to have been direct in the case of Southern Kurdish Bijar and probably indirect through Southern Kurdish in Gorani Gawraǰu. Alternatively, it may also reflect the earlier influence of Aramaic, which was previously much more widely spoken in the region.

4 Areal/contact issues

According to Figure 3, much of the post-predicate syntax is similar in the investigated dialects. Thus, each of Goals, Caused Goals, Recipients, Complements of 'become,' "other Obliques" (instruments, sources, and comitatives), Complements of copular verbs, and Direct objects occur with more or less the same proportion in the post-predicate position. Near-categorical post-verbal placement of goals and recipients is common to all three varieties investigated here, and appears to be a feature shared by all varieties of Kurdish, Gorani and Zazaki.

The major area of differentiation is the position of Addressees, and locational complements of copula verbs. These two traits bring together Central Kurdish Sanandaj and Gorani Gawraǰu against the Southern Kurdish Bijar, which is situated further to the north. Given that much of the post-predicate syntax of these varieties is shared, the question is why Southern Kurdish Bijar prefers pre-predicate positioning of Addressees and locational copula complements. Some scenarios can be outlined here:

First, as discussed in the introduction, recent research on the Central Kurdish dialect of the Sanandaj region has uncovered evidence for a significant Gorani substrate, which is attributed to an earlier shift from Gorani to Kurdish, or at least a high level of Kurdish-Gorani bilingualism among Kurds (see Mohammadirad 2024 for a recent discussion). Historically, the Sanandaj area used to be part of an earlier and more extensive Gorani heartland. Language shift from Gorani to Kurdish in the region is documented, for instance, in the introduction to the book *Les dialectes d'Awroman et de Pawa*, which reports on the the linguistic situation at Sanandaj in 1900. The authors note that "learned people" in the city knew and spoke Maço (the epithet for Gorani/Hawrami/Awromānī, meaning 'S/he says'):

> À Sänä où le kurde est maintenant la langue commune hors des communautés persane, juive et syrienne, on prétendait que l'awromānī y avait été communément entendu autrefois [In Sänä (Sanandaj, Kurd. Sine), where Kurdish is now the common language outside the Persian, Jewish and Syriac communities, it was claimed that Awromānī had been commonly heard there in the past] (Christensen & Benedictsen 1921: 5)[10]

Assuming the shift scenario to be true, V-Addressee and copula-location orders in Central Kurdish Sanandaj can be instances of constructional calque (or

---

[10]See Khan & Mohammadirad (2024) for a detailed account of language shift in Sanandaj. Likewise, (Mahmoudveysi 2016: 3) reports that the vernaculars of speakers of Bēwänīǰī, Riǰābī, and Gähwārä̈ī localities around Kerend (Iran), which were investigated by Mann & Hadank (1930) as Gorani dialects, have now shifted to vernaculars of Southern Kurdish.

"metatypy" in terms of Ross 2019), meaning that the post-verbal placement of the mentioned constructions was calqued into the type of Kurdish in Sanandaj region to which Gorani speakers shifted. Additional support for a Gorani substrate in the word order domain comes from the opposing directionality of light-verb complements in Central Kurdish dialects, resulting in post-predicate linearisation of light-verb complement in Southern dialects (see §3.18). Note that in the case of Addressees and locational copula complements, Central Kurdish Sanandaj has extended post-verbal placement to a greater degree than Gorani Gawraǰu. The reason could perhaps lie in the fact that Gorani Gawraǰu probably does not faithfully represent the actual substrate variety of Gorani that must have been spoken in Sanandaj. It is geographically far from Sanandaj, quite isolated from other varieties of Gorani, and it has itself been Kurdicised to a large extent. Gorani dialects, geographically closer to Central Kurdish Sanandaj, would probably yield more interesting correlations; we consider this option below.

This brings us to the second scenario, which concerns the geographical distribution of features. If we assume that the Central Kurdish Sanandaj values on the features of V-Addressee and copula-place are indeed due to a Gorani substrate, we would expect to find similar values in a conservative variety of Gorani, particularly if geographically close to Central Kurdish Sanandaj. Similarly, if Southern Kurdish Bijar is more generally representative of Kurdish spoken further from the core of the earlier Gorani speech zone, we would expect the Bijar values to be closer to Kurdish varieties spoken further to the north. In order to test these predictions, Hawrami Takht, geographically close to Central Kurdish Sanandaj, and Central Kurdish Mukri (and Central Kurdish Bingird), geographically close to Southern Kurdish Bijar, were selected as control languages (see Figure 1).

To start with the Addressees, I tested the post-predicate realisation of Addressees of 'say/tell,' including nominal Addressees only, in the following varieties: Hawrami Takht (20 clauses, Mohammadirad In review), Central Kurdish Bingird (14 clauses, MacKenzie 1962: 136–170), and Central Kurdish Mukri (12 clauses, Öpengin 2016). Figure 3 exhibits the ratio of post-predicate nominal Addressees in the sample.

The resulting data confirms our hypothesis that the postverbal realisation of nominal Addressees is areally confined to the south of the Central Kurdish speech zone, where we assume a Gorani substrate. Interestingly, in the NENA dialect of Sanandaj, Addressee arguments of 'say/tell' are 100% post-predicate (see Noorlander 2022), suggesting further that the word order profile of Addressees is areally defined. An areally-mediated shift in Addressee placement is also documented for Northern Kurdish in Haig (2022). In the northern Central Kurdish
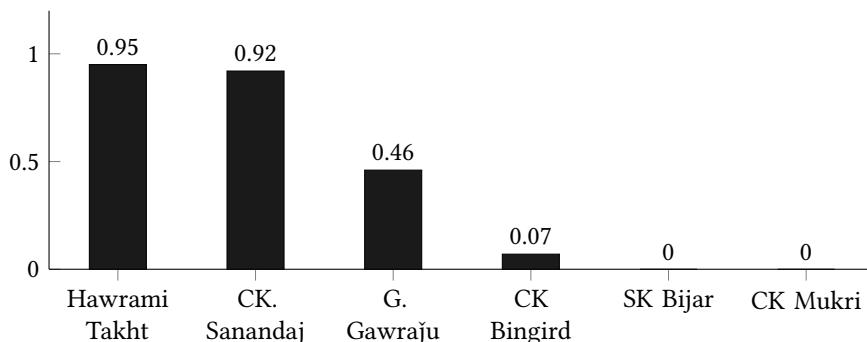
Figure 3: Post-predicate ratio of nominal Addressees

dialects of Mukri and Bingird, the reverse order Addressee-Verb is prevalent, tying in with the ordering in the geographically neighbouring Southern Kurdish Bijar.

It is also notable that Central Kurdish Sanandaj shows much closer correspondence with the neighbouring Hawrami Takht than with the Kurdicised Gorani Gawraǰu. The latter is not really a good representative of the original assumed state of, e.g. Gorani as once spoken in Sanandaj, and it is outside of the Central Kurdish region. Indeed, Central Kurdish Sanandaj better reflects the original Gorani word order than Gorani Gawraǰu. The real conclusion seems to be the structural proximity of Central Kurdish Sanandaj and Hawrami Takht, in line with the assumption of a Gorani substrate in Central Kurdish Sanandaj.

Relatedly, the opposing directionality in the placement of place arguments of copula construction s in Central Kurdish Sanandaj and Southern Kurdish Bijar is matched by the same tendencies in immediate neighbouring languages,[11] thus in the southern Central Kurdish speech zone place arguments of copula verbs are predominantly post-predicate, whereas the reverse ordering holds in the north, see Figure 4.

An investigation of these minor word order features thus reveals commonalities between southern Central Kurdish dialects, here represented by Central Kurdish Sanandaj and Gorani (represented by Hawrami Takht). I propose that these differences can be most plausibly explained through the greater influence of Gorani in the southern part of the Central Kurdish speech zone, particularly due to Gorani speakers shifting to Kurdish. Northern Central Kurdish dialects, like Central Kurdish Mukri, in which contact with Gorani was probably not as

---

[11]The number of test clauses is 12 for Central Kurdish Mukri, and 10 for Hawrami Takht.
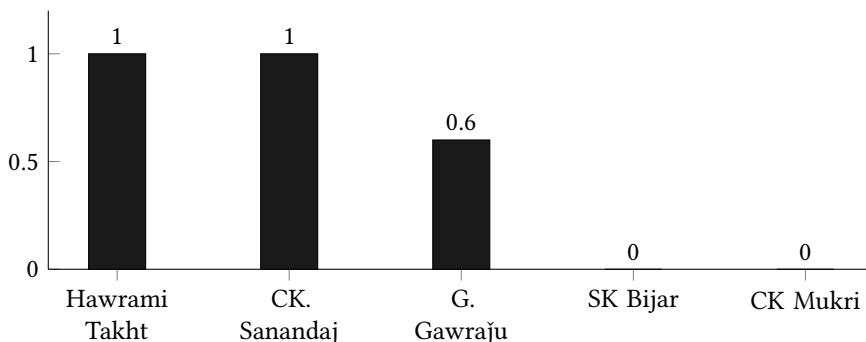
Figure 4: Post-predicate ratio of place constituents in copular constructions

intense as it was in the south, lack the effects documented here (see Moham-madirad 2024 for other features which highlight the impact of Gorani substrate in creating north/south division of Central Kurdish dialects). The findings also suggest, conversely, that the remnant Gorani variety of Gawraǰu has diverged from a presumably more conservative state of Hawrami Takht and drawn closer to the more widespread pattern found in Kurdish.

Note that the minor word order patterns considered in this section have gen-erally gone under the radar of the larger-scale approaches to Kurdish word order and language contact but illustrate the importance of more detailed case studies in identifying local patterns of contact.

5 Conclusion

Among Iranian languages, Kurdish varieties are the westernmost outlier at the intersection with VO languages. This has resulted in the preponderance of head-initial word order configurations in this group of varieties, as documented in Haig (2015) and subsequent literature. This study highlighted the word order profile of oblique arguments in three Kurdic dialects, namely Central Kurdish Sanandaj, Gorani Gawraǰu, and Southern Kurdish Bijar. Major patterns of con-stituent ordering in these languages match to a large extent; for example, they all have rigid object-word order: Goals and nominal Recipients are predominantly postverbal. However, these dialects exhibit microvariation concerning the posi-tioning of Addressees and light-verb complements, and locational copula comple-ments. These differences were claimed to represent areal patterns and warrant a north-south distinction of Central Kurdish dialects triggered by the Gorani sub-

strate in the southern Central Kurdish dialects. It appears that the northern Central Kurdish dialects have preserved the generally assumed Old Iranian pattern of preverbal realisation of Addressees, reinforced through contact with Azeri Turkic varieties. In contrast, the southernmost dialects have shifted to post-verbal Addressees and post-verbal place complements in copula constructions.

Abbreviations

1	1st person	NEG	negative
2	2nd person	NUM	numeral
3	3rd person	N	Noun
A	transitive subject	OBL	oblique
ADJ	adjective	O	Direct Object
AUX	auxiliary	PERF	perfect
CK	Central Kurdish	PL	plural
CLF	classifier	PN	Proper Noun
COMP	complementizer	POS	possessor
COMPL	completive	POST	postposition
COP	copula	PROH	prohibitive
DEF	definite	PROX	proximal
DEIC	deictic	PRS	present
DEM	demonstrative	PST	past
DIST	distal	PTCL	particle
DRCT	directional particle	PTCP	participle
EXIST	existential particle	R	Flagged oblique argument
EZ	ezafe	S	Intransitive Subject
G	Gorani	SBJV	subjunctive
IND	indicative	SG	singular
INDF	indefinite	SK	Southern Kurdish
IPFV	imperfective	SPEC	specific

Acknowledgements

References

Asadpour, Hiwa. 2022a. *Typologizing word order variation in northwestern Iran.* Frankfurt am Main: Goethe-Universität. (Doctoral dissertation).

Asadpour, Hiwa. 2022b. Word order in Mukri Kurdish - The case of incorporated targets. In Hiwa Asadpour & Thomas Jügel (eds.), *Word order variation: Semitic, Turkic and Indo-European languages in contact*, 63–87. Berlin: De Gruyter Mouton.

Bailey, Denise. 2018. *A grammar of Gawraju Gurani.* Georg-August-Universität Göttingen. (Doctoral dissertation).

Christensen, Arthur & Åge Meyer Benedictsen. 1921. *Les dialectes d'Awroman et de Pawa.* French. Kobenhavn: B. Lunos.

De Morgan, Jacques. 1904. *Mission scientifique en perse. études linguistiques. dialectes kurdes, tome 5.* Paris: E. Leroux.

Dryer, Matthew S. 1992. The Greenbergian word order correlations. *Language* 68. 81–138.

Fattah, Ismaïl Kamandâr. 2000. *Les dialectes kurdes méridionaux: étude linguistique et dialectologique (Iranica 37).* Leuven: Peeters.

Haig, Geoffrey. 2015. Verb-goal (VG) word order in Kurdish and Neo-Aramaic: Typological and areal considerations. In Geoffrey Khan & Lidia Napiorkowska (eds.), *Neo-Aramaic and its linguistic context*, 407–425. Piscataway, NJ: Gorgias Press.

Haig, Geoffrey. 2017. *Post-predicate arguments in Iranian languages. Keynote lecture, seventh international conference on Iranian languages (ICIL7), Lomonosov Moscow State University, 28-30th August 2017.*

Haig, Geoffrey. 2022. Post-predicate constituents in Kurdish. In Yaron Matras, Geoffrey Haig & Ergin Öpengin (eds.), *Structural and typological variation in the dialects of Kurdish*, 335–377. Cham: Springer International Publishing. DOI: 10.1007/978-3-030-78837-7_8.

Haig, Geoffrey, Paul M. Noorlander & Nils N. Schiborr. In press. Which word order features are stable in a contact setting? Corpus-based evidence from the Western Asian transition zone. In Jeroen Darquennes, Joe Salmons & Wim Vandenbussche (eds.), *Language contact: An international handbook*, vol. 2. Berlin: De Gruyter Mouton.

Haig, Geoffrey, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils Schiborr. 2024. Post-predicate elements in the Western Asian Transition Zone: Data, theory, and methods. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based ap-*

proach to areal typology, 3–54. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266331.

Hawkins, John A. 2008. An asymmetry between VO and OV languages: The ordering of obliques. In Greville Corbett & Michael Noonan (eds.), *Case and grammatical relations: Studies in honor of Bernard Comrie*, 167–190. Amsterdam: Benjamins. DOI: doi.org/10.1075/tsl.81.08ana.

Khan, Geoffrey. 2009. *The Jewish Neo-Aramaic dialect of Sanandaj*. Piscataway, NJ: Gorgias Press.

Khan, Geoffrey & Masoud Mohammadirad. 2024. *A study of the impact of iranian on neo-aramaic*. Berlin, Boston: De Gruyter Mouton. DOI: doi:10.1515/9783111209180.

Khan, Geoffrey, Masoud Mohammadirad, Dorota Molin & Paul M. Noorlander. 2022. *Neo-Aramaic and Kurdish folklore from Northern Iraq*. Cambridge: Open Book Publishers. DOI: 10.11647/OBP.0307.

MacKenzie, David N. 1962. *Kurdish dialect studies Vol. II*. London/New York: Oxford University Press.

Mahmoudveysi, Parvin. 2016. *The meter and the literary language of Gūrānī poetry*. Hamburg: Universität Hamburg. (Doctoral dissertation).

Mahmoudveysi, Parvin & Denise Bailey. 2013. *The Gorani language of Zarda, a village of west Iran: Texts, grammar, and lexicon*. Wiesbaden: Reichert.

Mahmoudveysi, Parvin, Denise Bailey, Ludwig Paul & Geoffrey Haig. 2012. *The Gorani language of Gawraǰū, a village of West Iran: Texts, grammar, and lexicon*. Wiesbaden: Reichert.

Mann, Oskar & Karl Hadank. 1930. *Mundarten der Gûrân, besonders das Kändûläî, Auramânî und Bâdschälânî, bearbeitet von Karl Hadank*. Berlin: Verlag der Preussischen Akademie der Wissenschaften in Kommission bei Walter De Gruyter & Co.

McCarus, Ernest. 2009. Kurdish. In Gernot Windfuhr (ed.), *The Iranian Languages*, 587–633. London & New York: Routledge.

Mohammadirad, Masoud. 2020. *Pronominal clitics in Western Iranian languages: Description, mapping, and typological implications*. Paris: Université Sorbonne Nouvelle Paris 3. (Doctoral dissertation). https://tel.archives-ouvertes.fr/tel-02988008/document.

Mohammadirad, Masoud. 2022a. Gorani (Gawraǰū). In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation*. Bamberg: Bamberg University Press. multicast.aspra.uni-bamberg.de/resources/wowa/.

Mohammadirad, Masoud. 2022b. Kurdish (Central, Sanandaj). In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation*. Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

Mohammadirad, Masoud. 2022c. Kurdish (Southern, Bijar). In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation*. Bamberg: Bamberg University Press. multicast.aspra.uni-bamberg.de/resources/wowa/.

Mohammadirad, Masoud. 2024. Gorani substrate within Kurdish: Evidence from southern dialects of Central Kurdish. *Journal of Historical Linguistics.* DOI: 10.1075/jhl.23001.moh.

Mohammadirad, Masoud. In review. A grammar of Hewramî.

Mohammadirad, Masoud. In press. Bound argument ordering across Central Kurdish and the Gorani substrate. In Jaffer Sheyholislami, Geoffrey Haig, Haidar Khezri, Salih Akin & Ergin Öpengin (eds.), *Oxford handbook of Kurdish linguistics.* Oxford: Oxford University Press.

Noorlander, Paul M. 2022. NE Neo-Aramaic (Jewish Sanandaj). In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA – Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation*. Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

Nourzaei, Maryam & Geoffrey Haig. 2024. Balochi: A cross-dialect investigation of post-verbal elements. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 121–154. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266337.

Öpengin, Ergin. 2016. *The Mukri variety of Central Kurdish: Grammar, texts, and lexicon.* Wiesbaden: Reichert.

Öpengin, Ergin & Masoud Mohammadirad. 2022. Pronominal clitics across Kurdish: Areal distribution, structural variation, and diachrony. In Yaron Matras, Geoffrey Haig & Ergin Öpengin (eds.), *Structural and typological variation in the dialects of Kurdish*, 181–237. Cham: Springer International Publishing. DOI: 10.1007/978-3-030-78837-7_5.

Parizadeh, Mehdi & Mohammad Rasekh-Mahand. 2024. Post-predicate elements in Early New Persian (10–13th Century CE). In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based*

approach to areal typology, 231–244. Berlin: Language Science Press. DOI: 10. 5281/zenodo.14266345.

Querry, Amédée. 1896. Le dialecte Guerruci. *Mémoires de la Société de Linguistique de Paris* 9. 1–23.

Rasekh-Mahand, Mohammad, Elham Izadi, Mehdi Parizadeh, Geoffrey Haig & Nils Schiborr. 2024. Post-predicate elements in modern colloquial Persian: A multifactorial analysis. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 197–230. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266343.

Ross, Malcolm. 2019. Syntax and contact-induced language change. In Anthony P. Grant (ed.), *The Oxford handbook of language contact*, 123–154. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780199945092.013.5.

Stilo, Donald. 2009. Circumpositions as an areal response: The case study of the Iranian zone. *Turkic Languages* 13. 3–33.

Part III

Case studies from the Caucasus and the Black Sea

Chapter 10

Post-predicate elements in Kartvelian and East Caucasian

Diana Forker

Friedrich Schiller University Jena

Kartvelian (or South Caucasian) and East Caucasian (or Nakh-Daghestanian) languages are usually described as "flexible SOV" languages which allow all logically possible word order permutations in main clauses. In this paper, I explore post-predicate elements in both language families and show that, in general, post-predicate elements are common in natural texts and influenced to various degrees by features such as genre/style, semantic role, information structure, heaviness and also language contact.

1 Introduction

Kartvelian (or South Caucasian) and East Caucasian (or Nakh-Daghestanian) languages are two of the three indigenous language families of the Caucasus. Kartvelian is the largest indigenous family in the Caucasus in terms of speakers, mainly due to Georgian, which is the national language of the Republic of Georgia. East Caucasian is the largest Caucasian family in terms of numbers of languages. In both language families word order has been studied, and the general consensus is that the languages have free word order at the clausal level, with SOV being in some way classified as basic. In this paper, I want to discuss post-predicate items in both families based on the available literature and corpus data. I will examine a number of features that influence the availability of elements after the verb:

- morphosyntactic properties (grammatical function, word class)

- pragmatic properties (heaviness, information structure)

- extralinguistic features (language contact)

I will concentrate on Georgian as representative of Kartvelian, but also include data from Laz and two other Kartvelian languages. With respect to East Caucasian I will rely on corpus data from Sanzhi Dargwa, Chirag Dargwa, Tabasaran and Hinuq. However, the corpus data originate from different sources and have been annotated in different ways, so the results are not always directly comparable. This paper pursues the gradient approach to word order as advocated by Levshina et al. (2023), and generally applied in this volume.

2 Post-predicate items in Kartvelian

2.1 Word order profile of Kartvelian languages

Kartvelian languages have head-final noun phrases with e.g., demonstratives and numerals preceding the noun, but admit some exceptions. In all languages, it is possessive pronouns that are most commonly positioned after the noun.

Megrelian (or Mingrelian) and to a lesser extent also Georgian can have postnominal modifiers such as adjectives and relative clauses and partially also modifying genitives (Aronson 1991, Harris 1991, Pourtskhvanidze 2015: 169–170), though postposed adjectives and genitives in Georgian are described as archaic and following Old Georgian patterns (Testelets 1998b). In the development of Modern Georgian from Old Georgian, a clear shift from head-initial to head-final order in the noun phrase has been observed (see the references in Testelets 1998b). In Laz, the tendency to postpose possessive pronouns is so strong that occasionally, possessives pronouns may even follow a postposition (Holisky 1991). In Svan, postposed modifiers, including possessive pronouns, are very rare and archaic and seem to be restricted to poetry and lyrics (Schmidt 1991: 537; Tuite 1998; see Testelets 1998b for examples).

Kartvelian languages have postpositions (with a few exceptions, for Georgian see Harris 2000). Auxiliaries follow the lexical verb, which is usually in a non-finite form, but the reverse order is not unusual (Harris 2000). The position of relative clauses depends on the formal type of relative clause. Relative clauses built with participles precede the head while those formed with a relative pronoun follow the head. Relative clauses with a gap may follow or precede the head. (Harris 2000). All Kartvelian languages have subordinating conjunctions in clause-initial, clause-second, or preverbal position (Boeder 2021), and Megrelian

also has a clause-final subordinator (Boeder 2005: 70; Testelets 2021: 522–523). As mentioned, Kartvelian languages are usually characterized as having free word order at the clausal level, with verb final order as the unmarked pattern (Boeder 2005: 64). In the following section I will examine word order in Georgian because it is by far the most studied language among the Kartvelian languages.

2.2 Georgian: Previous studies on word order at the clausal level

After a series of elicitation tests Skopeteas (2021) comes to the conclusion that certain asymmetries between V-medial and V-final orders suggest that the basic word order is verb-final. One of his arguments is the position of prepositional complements such as comitatives and themes which are preferably placed as expected for verb-final languages (comitatives before themes) and have rigid scope in the basic order.

However, apart from a few exceptional cases, both V-O and O-V orders occur in free variation and V-O is not triggered by any special pragmatic or semantic configurations (Testelets 1998b, Asatiani & Skopeteas 2012, Skopeteas et al. 2009, Boeder 2021). All other orders are also attested in elicitation and natural texts (see, e.g. Pourtskhvanidze 2015: 161–162 for examples). With respect to frequency, based on data from the internet Skopeteas (2021) found that in a total of 925 non-idiomatic VPs 64.1% have O-V order while 35.9% gave V-O order. Other authors present different numbers. Vogt (1971) counted 50 randomly chosen pages in the influential novella, *jaq'os xiznebi* 'Jaqo's dispossessed' by Mikheil Javakhishvili, which was first published in 1924. S-O-V order is attested in 75% of the sentences; S-V-O in one sixth of them (Vogt 1971: 222). He also counted traditional folk tales (published in 1958) and got slightly different results: the subject occupies the initial position in two third of the examples; the direct object precedes or follows the verb with roughly the same frequency, which means that O-V and V-O are equally frequent. Stilo (2014: 421) counted indefinite and definite direct objects in a small corpus of colloquial (spoken) Tbilisi Georgian that consisted of 500–600 clauses. He found that 60.2% of the definite objects occurred before the verb (O-V) while 39.8% occurred after the verb (V-O). For indefinite direct objects the differences increase a bit: O-V is found in 64.9% of the clauses whereas V-O in only 35.81%. These numbers are comparable to Skopeteas counts of internet texts. In another unpublished data set of conversational informal Georgian the total number of nominal direct objects (excluding pronouns) that are V-O is 160 out of 364 (around 44%), (Stilo 2018). For definite objects, the V-O figure is 96 of 206 (around 47%) and for indefinite objects V-O is found for 64 out of 158 (about 40%).

Skopeteas et al. (2009) also suggest that stylistic factors might have an effect, which is, in fact, also suggested by Vogt's data, since he found a difference between a classical written novella and folk tales that belong to traditional oral literature. Finally, Skopeteas (2021) cites a study by Apridonidze (1986) that shows that the more constituents a clause has the more likely it is that the verb does not occur in final position but that the clause contains post-predicate elements.

In sum, Georgian seems to have S-O-V as basic word order in terms of a rather formal theory of grammar and a certain degree of variation in terms of actually attested patterns in natural data. For the latter, style/genre and number of constituents in the clause play a role, but possibly also other factors such as grammatical function, definiteness, pragmatics and heaviness of constituents.

2.3 Georgian word order in a small corpus study

For this paper, I recorded nine oral texts and conducted a small corpus study. The texts have been elicited from nine speakers of Georgian by means of the Pear Story Movie (Chafe 1980).[1] The corpus contains 2,901 words, and 644 clauses, of which 556 contain a verb and at least one argument or other constituent with a relevant grammatical function for this study. The other 88 clauses have been left out because they only consist of the verb, or verb with particle or adverb outside the purview of this study.

I classified arguments and adjuncts into the following grammatical functions, based on semantics rather than formal marking:

- S: subject of monovalent verb

- A: subject of bivalent verb

- P: object of bivalent verb

- Goal: spatial Goal, objects such as addressee, recipient

- location: spatial location

- source: spatial origin, source

- instrument: instrument or tool of any sort

- beneficiary

---

[1]The film is available at https://www.youtube.com/watch?v=bRNSTxTpG7U and https://drive. google.com/file/d/1jF6vtUdOlBN9LFzJ_NujWfP9Ef6ObK9K/view.

I counted only nominal referents fulfilling these grammatical functions (including postpositional phrases) and excluded temporal adverbials and some other types of adjuncts as well as adverbs and particles. I included various types of subordinate clauses, in particular relative clauses, complement clauses and purpose clauses and also non-declarative utterances (of which there were only very few).

As a starting point, I compared my counts with Apridonidze's (1986) findings, according to which the probability of a constituent occurring after the verb increases when there are more constituents in the clause. In fact, this is not surprising, since the more constituents a clause has, the more possibilities there are to put at least one after the verb. However, my own data are not as neat as his data and also differ quantitatively. Table 1 shows my data for clauses with at least two and up to five constituents (including the verb). I counted all nominal constituents, but did not include pronouns, particles or adverbs. There is no monotonic decrease in the probability for verb-final order, but a clear tendency. Of clauses with only two constituents (e.g. a verb and S or a verb and P), around 40% have X-V order and 60% V-X order. For clauses with three constituents (e.g. verb, A and P or verb S and Goal) the percentage of V-X order increases to 73%.

Table 1: Word order at the clausal level in relation to the number of constituents

# constituents	2	%	3	%	4	%	5	%
X-V	122	40.67	48	26.82	7	11.11	2	15.38
V-X	178	59.33	131	73.18	56	88.89	11	84.61
total	300		179		63		13	

Table 2 summarizes first the constituent order patterns of all clauses and, second, the position of S, A and P (including relative pronouns and demonstrative pronouns). In my data, around two third of the clauses are not verb-final but contain at least one nominal or pronominal constituent after the verb. Compared to the other studies cited above this is a considerably higher amount of post-predicate items. Subjects of monovalent verbs (S) and to an even larger extent nominal subjects of bivalent verbs (A) overwhelmingly precede the verb, whereas objects of bivalent verbs (P) tend to follow the verb. These numbers are astonishing in the light of the data cited above. The only hypothesis that comes to my mind is that this might reflect a difference in written vs. spoken language (see Haig et al. 2024 [this volume] and Rasekh-Mahand et al. 2024 [this volume]

for similar findings), but Stilo's data also represent spoken language, albeit not elicited monologues.

Table 2: The position of S, A and P

	clauses	%	S	%	A	%	P	%
X-V	186	33.45	121	69.12	93	93.00	100	38.76
V-X	370	66.55	54	30.85	7	7.00	158	61.24
total	556		175		100		258	

In order to check whether the part of speech had an influence on the position of arguments and adjuncts in the surveyed functions, first of all, I excluded all relative pronouns from the counts. Relative pronouns in my corpus are placed before the verb, which is their normal position (Harris 2000), and the majority of relative pronouns occur in the function of S or A. In a second step, I omitted all demonstrative pronouns used in anaphoric function from the counts. There were no personal pronouns for first and second person due to the type of stimulus used for the narratives. I thus cannot make any statements on the position of those pronouns. Table 3 summarizes the counts for nominal S, A and P arguments. As a comparison with Table 2 shows, there are only small differences in the percentages. Furthermore, there are not many demonstrative pronouns in my small corpus and from Table 4 it is clear that those pronouns do not differ much in their position/function from the nouns (and the very few indefinite pronouns).

Table 3: The position of S, A and P (excluding relative pronouns and demonstrative pronouns in anaphoric function)

	S	%	A	%	P	%
X-V	100	66.23	66	91.67	93	37.96
V-X	51	33.77	6	8.33	152	62.04
total	151		72		245	

Goals are most commonly placed after the verb, with an overall higher frequency of post-verbal placement than any other argument type investigated here. This is a feature that my Georgian corpus shares with all other spoken-language corpora in the WOWA data base (see Haig et al. 2024 [this volume]). Instruments

Table 4: The position of arguments and adjunctions expressed as demonstrative pronouns

	S	A	P	Goal	location	source	instrument	beneficiary
X-V	9	13	3	1	0	1	0	4
V-X	3	1	6	3	0	0	3	2
total	12	14	9	4	0	1	3	6

are also more frequently in post-verbal than in preverbal position. By contrast, referents expressing locations are usually found before the verb; cf. Table 5 and Table 6 for relevant figures. In (1) instrument, Goal and (metaphorical) location appear after the verb. For source and beneficiary, the distribution is roughly half-half and no clear tendency could be detected. The corpus also contains 38 complement clauses of which 35 are in a position after the matrix predicate.

Table 5: The position of other types of arguments and adjuncts

	Goal	%	loc	%	src	%	instr	%	ben	%
X-V	22	25.88	52	61.18	21	46.67	18	35.29	8	53.33
V-X	63	74.12	33	38.82	24	53.33	33	64.71	7	46.67
total	85		85		45		51		15	

Table 6: The position of other types of arguments and adjuncts (excluding relative pronouns and demonstrative pronouns in anaphoric function)

	Goal	%	loc	%	src	%	instr	%	ben	%
X-V	20	25.00	44	57.14	16	40.00	18	37.5	4	44.44
V-X	60	75.00	33	42.86	24	60.00	30	62.5	5	55.56
total	80		77		40		48		9	

(1) V-INST-GOAL-LOC[2]

Georgian (Georgian Pear Story Corpus)

mi-e-q'rdn-ob-a   *zurg-it* *k'ibe-s* *albat* *im-is*
PV-APPL-lean_on-TM-S.3SG back-INST stairs-DAT probably DEM.DIST-GEN

pikr=ši
thought=LOC

'(He) leans with the back to the stairs, probably in his thoughts.'

Then I looked into the position of new referents, more specifically of new human referents which are introduced into the Pear Stories one by one. Most speakers mentioned five human referents, and there was a very clear tendency to express them either as S or as P and put them in a position after the verb (Table 7). In example (2), the main protagonist of the film, a young boy, is introduced into the narrative in the function of S occurring as the last item in the clause.

(2) TIME-V-INST-S

Georgian (Georgian Pear Story Corpus)

cot'a *xan=ši*   *ga-mo-čn-d-eb-a*   *velosip'ed-it bič'i*
a_little period=LOC PV-PV-appear-INTR-TM-S.3SG bike-INST   boy

'After a little while a boy with a bike will appear.'

Then I counted the position of light versus heavy noun phrases with one, two, three, four or more words. In Table 7 relative clauses are excluded, i.e. all noun phrases that head relative clauses have been omitted from the counts and only noun phrases with demonstratives, adjectives and the like have been included. However, there are basically no differences between noun phrases that consist of one, two or three words. Only noun phrases containing four or more words show an increased tendency for a position after the verb.

Finally, Table 8 presents first of all the position of relative clauses and their heads. 29 out of a total of 34 nominal heads of relative clauses occur in a position after the verb. Second, I added the noun phrases with relative clauses to the counts in Table 7. To illustrate that with an example, we can look at (3): the NP 'that lower part of a tree' consists of two constituents before the head noun, namely a demonstrative and a genitive, and a relative clause following it. For the manner of counting displayed in Table 8 this NP has four constituents (demonstrative, genitive, head noun and relative clause). These numbers suggest that

---

[2]In this study of spoken Georgian, the concepts of GOAL and INST are used slightly differently from how they are used in other papers of this volume. I used case marking as a major indicator, i.e. dative marking for GOAL and instrumental case marking for INST.

Table 7: Newness (human referents) and heaviness of constituents (without relative clauses)

	new		4 (+) words		3 words		2 words		1 word	
X-V	11	25%	12	42.86%	30	53.57%	150	55.76%	233	54.57%
V-X	33	75%	16	57.14%	26	46.43%	119	44.24%	194	45.43%
total	44		28		56		269		427	

the strongest effect on the position is the presence of a relative clause in the NP, which leads to >80% post-verbal placement - and this seems to be irrespective of how many other constituents there are in the NP. Similarly, the corpus contains 38 complement clauses of which three occur before the verb and 35 after the verb. This can be generalized: if an NP contains a clausal constituent, it is nearly categorically likely to be post-verbal. The effect of number of constituents, on the other hand, is quite small by comparison, and is only really significant for +4 constituents vs. 1 constituent.

Table 8: Position of head noun of relative clause and heaviness (including relative clauses; all semantic roles)

	relative clause		4(+) constituents		3 constituents		2 constituents		1 constituent	
X-V	5	14.71%	15	37.50%	30	45.45%	151	54.12%	232	54.08%
V-X	29	85.29%	25	62.50%	36	54.55%	128	45.88%	197	45.92%
total	34		40		66		279		429	

(3) V-GOAL-REL
     Georgian (Georgian Pear Story Corpus)
     ga-i-vl-ian   am   x-is   ʒira-s
     PV-REFL-go-S.3PL DEM.PROX tree-GEN lower_bottom-DAT
     [roml=idana=c   uk'-ve   ʒirs   ar-is   ča-mo-sul-i
     which=ABL=ADD already down be-S.3SG PV-PV-go.PTCP-NOM
     mama-k'ac-i]
     father-man-NOM
     '(The boys) pass by the foot of this tree, from which already has come down the old man.'

Summarizing we can state that all kinds of arguments and adjuncts can occur after the verb, but direct objects (4), goals (1), (2) including indirect objects, and instruments (1), (2) are particularly prone to be placed after the verb, which means that the grammatical function has an impact on the position of the respective item. Furthermore, pragmatics plays a role: newly introduced (human) referents mostly follow the verb (2) (non-human referents have not been counted). Very heavy noun phrases and nouns heading a relative clause also tend to be positioned after the verb (3), but demonstrative pronouns in anaphoric function do not differ in their preferences from nominals.

(4) V-A-P
     Georgian (Georgian Pear Story Corpus)
     šemdeg   da-i-berṭq'-d-a   am   bič'-ma
     afterwards pv-refl-shake_out-impf-s.3sg dem.prox boy-erg
     šarval-i
     trousers-nom
     'Then this guy shook out his trousers.'

Due to the limits of my corpus, further research is needed that targets also first- and second-person pronouns, examines the impact of (in)definiteness and the positional properties of subordinate clauses with non-finite verbs such as participles and masdars.

2.4 Megrelian, Svan and Laz

What concerns Megrelian and Svan, post-predicate elements do not seem to be rare. We find subjects, direct objects, indirect objects, obliques such as instruments and others, temporal and spatial adverbials both in nominal as well as pronominal form in all grammatical descriptions surveyed (Harris 1991, Holisky 1991, Rostovtsev-Popiel 2021, Tuite 1998, Schmidt 1991).

Laz is the least flexible Kartvelian language with respect to word order, even though the claim by Testelets (2021: 518) that Laz is relatively strict verb final with only very few constructions allowing for a restricted range of post-predicate elements has to be rejected. In contrast to the other three Kartvelian languages, Laz is mainly spoken in Turkey and thus under heavy Turkish influence. Lacroix (2009: 737) makes some generalizations about post-predicate elements in Laz. They are mostly (i) known / topical, or (ii) part of an idiomatic expression, or (iii) new referents in introductory sentences, or (iv) specify a referent that has already been mentioned in the sentence. Laz has a much lower frequency of post-predicate elements when compared to Georgian. This is certainly true for the

texts in Kutscher & Genç (1998) and in Stilo & Lacroix (2021). However, otherwise it seems that roughly the same range of elements are allowed as in the other Kartvelian languages, both in elicitation and in natural texts (e.g. Kutscher & Genç 1998, Lacroix 2009): subjects, objects, obliques, adverbials (goals, locations). In the 11 Arhavi Laz texts collected by Lacroix, published in his grammar (Lacroix 2009) and coded for WOWA by Don Stilo (Stilo & Lacroix 2021), 400 items (noun phrases and adverbials) have been categorized with respect to their position: 391 occur before the verb (around 98%) and only 9 after the verb (around 2%). Items occurring after the verb serve as direct objects, addressees, locations, goals, and one is a possessed referent in a possessive construction. With these numbers, Laz is among the most consistently verb-final languages in the entire WOWA data set.

As shown for Georgian and just mentioned, in introductory sentences or, more generally, in contexts in which new referents are introduced into a narration, the new referents often follow the verb. These new referents are usually either subjects or direct objects as in (5) from Svan.

(5) Svan (Schmidt 1991: 539)
    ašxwin ləcte   otzəzax   bepšw
    once   water.to they.apparently.sent child.NOM
    'Once (they) sent a child to the water.'

But postverbal items can be topical, too. Example (6) from Megrelian and example (7) from Laz illustrate postverbal subjects that encode established referents.

(6) Megrelian (Rostovtsev-Popiel 2021: 557)
    k'in=i   mida-rt-es   o-nadir-u-ša   boš-ep-k
    back=EV PV-go-3SG.PST SUPINE-hunt-SUPINE-all boy-PL-ERG
    'The boys left for hunting again.'

(7) Laz (Holisky 1991: 469)
    i.bgar-u do xolo meyoč-u oxorža-k
    cry-3SG and again curse-3SG wife-ERG
    'The wife cried and cursed again.'

Heaviness might play a role. In (8), from Svan, the first main clause contains a postverbal focal object whereas the focal object in the second main clause is in preverbal position. The postverbal object of the first clause does not even directly follow the verb, but is separated from it by an inserted subordinate conditional clause. It is heavy, consisting of a participial relative clause and an adjective, which might be a reason for its postverbal position.

(8) Svan (Tuite 1998: 19)

eče-ži a-d-isg-x, [xoxra bepšw-ild-ær axa
there-at VER-put-SM-PL little child-DIM-PL.NOM if

*æt-[i]-dagr-i-w-x], **ežær-e le-pane xoxra dir-ild-ær-s***
PV-VER-die-SM-IMP-PL 3PL-GEN PTCP-consecrate little bread-DIM-PL-DAT

i let'wra a-t'wr-e-x ečeču
and candle.DAT VER-light-SM-PL there

'If small children from the household have died they set there little loaves of bread consecrated to them, and light a candle.'

I cannot say whether grammatical functions have an impact on the likelihood or frequency of being placed after the verb. The following two examples show an inanimate Goal (9) as well as an animate direct object, an animate indirect object plus animate adverbial (10).

(9) Laz (Holisky 1991: 409)

igzal-es bee-pe diška-ša
go-3PL child-PL firewood-all

'The children went for firewood.'

(10) Megrelian (Harris 1991: 374)

mapa-k kimeč tina mec'amale-s čil-o
king-ERG gave 3SG doctor-DAT wife-ADV

'The king gave her to the doctor as [his] wife.'

In Laz, indefinite postpredicate items are also a feature of some idiomatic expression such as 'set the table', 'make someone's wedding' and 'drink tea' (11).

(11) Laz (Lacroix 2009: 741)

hek do-v-es didi duğuni
there PV-make-AOR.I.3PL big wedding

'There they made big weddings.'

In sum, in Kartvelian SOV is a common and possibly the basic word order, but other orders are also possible and attested in texts. There are no hard constraints concerning the grammatical function or role of postverbal arguments or adjuncts or their parts of speech. Laz differs from all other Kartvelian languages in terms of actual frequency of postpredicate items in natural texts, which is likely due to a substantial impact of Turkish. Note for example that the Laz corpus in Stilo & Lacroix (2021) exhibits less than 5% post-verbal Goals (cf. the figure of 75%

from spoken Georgian (Tables 5 and 6 above), and comparable figures across the WOWA sample). It is possible that the texts in the Laz corpus of Stilo & Lacroix (2021) have been edited in some manner; this remains to be clarified.

3 Post-predicate items in East Caucasian

3.1 Word order profile of East Caucasian

Noun phrases are normally head-final (Ganenkov & Maisak 2021). However, various types of modifiers (except for demonstratives) can occur after the head noun and there is some indication that in many cases the postponed modifier does not form one NP with the preceding nominal, but rather makes up its own NP, e.g. because it needs to be case marked, nominalized or bear other types of special marking (e.g. Dargwa languages, Akhvakh). Testelets (1998a: 274) characterizes postposed modifiers as focused, contrasted, or restrictive. It seems that in natural texts genitives, in particular possessive pronouns, are postposed more commonly than any other type of modifier (see examples below).

East Caucasian languages have postpositions. Auxiliaries follow the lexical verbs. Major complementation strategies are non-finite verb forms (infinitive, masdar, participles, converbs), quotative particles, which are usually placed to the right of the clause, or enclitics and zero marking. Complementizers, which are often loans, play only a marginal role. Complement clauses may precede or follow the matrix verb.

As the other two indigenous families of the Caucasus, East Caucasian languages are predominantly head-final (SOV), but allow for all logically possible orders. Thus, we find postverbal arguments and adjuncts of all kinds in the literature (Testelets 1998a, van den Berg 2005) and in natural texts they are common. Word order in subordinate clauses is more restricted. For instance, in Sanzhi Dargwa relative clauses are verb-final with very few exceptions; complement clauses and adverbial clauses show a stronger tendency for verb-final order than main clauses, but far less than relative clauses (Forker 2020). A similar distribution is found in Hinuq: relative clauses are strictly head-final whereas complement and adverbial clauses occasionally contain post-predicate elements (Forker 2013).

3.2 Post-predicate elements

Based on the literature and on counts from the Multicast corpora for Chirag Dargwa, Sanzhi Dargwa (Forker & Schiborr 2019) and Tabasaran (Bogomolova

et al. 2021) a few generalizations concerning the conditions for post-predicate elements are possible. In general, they are far more frequent than in Adyghe (see Haig et al. 2024 [this volume], Table 9). Grammatical functions play a role in all three languages. Chirag Dargwa and Sanzhi Dargwa have far more post-predicate subjects than any other kinds of elements whereas for Tabasaran the difference between subjects and objects is relatively small. Goals, which include addressees, are more often found in postverbal position than obliques (= indirect objects, beneficiaries, instruments, sources, etc.) and locations (Table 9). For instance, in Sanzhi 57.9% of the 57 goals in main clauses occur after the verb (33 items).

Table 9: Post-predicate elements in Chirag, Sanzhi and Tabasaran

	Chirag Dargwa		Sanzhi Dargwa		Tabasaran	
texts	11		8		5	
words	5347		3857		5450	
main clauses	1183		945		1210	
all clauses	1377		1066		1383	
post-predicate elements in main clauses (nouns, pronouns, other items)						
subject	65	(23.81%)	62	(26.05%)	141	(28.54%)
object	31	(28.44%	38	(41.76%)	77	(30.68%)
Goal + addressee	20	(32.79%)	33	(57.89%)	64	(54.7%)
oblique	21	(34.42%)	18	(35.29%)	21	(31.34%)
location	12	(26.53%)	18	(35.29%)	28	(45.9%)
total per clause (all roles)	150	(12.68%)	169	(17.88%)	331	(27.36%)

Chechen and Ingush can be added to the East Caucasian languages for which we know that postverbal items are common in natural texts. For Ingush, Nichols (2011: 678) states that in "main clauses, other than episode-initial and other all-new ones, verb-second order is most common." According to Nichols (1994), in Chechen OVS is not uncommon in elicited sentences (see also Komen 2007: 32 for a similar assessment).

When comparing preverbal objects (OV) to postverbal objects (VO) in Chirag, Sanzhi, Tabasaran and Hinuq in main (Table 10) it turns out that almost between 30 and 40% of the objects occurs after the predicate. This is less than in Georgian (Tables 2 and 3), but still much more than in the Iranian and Turkic verb-final languages in the WOWA sample, which generally exhibit >80% OV order (see

Table 10: Postverbal objects (VO) in Chirag, Sanzhi, and Tabasaran in main clauses

	Chirag Dargwa	Sanzhi Dargwa	Tabasaran
O (all Os)[a]	109	91	251
VO	31	38	77
%	28.44%	41.76%	30.68%
NP	79	71	204
VO for NPs	19	29	68
%	**24.05%**	**40.85%**	**33.33%**
pro	20	13	22
VO for pro	9	9	8
%	**45.00%**	**69.23%**	**36.36%**

[a]Note that all Os consist of lexical NPs, pronouns and other items. In the table, only lexical NPs and pronouns are listed separately.

Haig et al. 2024 [this volume]). This is suggestive of a distinct kind of OV for these languages (and Georgian as well). Furthermore, in Sanzhi Dargwa and Chirag Dargwa, and to a small degree also in Tabasaran, pronominal objects have a greater tendency to be placed after the verb than nominal objects; for obliques, goals and locations no such tendencies can be observed (Table 10).

I was not able to systematically check for heaviness and the position of headed relative clauses. Instead, I will examine the literature on information structure and the placement of arguments and adjuncts in post-verbal position. Ganenkov & Maisak (2021: 129) state "The postverbal field is reserved for background information—that is, those arguments that are recoverable from the context but still mentioned for the sake of clarity." This generalization can be made for Hinuq, Dargwa languages, Archi, Avar, Lak, Ingush and probably more East Caucasian languages (Forker & Belyaev 2016; Komen & Bugenhagen 2017, Testelets 1998a: 260–261). In particular when the verb is focused topical arguments can follow it (12).

(12) [Then the wife of a student hears about the news.]
  Lak (Khalilova 1976: 204–205)
  [mu=gu]^{TOP} maħattal x̂-unu d-ur [wa iš-ira-j]^{TOP}
  DEM.PROX=ADD amazed become-PST.GER II-COP this issue-OBL-SPR
  'She also got amazed because of this issue.'

Verb fronting is a typical way of marking predicate focus and leads to post-predicate elements that are either topical or can also be focal. In (13), the verb is located in the clause-initial position while the argument NPs retain their unmarked SO order.

(13) [Husband and wife fought and a scandal happened and]
      Sanzhi Dargwa (Forker 2020: 523)
      [b-aˤq-ib   ca-b]^FOC sub-li   x:unul-li-j
      N-hit.PFV-PRET COP-N husband-ERG woman-OBL-DAT
      'The husband hit the wife.'

Focus, in particular wide focus and contrastive focus can also occur after the verb. In some languages, it is especially common with goals, including spatial goals, addressees, recipients, etc. It is possible to have simultaneously pre- and postverbal wide focus. In (14) from Budukh, we have contrasted focal elements in clause-initial position as well as in clause-final position.

(14) Budukh (Talibov 2007: 273)
      [q'aʒir-a]^FOC suˤre-rber č-aʁ-ar   [q:iˤš:laχ-ʒ-e]^FOC, [jaz-ʒ-e]^FOC
      winter-LOC   herd-PL SUB-go-MSD qishlaq-OBL-LOC autumn-OBL-LOC
      ʕoˤšχ-ar-i   [daʁ-ʒ-a]^FOC
      return-MSD-PRS mountain-OBL-LOC
      'In winter the herds go to the qishlaqs, in autumn they return to the mountains.'

In Hinuq (15), postverbal topics tend to precede postverbal foci (i.e. V-TOP-FOC) rather than the other way around (V-FOC-TOP), which is also the usual order for preverbal topics and foci and corresponds to what has been observed for many languages: known information precedes new information. However, apart from those sentences in which one of the NPs is a Goal, two NPs following the verb are not frequently found.

(15) Hinuq (Forker 2013: 759)
      Ø-ežinnu uži-ž   r-aš-a   goł [hayłoz]^TOP [nasibaw   žo]^FOC
      I-old   son-DAT v-find-INF be he.DAT   predestined thing(V)
      'The oldest son will find the thing predestined for him.'

A special context that leads to the occurrence of post-predicate elements are floating modifiers of nouns that are separated from the head noun by other constituents. The head nouns (possibly in combination with other modifiers) are often focal and occur in preverbal position while the floating modifier is displaced

postverbally. Especially common are floating genitives in the form of topical personal pronouns and demonstratives as in example (16) (see also Forker 2020: 410, 512–518, Creissels 2013, Komen & Bugenhagen 2017 for more examples). Creissels (2013) analyzes such constructions in Akhvakh. In contrast to genitives occurring in their canonical prenominal position floating genitives agree with the head noun in gender and fulfill "a possessive framing function, in the sense that the floating genitive identifies the personal sphere of its referent as the frame within which the predication expressed by the clause holds" (Creissels 2013: 333).

(16) Icari Dargwa (Sumbatova & Mutalov 2003: 160)
      č'ug   q:at:a-d   ħaˤjwan-ti d-ir-iri   niš:a-la
      down canyon-N.PL.INESS cattle-PL   N.PL-become-HAB.PST 1PL-GEN
      'Down in the canyon there was our cattle.'

When comparing postverbal subjects with postverbal objects, it seems that the former are more influenced by information structure than the latter. Komen & Bugenhagen (2017), based on counts in a corpus of Chechen, found that one third of Chechen subjects occur after the finite verb in main clauses, and of those post-verbal subjects one third are pronominal. Postverbal subjects occur in utterances with presentational focus to introduce new referents by means of NPs, in existential clauses, to express paragraph-internal cohesion, i.e. with topical and pronominal subjects, and in reported speech constructions. This is a phenomenon also found in other East Caucasian languages as well as in Northwest Caucasian (Forker 2024 [this volume]) and Kartvelian (Section 2). Intransitive thetic sentences show very clear word order preferences. Presentational sentences that introduce new referents (usually human, but sometimes also non-human, e.g. in fairy tales) frequently place the new referent in post-predicate position (17).

(17) Ingush (Komen & Bugenhagen 2017)
      Qoalagh='a [qeachaav cwalxa cwa bearii]^{FOC}.
      third=and   arrived   alone   one horseman
      'A third lone rider arrived.'

Direct speech constructions where the verb of speech follows the quote often have postverbal subjects (18), and this type of construction is also common in the Northwest Caucasian language Adyghe (Haig et al. 2024 [this volume]).

(18) Chechen (Komen & Bugenhagen 2017)
"t'aaqqa ishkoliehw diesha   a   aatta xir du," oolura   txan
then   school.LOC learn.INF add easy will be   say.IMPF 1PL.GEN
neenavashas
uncle.ERG
' "Then he would learn more easily at school" our uncle said.'

A last factor influencing the likelihood at least for postverbal objects is language contact. Table 11 summarizes counts in four different text collections in Hinuq (Forker & Belyaev 2016, Forker 2019). There is almost no difference concerning the position of the direct object between the older published texts and my own texts recorded 60 years later (17%). The pear stories collected with the same stimulus as the Georgian texts discussed in Section 2.3 have more postverbal objects (25%). The frog stories produced by speakers under 30 years living in the ethnolinguistically mixed village Monastirski and Shamkhal in the lowlands show an even larger amount of postverbal objects (43%). This can possibly be attributed to the greater influence of Russian and ongoing language shift among young speakers in the lowlands and resembles what has been said about Laz in Turkey when compared to the other Kartvelian languages in Georgia (Section 2.3).

Table 11: O-V vs. V-O in Hinuq texts (Forker & Belyaev 2016, Forker 2019)

# words	age of speakers	place of recording	year	OV	VO	total
old published texts						
1,507	14–29	Chechnya	1950	139 (82.74%)	29 (17.26%)	168
new traditional texts						
2,503	12–62	Hinuq	2006–2009	137 (82.53%)	29 (17.47%)	166
pear stories						
1,583	13–30	Hinuq	2006–2007	125 (74.40%)	43 (25.60%)	168
frog stories						
2,033	19–29	Monastirski, Shamkhal	2013	72 (57.14%)	**54 (42.86%)**	126

Russian is usually assumed to have free word order, but with an underlying SVO structure (Tomlin 1986, though see the debate in Theoretical Linguistics 48(1–2) 2022, in particular Haider & Szucsich (2022)). Corpus studies come to different results, but it seems that V-O is more frequent than O-V. For instance, Bazhukov et al. (2021) count the order of DO, IO and V for ditransitive verbs in the SynTagRus corpus and get 1420 O-V clauses vs. 4978 V-O clauses. Billings (2015) analyzed 500 clauses in the Russian National Corpus (RNC). The most numerous patterns were SVO (448) and SOV (22 clauses). However, Levshina et al. (2023: 856–857) compared 100 sentences of spoken Russia to 100 sentences of written Russian (Fiction and News) and found remarkable differences between the modalities: the conversations contained 61 examples of OV, and only 39 examples of VO, whereas both the fiction and news contained 17 examples of OV and 83 examples of VO each. Hinuq speakers are exposed to written standard Russian through the educational system, through the media, etc., but also to other forms of Russian such as oral (colloquial and standard) Russian through the media and non-standard Russian as spoken in the Caucasus. Russian impact on word order patterns in a similar vein as it is possibly found in Hinuq (SOV > SVO) has been document for Sakha (Turkic) (Grenoble et al. 2019) and Udmurt (Uralic) (Asztalos 2021).

In sum, postverbal items in East Caucasian occur relatively frequently. They fulfil various grammatical functions. In particular goals are prone to occur after the verb. There are also indications that in some languages (Sanzhi Dargwa, Chirag Dargwa, Chechen) part of speech plays a role in the sense that nominals and pronouns do not behave alike when it comes to their position with respect to the verb; see Forker 2024 [this volume] for additional parallels. There are two constructions in which postverbal items are frequently found in many East Caucasian languages, namely thetic introductory sentences and reported speech constructions, and similar constructions have been identified in Kartvelian and Adyghe. Information structure affects word order at the clausal level, but it is not possible to identify any strict rules. This means that topical as well as focal items can appear after the verb. Furthermore, East Caucasian languages have a special construction in which modifiers split from their head and appear in a postverbal position. Finally, data for Hinuq suggest that Russian has an impact on the frequency of postpredicate items, in particular with younger speakers living in ethnically mixed places in the Dagestanian lowlands.

4 Discussion

All indigenous Caucasian language families (Kartvelian, East Caucasian, but also Northwest Caucasian) are more rigid with respect to word order in noun phrases and subordinate clauses and declarative main clauses enjoy the most flexibility.

Post-predicate items in Georgian (Kartvelian) and East Caucasian are relatively common (when compared to Northwest Caucasian) and can be triggered by

- certain constructions such as thetic utterances and general information structure

- certain semantic roles (e.g. in Georgian, goals, and in Sanzhi Dargwa and Tabasaran, goals and addressees) show a greater preference than other semantic roles

- heaviness and the presence of relative clauses in Georgian (no data for East Caucasian available)

As the data from Laz (Kartvelian) show, language contact has a strong impact on the flexibility of constituent order at the clausal level and on the presence vs. absence of postpredicate elements. Laz is the only one of the Kartvelian languages mainly spoken in Turkey and resembles Turkic with respect to word order patterns.

When comparing the three indigenous language families of the Caucasus (and excluding Laz), it turns out that Northwest Caucasian languages are the least flexible languages. One is tempted to hypothesize that this is due to their head-marking profile. They have little to no case marking but richer verbal indexing than the other two families. Studies have found a robust negative correlation between rigid word order and case marking (Sinnemäki 2014, Levshina 2021).

The label "flexible SOV" for Kartvelian and East Caucasian is very coarse-grained and corpus data from different sources show a spectrum of different word order patterns and varying degrees of frequencies. Levshina et al. (2023) show that word order patterns are subject to influence by many factors, some of them competing with each other, so that word order flexibility is a common outcome. In their study they mention one East Caucasian language, Avar, as an "SOV flexible" languages with a higher degree of flexibility than other languages in the same study (Malayalam, Hindi, Spanish, Korean, and English.), which fits to the data from Chirag, Sanzhi and Tabasaran in this paper.

Abbreviations

1	first person	IO	indirect object
3	third person	LOC	locative
A	agent	MSD	masdar
ABL	ablative	N	neuter
ADD	additive	NOM	nominative
ADV	adverbial	OBL	oblique
AOR	aorist	PFV	perfective
APPL	applicative	PL	plural
cop	copula	PRET	preterite
DAT	dative	PROX	proximate
DEM	demonstrative	PRS	present
DIM	diminutive	PST	past
DIST	distal	PTCP	participle
DO	Direct object	PV	preverb
ERG	ergative	REFL	reflexive
EV	euphonic vowel	S	subject (single argument of an intransitive verb)
GEN	genitive		
GER	gerund	SG	singular
I	masculine gender	SM	series marker
II	feminine gender	SPR	superessive
IMP	imperative	SUB	subjunctive
IMPF	imperfective	SUPINE	supine
INESS	inessive	TM	thematic marker (present stem formant)
INF	infinitive		
INST	instrumental	V	gender V
INTR	intransitive	VER	version

References

Apridonidze, Šukia. 1986. *Sit'q'vatganlageba axal kartulši [Word order in modern Georgian]*. Tbilisi: Mecniereba.

Aronson, Howard. 1991. Modern Georgian. In Alice C. Harris (ed.), *The indigenous languages of the Caucasus. vol. I: The Kartvelian languages*, 219–312. Delmar, NY: Caravan Books.

Asatiani, Rusudan & Stavros Skopeteas. 2012. The information structure of Georgian. In Manfred Krifka & Renate Musan (eds.), *The expression of information structure*, 127–158. Berlin: De Gruyter Mouton.

Asztalos, Erika. 2021. From head-final towards head-initial grammar: Generational and areal differences concerning word order usage and judgement among Udmurt speakers. In Diana Forker & Lenore A. Grenoble (eds.), *Language contact in the territory of the former Soviet Union*, 143–182. Amsterdam: Benjamins. DOI: 10.1075/impact.50.06asz.

Bazhukov, M. O., L. I. Chubarova, N. A. Slioussar & S. Yu. Toldova. 2021. The order of objects in Russian: A corpus study. *Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialogue"* 20. 68–78. DOI: 10.28995/2075-7182-2021-20-68-78.

Billings, Stephanie Kay. 2015. *A corpus-based analysis of Russian word order patterns.* Brigham Young University. (MA thesis). https://scholarsarchive.byu.edu/etd/5624.

Boeder, Winfried. 2005. The South Caucasian languages. *Lingua* 115(1-2). 5–89.

Boeder, Winfried. 2021. Modern Georgian.

Bogomolova, Natalia, Dmitry Ganenkov & Nils N. Schiborr. 2021. Multi-CAST Tabasaran. In Geoffrey Haig & Stefan Schnell (eds.), *Multi-CAST: Multilingual corpus of annotated spoken texts.* Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/#tabasaran (27 July, 2023).

Chafe, Wallace. 1980. *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production.* Norwood, NJ: Ablex.

Creissels, Denis. 2013. Floating genitives and possessive framing in Northern Akhvakh. In Anne Carlier & Jean-Christophe Verstraete (eds.), *The genitive*, 333–354. Amsterdam: Benjamins.

Forker, Diana. 2013. *A grammar of Hinuq.* Berlin: De Gruyter Mouton.

Forker, Diana. 2019. The impact of language contact on Hinuq: Phonology, morphology, syntax, and lexicon. *Language Typology and Universals* 71. 29–62.

Forker, Diana. 2020. *A grammar of Sanzhi Dargwa.* Berlin: Language Science Press.

Forker, Diana. 2024. Post-predicate elements in Adyghe. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 309–336. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266351.

Forker, Diana & Oleg Belyaev. 2016. Word order and focus particles in Nakh-Daghestanian languages. In M. M. Jocelyne Fernandez-Vest & Robert D. Van Valin (eds.), *Information structuring of spoken language from a cross-linguistic perspective*, 239–262. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110368758-012.

Forker, Diana & Nils N. Schiborr. 2019. Multi-CAST Sanzhi Dargwa. In Geoffrey Haig & Stefan Schnell (eds.), *Multi-CAST: Multilingual corpus of annotated spoken texts*. Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/#sanzhi.

Ganenkov, Dmitry & Timur A. Maisak. 2021. Nakh-Dagestanian languages. In Maria Polinsky (ed.), *The Oxford handbook of languages of the Caucasus*, 86–145. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780190690694.013.4.

Grenoble, Lenore A., Jessica Kantarovich, Irena Khokhlova & Liudmila Zamorshchikova. 2019. Evidence of syntactic convergence among Russian-Sakha bilinguals. *Suvremena Lingvistika* 45(7). 41–57.

Haider, Hubert & Luka Szucsich. 2022. Slavic languages – "SVO" languages without SVO qualities? *Theoretical Linguistics* 48(1-2). 1–39. DOI: 10.1515/tl-2022-2035.

Haig, Geoffrey, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils Schiborr. 2024. Post-predicate elements in the Western Asian Transition Zone: Data, theory, and methods. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 3–54. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266331.

Harris, Alice C. 1991. Mingrelian. In Alice C. Harris (ed.), *The indigenous languages of the Caucasus. vol. I: The Kartvelian languages*, 314–394. Delmar, NY: Caravan Books.

Harris, Alice C. 2000. Word order harmonies and word order change in Georgian. In Rosanna Sornicola, Erich Poppe & A. Sisha-Halevy (eds.), *Stability, variation and change of word order patterns over time*, 133–163. Amsterdam: Benjamins.

Holisky, Dee A. 1991. Laz. In Alice C. Harris (ed.), *The indigenous languages of the caucasus. vol. I: The Kartvelian languages*, 395–472. Delmar, NY: Caravan Books.

Khalilova, Kh. M. 1976. Lakskie teksty [Lak texts]. In A. A. Axlakov & Kh. M. Xalilov (eds.), *Satira i jumor narodov Dagestana [Satire and humor of the Daghestanian people]*, 201–258. Makhachkala: Daginogoizdat.

Komen, Erwin R. 2007. *Focus in Chechen.* Leiden University. (MA thesis).

Komen, Erwin R. & Robert D. Bugenhagen. 2017. Post-verbal pronominal subjects in Chechen and Ingush. Paper presented at the *50th annual meeting of the societas linguistica europaea,* 10 September - 13 September, Zürich, Switzerland.

Kutscher, Silvia & Nuran Sevim Genç. 1998. *Ardeşen narrates - Ardeşeni na isinapinenpe. A collection of Laz spoken texts with glosses and translations into English, German, Turkish.* Munich: Lincom Europa.

Lacroix, René. 2009. *Description du dialecte Laze d'Arhavi (Caucasique du Sud, Turquie)*. Lyon: Université Lumière Lyon 2. (Doctoral dissertation).

Levshina, Natalia. 2021. Cross-linguistic trade-offs and causal relationships between cues to grammatical subject and object, and the problem of efficiency-related explanations. *Frontiers in Psychology* 12. 648200. DOI: 10.3389/fpsyg.2021.648200.

Levshina, Natalia, Savithry Namboodiripad, Marc Allassonnière-Tang, Mathew Kramer, Luigi Talamo, Annemarie Verkerk, Sasha Wilmoth, Gabriela Garrido Rodriguez, Timothy Michael Gupton, Evan Kidd, Zoey Liu, Chiara Naccarato, Rachel Nordlinger, Anastasia Panova & Natalia Stoynova. 2023. Why we need a gradient approach to word order. *Linguistics* 61(4). 825–883. DOI: 10.1515/ling-2021-0098.

Nichols, Johanna. 1994. Chechen. In Rieks Smeets (ed.), *The indigenous languages of the Caucasus. vol. 4: The North East Caucasian languages II*, 1–77. Delmar, NY: Caravan Books.

Nichols, Johanna. 2011. *Ingush grammar*. Berkeley: University of California Press.

Pourtskhvanidze, Zakharia. 2015. *Fokuspartikeln und Wortstellung im Georgischen*. Wiesbaden: Reichert.

Rasekh-Mahand, Mohammad, Elham Izadi, Mehdi Parizadeh, Geoffrey Haig & Nils Schiborr. 2024. Post-predicate elements in modern colloquial Persian: A multifactorial analysis. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 197–230. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266343.

Rostovtsev-Popiel, Alexander. 2021. Megrelian. In Maria Polinsky (ed.), *The Oxford handbook of languages of the Caucasus*, 529–569. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780190690694.013.8.

Schmidt, Karl H. 1991. Svan. In Alice C. Harris (ed.), *The indigenous languages of the Caucasus. vol. 1: The Kartvelian languages*, 471–556. Delmar, NY: Caravan Books.

Sinnemäki, Kaius. 2014. Complexity trade-offs: A case study. In Frederick J. Newmeyer & Laurel B. Preston (eds.), *Measuring grammatical complexity*, 179–201. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199685301.003.0009.

Skopeteas, Stavros. 2021. Georgian: V-final or V-medial. In Christiane Bulut, Anaïd Donabédian-Demopoulos, Geoffrey Haig, Geoffrey Khan, Pollet Samvelian, Stavros Skopeteas & Nina Sumbatova (eds.), *Glottothèque: Languages of the Anatolia-Caucasus-Iran-Mesopotamia online*. Bam-

berg/Cambridge/Göttingen/Moskow/Nicosia/Paris: LACIM network. https://spw.uni-goettingen.de/projects/lacim/lng-geo.html.

Skopeteas, Stavros, Caroline Féry & Rusudan Asatiani. 2009. Word order and intonation in Georgian. *Lingua* 119(1). 102–127. DOI: 10.1016/j.lingua.2008.09.001.

Stilo, Donald. 2014. Areal factors and postverbal patients in spoken Azerbaijani: A corpus-based study. In Nurettin Demir, 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. Paper presented at the *Conference Anatolia-Caucasus-Iran: Ethnic and Linguistic Contacts* 10-12 May 2018. Yerevan University, Yerevan, Armenia.

Stilo, Donald & René Lacroix. 2021. Laz (Arhavi). In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation*. Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/ (27 July, 2023).

Sumbatova, Nina R. & Rasul O. Mutalov. 2003. *A grammar of Icari Dargwa*. München: Lincom Europa.

Talibov, Bukar B. 2007. *Buduxskij jazyk*. Moscow: Academia.

Testelets, Yakov G. 1998a. Word order in Daghestanian languages. In Anna Siewierska (ed.), *Constituent order in the languages of Europe*, 257–280. Berlin: De Gruyter Mouton.

Testelets, Yakov G. 1998b. Word order in Kartvelian languages. In Anna Siewierska (ed.), *Constituent order in the languages of Europe*, 235–256. Berlin: De Gruyter Mouton.

Testelets, Yakov G. 2021. Kartvelian (South Caucasian) languages). In Maria Polinsky (ed.), *The Oxford handbook of languages of the Caucasus*, 491–528. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780190690694.013.5.

Tomlin, Russell. 1986. *Basic word order: Functional principles*. London: Croom Helm.

Tuite, Kevin. 1998. *A short descriptive grammar of the Svan language*. Montréal: Université de Montréal. https://archive.org/details/060-svan.

van den Berg, Helma. 2005. The East Caucasian language family. *Lingua* 115(1-2). 147–190.

Vogt, Hans. 1971. *Grammaire de la Langue Géorgienne*. Oslo: Universitaetsvorlaget Oslo.

Chapter 11

Post-predicate elements in Adyghe

Diana Forker

Friedrich Schiller University Jena

In this chapter, I study post-predicate elements in the Northwest Caucasian language Adyghe. In the literature, Adyghe is characterized as having SOV as its basic pattern, but as being in principle a "free" word order language. There are no corpus-based studies on word order in Adyghe (or any other Northwest Caucasian language) up to now, so this study is a step towards filling this research gap.

I first examine examples of post-predicate elements in the literature on Adyghe, which confirm the expectations and exemplify various types of arguments and adjuncts as well as subordinate clauses that can appear after the verb. In a second step, I identify and count post-predicate elements in 20 Adyghe texts collected between 1969 and 2017 among the Adyghe people in the Caucasian homeland and the Turkish diaspora by various researchers. Only around 10% of the main clauses contain post-predicate elements of which the majority are pragmatic particles, but post-predicate subject, objects and adjuncts are also attested. Most post-predicate referents are topical, but focal referents can also be found. Differences in genre play a relatively big role: personal accounts contain around as twice as many instances of post-predicate elements than traditional narratives. Furthermore, the texts from Adygea show a greater frequency of post-predicate elements than those from Turkey which might be due to the influence of two typologically and genealogically different contact languages (Russian and Turkish).

1 Introduction

In this contribution, I outline the main facts of word order in Adyghe and then analyze in more detail post-predicate elements.

Adyghe belongs to the Circassian branch of the Northwest Caucasian language family. The homeland of the Adyghe people is the northwestern Caucasus

region (Russian Federation). Since the conquest of the Caucasus by Russia in the second half of the 19th century there are diaspora communities in Turkey, Jordan, Israel, Syria and other countries.

The main goal of my contribution is to identify first of all grammatically acceptable post-predicate elements and examine them with respect to their syntactic functions, grammatical roles and their information structural properties. My second goal is to study post-predicate elements in natural texts by comparing data from the Adyghe homeland with data from the diaspora community in Turkey. Adyghe diaspora communities in Turkey (and other places) have been relatively isolated from the original speech community for more than 150 years. The two communities are under the influence of two distinct languages of wider communication, Turkish and Russian respectively, and the data suggest that these differing contact scenarios have led to divergence.

Like all Northwest Caucasian languages, Adyghe is polysynthetic with highly complex verbal morphology. Verb forms contain pronominal prefixes indexing all syntactic arguments of the predicate, i.e. intransitive and transitive subjects, direct object, indirect object, etc. Parts of speech are not always clearly differentiated in terms of inflection. A wide range of grammatical markers for person, tense, number, modality, and negation can be added to any content word. Adyghe has ergative alignment that shows up in case marking and agreement. The suffix *-r* ("absolutive") marks intransitive subjects and direct objects. The suffix *-m* ("oblique") marks transitive subjects (agents), as well as indirect objects, certain adverbials (temporal, spatial), and adnominal possessors. Nonspecificity and indefiniteness are indicated by the omission of case suffixes. Proper nouns and first and second pronouns do not distinguish absolutive and oblique. Clauses contain at least a predicate, which can be verbs, nouns, pronouns, adjectives or even postpositions. Predicates take pronominal prefixes and tense morphology. Copula clauses consist of a copula complement and the copula verb. Overt argument NPs are optional as it is expected for polysynthetic languages (Testelets & Lander 2017). This property makes the study of word order patterns at the clausal level based on natural texts somewhat difficult because arguments are recurrently only expressed through pronominal prefixes.

Major grammatical descriptions of Adyghe are Jakovlev & Ashkhamaf (1941), Rogava & Kerasheva (1966) and Arkadiev et al. (2009). Kumakhov & Vamling (2009) analyze Circassian clause structure, including word order. There are no corpus-based studies on word order in Adyghe. But there is a corpus of Standard Adyghe (Arkhangelskiy et al. 2018–2023), available on the Internet (http://adyghe.web-corpora.net/).

This paper is mainly based on 20 Adyghe texts (6,146 words) recorded in Adygea (Caucasus) and Turkey.[1] All examples are marked by [H] for 'homeland' and [D] for 'diaspora.' Some of these texts have been published in Höhlig (1997), Paris (1974) and Feer (2019), and the texts from Paris (1974) can also be found in the online Pangloss Collection (see references). Other texts collected by Monika Höhlig in the 1990s and by Feer between 2016 and 2017 were kindly provided to me by both researchers. The texts are monologues that can be roughly divided into two types, namely (i) traditional narratives such as legends, fairy tales and anecdotes and (ii) personal accounts/autobiographies. Following Höhlig (1997), I also classified the texts according to the age of the speakers into old, middle and young generation. The texts have been chosen such as to roughly equally represent Adyghe from the homeland in contact with Russian and Adyghe from the diaspora in contact with Turkish. A further criterion was genre.

A full list with sources can be found in the appendix.

2 Word order profile of Adyghe

2.1 Word order patterns in noun phrases and other constituents

In this section and the following section, I present an overview of word order patterns in Adyghe noun phrases and clauses including a few examples with post-predicate elements. A more detailed discussion of post-predicate elements in texts will be given in Section 3.

Constituent order within the NP is mixed. Adjectives, simple cardinal numerals except for the numeral 'one' and resultative verbs follow the noun. The numerals are suffixed by means of a linking morpheme as in the following example (1).

(1)   noun + adjective-numeral + adjective
      Adyghe (courtesy of Y. Lander) [H]
      ha   ʁʷež'-jə-ṭʷ   gʷere
      dog yellow-LNK-two certain
      'two certain yellow dogs'

Demonstratives (2), non-referential modifying nouns and appositive names (3), relative clauses (4), possessors including possessive prefixes (2), (3), the cardinal numeral 'one' (4) and ordinal numerals precede the nominal head. Lander (2017)

---

[1]There is no WOWA-data set for Adyghe, or any other Northwest Caucasian language to date.

labels the prenominal modifiers "non-adjectival" and notes that they are ungradable. Postnominal modifiers can but need not be gradable. Very commonly modifiers enter into a close connection with the modified noun and are pronounced and written together such as the noun and the adjective in the second noun phrase 'his grandmother' in example (2). These units are called "nominal complex" in Lander (2017), who argues that they form a single word, based on their morphosyntactic properties. For instance, case marking and plural marking occur only once per unit (8).

(2)   demonstrative + noun + adjective and possessive-noun-adjective, postposition
      Adyghe (courtesy of M. Höhlig) [H]
      [mwe č'ele çəč'ə-r]   *[ja-ne-ẑ]*   *djə qə-zə-ḳʷe-č'e*   ...
      that   boy   little-ABS POSS-mother-old to   DIR-REL.TEMP-go-INST
      'when that little boy went to his grandmother ...'

(3)   possessive construction, postpositional phrase
      Adyghe (Paris 1974) [D]
      [[qahraman gʷaše-m]   *jə-šə-šxa-pʷe]*   *jə-dež'-g'e*
      Kahraman   princess-OBL POSS-horse-eat-place POSS-to-INST
      possessor   possessed   postposition
      'to the manger of the horses of princess Kahraman'

(4)   relative clause + numeral-noun
      Adyghe (Paris 1974) [D]
      dečʼːʼəɣəməqːʷe   *pšəpəjə-r*   *[...] [xeʁegʷə-m   jə-sə-xe-me*
      Detcheghemeqo Pshepeye-ABS   country-OBL LOC-live-PL-OBL.PL
      ʔape-g'e   *q:-ja-ʁa-λaʁʷe-w]*   *zə-çəfə-ʁ*
      finger-INST DIR-3PL.A-CAUS-see-ADV one-human.being-PST
      'Detcheghemeqo Pshepeye [...] was a person whom the inhabitants (lit. 'the ones living there') of the country respected.' (lit. 'one human being that they pointed at with their fingers')

Adyghe has exclusively postpositions, which have mostly been grammaticalized from nouns (Arkadiev & Maisak 2018). The postpositional complement is often additionally expressed via a possessive prefix (3).

In complex verb forms, auxiliaries follow the lexical verb and some have already grammaticalized into suffixes (Kimmelman 2011, Arkadiev & Maisak 2018). They mostly express aspectual meanings and epistemic modality (probability and

necessity) and can also occur in conditional clauses. In the following example, the auxiliary verb *χʷə-* 'be, happen' bears the conditional suffix *-me* and forms a complex verb together with the lexical verb *ʔʷe-* 'say, speak, tell, talk'.

(5) Adyghe (courtesy of M. Höhlig) [H]
gʷəxeλ-ew jə-ʔe-r   qə-mə-ʔʷa-xe   χʷə-me,   …
intent-ADV LOC-be-ABS DIR-NEG-speak-TRM happen-COND
'if (the guest) did not tell the intentions that he had …'

2.2 Word order at the clausal level

At the level of the main clause, Adyghe is, as other Northwest languages, at the same time left-branching / verb-final, with SOV being considered as a kind of default pattern, though tolerating a fair degree of flexibility (Jakovlev & Ashkhamaf 1941, Rogava & Kerasheva 1966, Kumakhov & Vamling 2009, Lander 2014, Testelets & Lander 2017). All logically possible orders are available. As typical for SOV languages, focal items occur in the preverbal position and contrastive items are said to occur sentence-initially (Arkadiev & Lander 2021).

Jakovlev & Ashkhamaf (1941: 91) list the following patterns for transitive verbs with overt subject, direct object, and an adjunct noun expressing location (6a-6f, 6g). Verb-final order with the subject preceding the direct object (SOV) is analyzed as basic and neutral with respect to emphasis and information structure (6a). The reversal of subject and direct object (OSV) illustrated in (6b) is characterized as also possible but less used. All other patterns are called "inverse" (6c-6f, 6g). They write that in verb-second patterns the final element is most highlighted and the penultimate element that immediately follows the verb is also highlighted, albeit to a lesser extent.

(6) a. Adyghe (Jakovlev & Ashkhamaf 1941: 91) [H]
        basic pattern: S-O-LOC-V [SOV]
        č'ale-me   baǯ'e-xe-r   mezə-m   š'a-λeʁʷə-ʁe-x
        boy-OBL.PL fox-PL-ABS forest-OBL LOC-see-PST-PL
    b. O-S-LOC-V [OSV]
        baǯ'e-xe-r č'ale-me   mezə-m   š'a-λeʁʷə-ʁe-x
        fox-PL-ABS boy-OBL.PL forest-OBL LOC-see-PST-PL
    c. S-V-O-LOC [SVO]
        č'ale-me   š'a-λeʁʷə-ʁe-x   baǯ'e-xe-r   mezə-m
        boy-OBL.PL LOC-see-PST-PL fox-PL-ABS forest-OBL

    d. LOC-V-O-S [VOS]

       mezə-m   *š'a-λɐʁʷə-ʁe-x*   *baǯ'e-xe-r*   *č'ale-me*

       forest-OBL LOC-see-PST-PL fox-PL-ABS boy-OBL.PL

    e. O-V-LOC-S [OVS]

       baǯ'e-xe-r *š'a-λɐʁʷə-ʁe-x* *mezə-m*   *č'ale-me*

       fox-PL-ABS LOC-see-PST-PL forest-OBL boy-OBL.PL

    f. V-S-O-LOC[VSO]

       š'a-λɐʁʷə-ʁe-x *č'ale-me*   *baǯ'e-xe-r* *mezə-m*

       LOC-see-PST-PL boy-OBL.PL fox-PL-ABS forest-OBL

    g. V-S-LOC-O [VSO]

       š'a-λɐʁʷə-ʁe-x *č'ale-me*   *mezə-m*   *baǯ'e-xe-r*

       LOC-see-PST-PL boy-OBL.PL forest-OBL fox-PL-ABS

       'The boys saw the foxes in the forest.'

In verb-initial patterns, again the final element is emphasized and the verb to a lesser degree (6f, 6g).

Kumakhov & Vamling (2009: 117) also illustrate the six available constituent order patterns for subject, direct object and verb (albeit with a pronominal subject). Kumakhov & Vamling (2009: 112) further state that constituent order varies with information structure. For example, in answers to questions that target the subject SVO is more common than SOV.

The position of the indirect object in pragmatically neutral clauses with nominal arguments is between the subject and the direct object (S-DO-IO-V) according to Kumakhov & Vamling (2009: 114–115) (12–13). In this pattern, the two arguments that bear identical case markers (*-m*) (S, IO) are separated by the direct object in the absolutive case (*-r*).

(7) Adyghe (Kumakhov & Vamling 2009: 114) [H]
    S-DO-IO-V

    č'ale-m   *txeλə-r*   *pŝaŝe-m* *r-jə-tə-ʁ*

    boy-OBL book-ABS girl-OBL OBL-3SG.A-give-PST

    'The boy gave the book to the girl.'

(8) Adyghe (courtesy of M. Höhlig) [H]
    S-DO-IO-V

    jež' ja-te=ja-ne-xe-m-jə   *ə-š-xe-m-jə*

    self POSS-father=POSS-mother-PL-OBL-ADD 3SG.POSS-brother-PL-OBL-ADD

    pŝeŝeǯəje-r [*zə-fe-mə-je-w*   *məλkʷə* *z-jə-ʔe*]   *wənaʁʷe*

    girl-ABS   REL.IO-BEN-NEG-want-ADV property REL.IO-LOC-be family

gʷere-m   r-a-tə-ʁ
certain-ᴏʙʟ ᴏʙʟ-3ᴘʟ.A-give-ᴘsᴛ

'Her parents and brothers gave the girl to a certain family with property who she did not want.'

However, S-IO-DO-V is also attested (9), in particular when either the subject or the indirect object is a personal pronoun.

(9) Adyghe (Rogava & Kerasheva 1966: 65) [H]
S-IO-DO-V
hač'e-m   č'ale-m   šə-r   r-jə-tə-ʁ
guest-ᴏʙʟ boy-ᴏʙʟ horse-ᴀʙs ᴏʙʟ-3sɢ.A-give-ᴘsᴛ
'The guest gave the horse to the boy.'

The addressee in (10) exemplifies another indirect object in preverbal position. Adjuncts such as instruments (11) or beneficiaries expressed by means of postpositional phrases (12) usually occur between the subject and the verb. Yet positions before the subject (11) and after the predicate are also allowed (Section 3).

(10) Adyghe (courtesy of M. Höhlig) [H]
šə-m   apəλə-m   r-a-ʔʷa-ʁ   mə-rə   mə-rə
horse-ᴏʙʟ stableman?-ᴏʙʟ ᴏʙʟ-3ᴘʟ.A-speak-ᴘsᴛ this-ᴘʀᴇᴅ this-ᴘʀᴇᴅ
ə-ʔʷe-re-r
3sɢ.A-speak-ᴅʏɴ-ᴀʙs
'They said to the stableman: "It is like this and like that."'

(11) Adyghe (courtesy of M. Höhlig) [H]
əpč'e adəʁe-me   ʔa-č'e   jaṭe-č'e
earlier Adyghe-ᴏʙʟ.ᴘʟ hand-ɪɴsᴛ clay-ɪɴsᴛ
qə-r-a-jə-č'ə-š'tə-ʁe-x
ᴅɪʀ-ᴅᴀᴛ-3ᴘʟ.A-smear-ᴇʟ-ᴀᴜx-ᴘsᴛ-ᴘʟ
[Now houses are built with plaster.] 'Earlier the Adyghe people smeared (the houses) with clay with the hands.'

(12) Adyghe (courtesy of R. Feer) [H]
ҫəf-me   a-paje we p-ŝə-ʁe
human.being-ᴏʙʟ.ᴘʟ 3ᴘʟ-for 2sɢ 2sɢ.A-do-ᴘsᴛ
'You built it for the people.'

Temporal and spatial adverbials including locations and goals frequently occur in clause-initial positions before the subject if there is any overtly expressed subject (13, 14) or otherwise directly before the verb (14, 15), but also occasionally after the verb (Section 3).

(13) Adyghe (courtesy of M. Höhlig) [H]
TIME-LOC-S-COP-PRED
a   zeman-m hakʷənhable-m   č'əle-m thamate ja-ʔa-ʁ   hakʷərəne
that time-OBL Hakunhable-OBL aul-OBL elder   LOC-be-PST Hakuren
təʁʷəẑ a-ʔʷe-w
Teguz 3PL.A-speak-ADV
'At that time in Hakunhable, the village-elder was called Hakuren Teguz.'

(14) Adyghe (courtesy of R. Feer) [H]
TIME-S-LOC-COP-PURP[DO-V]
aj   ə-pe-re   mafe-xe-m χʷaʒ'e-r   beʒerə-m
that.OBL 3SG.POSS-earlier-ADJ day-PL-OBL hodja-ABS bazar-OBL
š'ə-ʔa-ʁ   [č'em qə-š'efə-n-ew]
LOC-be-PST cow   DIR-buy-MOD-ADV
'The days before, the Hodja was at the market to buy a cow.'

(15) Adyghe (courtesy of M. Höhlig) [H]
TIME-GOAL-CV
wtoroj   klass   nes škʷelə-m   sə-ḵʷa-ʁ
second[R] class[R] to   school[R]-OBL 1SG.ABS-go-PST
'Until the second class I went to school.'

Complement clauses are marked by a variety of strategies among which the most frequent ones are the bare verbal stem without any tense or other markers, the modal/potentialis form with or without additional case suffixes, a specialized factive form with the prefix *zere-*, case markers (adverbial case as in (4), (14), (16) and instrumental case as in (2)) and the conditional suffix in combination with the additive (17) (see Serdobolskaya 2016 for a detailed analysis). There are no complementizers. Complement clauses usually precede the matrix clause, (16), (17) but they can also follow or be embedded.

Reported speech can be marked with a quotative particle that has been grammaticalized from a non-finite form of the verb of speech *ʔʷe-* 'say, speak, tell, talk' (which, however, retains its person prefixes). The quotative particle follows the quote (10), (17). The clause expressing the quote follows the matrix clause with the verb of speech (10) or precedes it as in the next two examples (16, 17).

(16) Adyghe (Feer 2019) [H]
      [təʁʷəẑəqʷe qəzbečʼ s-λeʁʷə-n-ew]   *sə-feja-ʁ,*
      Tuguzhuko Kyzbech 1sg.A-see-mod-adv 1sg.abs-want-pst
      [sə-de-gʷəšʼəʔe-n-ew]   *sə-faj*   *ə-ʔʷa-ʁ*
      1sg.abs-com-talk-mod-adv 1sg.abs-must 3sg.A-speak-pst
      '"I would like to see Tuguzhuko Kyzbech, I want/need to speak with him,"
      he said.'

(17) Adyghe (courtesy of R. Feer) [H]
      [a-fede   šʼə-ʔe-m-jə]   *s-ŝa-xe-re-p*   *nəʔa*
      dat-similar loc-be-cond-add 1sg.io-know-trm-dyn-neg only
      ə-ʔʷa-ʁ   *je-ʔʷe*   *pŝaŝe-m*
      3sg.A-speak-pst dat-speak girl-obl
      'The girl said, "I don't know at all if there is somebody similar."'

Adverbial subordination is expressed through specialized and general con-
verbs, relativization, and the additive suffix, but not by means of subordinating
particles (Forker In Press). Adverbial clauses, in particular chaining clauses, pre-
cede the main clause, but a position after the main clause is also possible (Section
3).

Here is a summary of the preferred ordering patterns:

- fixed word order, but not consistent across different NP types

- only postpositions

- auxiliaries after lexical verb

- flexible word order at clausal level with preference for verb-final patterns,
  in particular for SOV

- complement and adverbial clauses precede main clauses

- relative clauses precede head nouns

From this follows that Adyghe has a certain preference for head-final patterns
within the clause and in clause combining.

3 Examining post-predicate elements

In this section, I will explore post-predicate elements in more detail and not discuss the relative positions of pre-predicate items with respect to each other. In particular, I will explore the relative frequency of post-predicate elements in homeland Adyghe vs. diaspora Adyghe and the impact of genre/style.

3.1 Post-predicate elements in elicitation and texts

As explained in Section 2.2, Adyghe has a tendency for head-final word order at the clausal level. But at the same time the word order is described as "free" and in the published works on Adyghe, one finds many instances of post-predicate elements in examples that have probably been elicited. Examples (6c-6f, 6g) show post-verbal subjects and direct objects. In sentences (18) and (19), we find post-verbal indirect objects functioning as recipients and causees respectively. In (19) also the direct object appears after the verb.

(18) Adyghe (Paris 1974: 220) [D]
      indirect object (recipient)
      λ̣ə-m   qʷə̂z r-j-e-tə   ŝʷəzə-m
      man-OBL pear DAT-3SG.A-DYN-give woman-OBL
      'The man gives / is giving a pear to the woman.'

(19) Adyghe (Letuchiy 2009: 388) [H]
      indirect object (causee)
      *t-ṣ̂ə-n   t-ŝe-re-p-ṣ̂ə,   qe-ẑʷə-ʁa-ʔʷə-ba   ade **wered***
      1PL.A-do-MOD 1PL.A-do-DYN-NEG-CS DIR-2PL.A-CAUS-say-PRT well song
      a-š'
      that-OBL
      'As for doing we will not do (anything), so he may sing a song!' (lit. 'You let him / you cause him to sing a song.')

The post-verbal elements can be definite or indefinite (which for nouns correlates with case marking, i.e. the omission of the case marking indicates indefiniteness), e.g. in (6c) and (6d) post-verbal direct objects bear the absolutive suffix -*r* and are definite, whereas in (19) the direct object is not marked for case and thus indefinite.

Not only arguments but also adjuncts can occur after the verb. In (6c) and (6e), the case-marked noun denoting a location is placed in post-verbal position. In the literature, one also finds examples of complement clauses that follow the matrix

clause with a variety of complement-taking predicates, e.g. 'want', 'need, must', 'fear', or 'know' (20) and some examples of post-verbal adverbial clauses (21).

(20)　Adyghe (Serdobolskaya & Motlokhov 2009: 533) [H]
　　　s-j-e-negʷəje　　　　*[ṭʷə qe-s-hə-n-ew]*
　　　1SG.IO-DAT-DYN-suppose two DIR-1SG.IO-carry.away-MOD-ADV
　　　'I suppose / fear that I get a two (= bad mark).'

(21)　Adyghe (Testelets 2009: 691) [H]
　　　pŝaŝe-r　qe-ʁə-ʁ　　　*[sə-z-de-gʷəš'əʔe-m]*
　　　girl-ABS DIR-cry-PST 1SG.ABS-REL.TEMP-COM-talk-OBL
　　　'The girl cried when I talked to her (= the girl or another female person).'

In short, Adyghe allows for post-verbal arguments and adjuncts with various syntactic functions and grammatical roles as well as for post-verbal complement and adverbial clauses. Based on the literature we cannot say whether post-verbal elements are a marginal phenomenon, and if some syntactic functions or grammatical roles are more frequently found there than others, due the lack of previous corpus studies.

Therefore, I examined post-predicate elements in declarative main clauses of 20 texts containing a total of 6,146 words distributed over 1,154 main clauses. The texts have been glossed by the researchers who recorded them (Paris, Höhlig and Feer) and by myself. In a second step, I manually annotated them for the presence or absence of post-predicate elements (see Section 3.2 below and Appendix for more information on the texts). I did not count all overt and covert arguments and all overt adjuncts and their positions with respect to the verb but only the post-predicate ones (and I am fully aware of the fact that this makes comparison with data from WOWA corpora impossible[2] and evaluation of frequency of post-predicate elements rather speculative).

The results of these counts are presented in Table 1. The first thing to notice is that for a language that is described as having "flexible word order", in natural texts the position after the verb is not frequently occupied despite being easily filled in elicitation. In around one out of ten main clauses, we find post-predicate elements of which the largest group are focus / modal particles.

The most common particle is *nah* 'more' (22), (29), others being *nəʔa* 'only' (17), *armərme* 'otherwise, if not this', *mewš'tew* 'like this' and also Russian loans such as *uže* 'already' and *daže* 'even' and the Turkish indirect evidential particle

---

[2]Furthermore, in WOWA only non-subject referential expressions are considered, rather than all kinds of post-verbal items (e.g. modal particles, clausal constituents).

Table 1: Grammatical and semanto-pragmatic functions of post-predicate elements in Adyghe texts (both diaspora and homeland)

Arguments	**33**
Subjects (independent of transitivity and semantic role)	20
Direct objects (of transitive and ditransitive verbs)	10
Indirect objects (2 addressees and 1 beneficiary; no recipients attested)	3
Adjuncts	**46**
Possessors	4
Instruments	3
Temporal adverbials	12
Spatial adverbials (13 locations, 6 goals, 2 sources)	21
Manner adverbials	5
Particles	**29**
Clauses	**19**
Adverbial clauses	16
Complement clauses	2
Relative clauses	1
Total	**126**

(je)məš and the particle *yani* 'that is, namely'. Pragmatic particles such as the ones listed are expected to have a great deal of freedom and thus do not really support the claim that the word order of Adyghe is flexible.

(22) Adyghe (Höhlig 1997: 218) [H]
       jež' djela-ʁe **nah**
       self fool-PST more
       'He was a fool.'

Due to the way in which I annotated and counted the data, I cannot make any statements concerning the relative frequency of certain types of postverbal arguments and adjuncts. I will instead present the attested post-predicate elements and discuss their information-structural properties, whenever possible taking into account audio recordings of some of the examples that were provided to me by Monika Höhlig.

First of all, there is one construction that regularly leads to post-predicate elements according to the literature and the examined texts, namely direct reported speech (Rogava & Kerasheva 1966: 395–402). When the matrix clause with the verb of speech interrupts the quote or follows it, the usual word order is reversed and the subject (17) or the addressee, if there is no overt subject (23) follows the verb of speech.

(23)   Adyghe (Rogava & Kerasheva 1966: 396) [H]
   səd-a   aməd wə-z-ʁegʷəmeč'ə-re-r?   j-e-wəpčə-ʁ
   what-Q Amid 2SG.PR-REL.TEMP-disturb-DYN-ABS 3SG.A-DYN-ask-PST
   bzəλfəʁe-m
   woman-OBL
   '"What is it that worries you, Amid?" (s/he) asked the woman.'

Occasionally one encounters thetic sentences in which the subject appears in clause-final position. A thetic utterance is fully focused with no topical constituent. Introductory clauses in traditional fairy tales or narratives about well-known personalities may follow this pattern, as in the following example from a story about a famous singer and composer. The V-S pattern for introductory thetic sentences is also attested in other verb-final languages from the Caucasus such as Kartvelian and East Caucasian (Forker 2021, Forker 2024 [this volume]). In example (24), the speaker makes a short break before uttering the subject encoded as personal name (and there is no falling intonation in contrast to the examples discussed below).

(24)   Adyghe (courtesy of M. Höhlig) [H]
   mə č'əle ẑə-m   de-sə-ʁ   [weredəʔʷe ʔaze-w]
   this aul   old-OBL COM-sit-PST singer   art.master-ADV
   [qebar-xe-r-jə   zeč'e qə-ʔʷate-w] [txədeẑ-xe-r-jə   qə-ʔʷate-w]
   story-PL-ABS-ADD all   DIR-tell-ADV legend-PL-ABS-ADD DIR-tell-ADV
   ḳʷaj zefes
   Kway Zefes
   'In this old village lived the master singer, story-teller and legend-teller, Kway Zefes.'

In introductory thetic sentences, also other constituents besides the subject can follow the verb, like the instrument in (25), which has a falling intonation towards the end of the sentence and no intonational break before the post-predicate element.

(25) Adyghe (courtesy of M. Höhlig) [H]
neməc-xe-r   *jewjə ž'ambeč'əje qə-da-ha-ʁ*   ***mašine-č'e***
German[R]-PL-ABS PRT   Dzhambechi DIR-LOC-enter-PST car[R]-INST
[beginning of narrative] 'The Germans came to Dzhambechi by car.'

The next example illustrates left dislocation of the subject combined with a post-predicate presumptive pronoun. In left dislocation, a referential constituent both precedes and is dislocated from a core clause with which it is associated. Within the core clause, there is an anaphoric co-referential resumptive pronoun (Westbury 2016). A typical function of left dislocation is to introduce referents that are not purely brand-new, but merely inactive. This means that the referent is assumed to be identifiable, but only minimally accessible, having been in one way or another evoked in the prior discourse or in the extra-linguistic context. This is what we find in (26): The story is about a group of boys, one of which is singled out by means of the left-dislocated element given in curly brackets and then resumed through the pronoun in subject function following the verb.

(26) Adyghe (Höhlig 1997: 234) [H]
{japλenere, z-jə-hatəq   *z-jə-haləʁʷə*
fourth   REL.IO-POSS-flat.bread REL.IO-POSS-bread
qə-z-ŝʷe-t-təʁʷə-ʁe-m}   *qe-ḳʷe-ž'ə-ʁ*   ***a-r-jə***
DIR-REL.TEMP-MAL-1PL.A-steal-PST-OBL DIR-go-RE-PST that-ABS-ADD
[All went back home from school. The three of us are standing there.]
'The fourth one whose flat bread, whose bread we had stolen, he also went home.'

In the following, I use the term "topic" in the sense of "aboutness topic" (e.g. Krifka 2007). The topical item is identified through the utterance and then some piece of information about it is provided in the comment. Post-predicate elements that function as aboutness topics in Adyghe are not emphasized by means of intonation. The pitch accent is usually somewhere at the beginning, and towards the end of the utterance the intonation it falls and becomes flat; the voice sometimes gets lower and quieter such that in some examples, the last syllable of the final post-predicate element is barely audible. For instance, in (27) the pitch accent of the second clause falls on the first verb (*šxe-n*).

(27) Adyghe (courtesy of M. Höhlig) [H]
a-xe-r-jə   *a*   *šxe-š't; šxe-n*   *faje-ba*   ***çəfə-r***
that-PL-ABS-ADD that eat-FUT eat-MOD must-PRT human.being-ABS
[Talking about the behavior of German and Soviet soldiers during WWII. The speaker is finishing her narration and after this she switches to a

different topic.] 'They (=Germans) as well will eat it; the human being must eat.'

Topical elements often convey given information and can, e.g., be expressed by means of pronouns. In all following examples, the post-predicate elements represent given information. For those examples that have personally been provided to me by Monika Höhlig (29, 31-33) I could clarify the intonation. In all examples, the post-predicate elements are deaccented by means of a falling intonation.

(28) Adyghe (Höhlig 1997: 236) [H]
  subject
  tawərəχ-ew š'ə-t-ep   **a-r**
  legend-ADV LOC-stand-NEG that-ABS
  [Here is what I want to tell you.] 'It is not an (old) legend.'

(29) Adyghe (courtesy of M. Höhlig) [H]
  discourse particle + subject
  ǯ'ə   zewže t-ŝʷeḳʷedə-ʁe-xe   **nah   a-xe-r**
  now all   1PL.IO-get.lost-PST-PL more.than that-PL-ABS
  [talking about various Adyghe traditions] 'Now we lost all those (i.e. traditions).'

(30) Adyghe (Höhlig 1997: 219) [H]
  direct object
  te   q-jə-p-hə-ʁ   **a   š'aλe-r**
  from.where DIR-LOC-2SG.A-carry.away-PST that bucket-ABS
  s-ʔʷ-əj   sə-kʷəwa-ʁ
  1SG.A-speak-ADD 1SG.ABS-shout-PST
  ' "From where did you take that bucket?" I said shouting.'

(31) Adyghe (courtesy of M. Höhlig) [H]
  direct object
  *... neməč' ha-xe-m   a-šxə-ʁ   **ṭʷə-jə**   ...*
    other   dog-PL-OBL 3PL.A-eat-PST two-ADD
  [Talking about the mysterious disappearance of turkeys and the fault of the dogs] '... the other dogs ate (our) two (turkeys) ...'

(32) Adyghe (courtesy of M. Höhlig) [H]
possessor
jəλes t̪ʷeč'ə-re   blə-re   ə-nəbž'ə-ʁ   č'etəwə-m
year twenty-COORD seven-COORD 3SG.POSS-age-PST cat-OBL
[The dogs were also afraid of it, it was a very good cat] 'The cat was 27
years old.'

(33) Adyghe (courtesy of M. Höhlig) [H]
manner adverbial (adjunct)
woot, nepeməč' tradicie-w   səd ž'ərjə t-xe-λə-n
well[R] other   traditions[R]-ADV what again 1PL.IO-LOC-lie-MOD
aj   fede-w
that.OBL be.similar-ADV
[talking about various Adyghe traditions] 'Well, which other traditions
do we have here like that.'

The postpredicate elements in examples (28-33) can hardly be said to be high-
lighted or emphasized, in contrast to what Jakovlev & Ashkhamaf (1941: 91) have
stated about the post-predicate elements in examples (6c-6f, 6g).

However, post-predicate elements are not always topical and do not always
convey given information. The postverbal adjuncts in (34) and (35) convey new
information and are part of the "presentational" or "information" focus of the
sentence in which they occur. Information focus expresses the most important or
new information in the utterance (Krifka 2007). Because I lack recordings of (34)
and (35) I cannot say anything about the intonational patterns. Yet I hypothesize
that the elicited examples (6c-6f, 6g) from Jakovlev & Ashkhamaf (1941: 91) should
be interpreted in a similar manner, i.e. the post-predicate elements as being part
of the information focus.

(34) Adyghe (Höhlig 1997: 255) [D]
spatial adverbial (Goal)
qe-ze-ḳʷe-ž'ə-m,   se sə-q-jə-č'ə-ʁ   mutfaqə-m
DIR-REL.TEMP-go-RE-OBL 1SG 1SG.ABS-DIR-LOC-go.out-PST kitchen[T]-OBL
[In the evening my husband came home.] 'When he came I went out to
the kitchen.'

(35)   Adyghe (Höhlig 1997: 274) [D]
       temporal and spatial adverbials (time span and location)
       *adəɣa   bze   te-gʷəš'aʔ   **zepət   wəne-m-jə***
       Adyghe language 1PL.A-speak all.time house-OBL-ADD
       kʷež'e-m-jə
       village[T]-OBL-ADD
       [talking about language knowledge and language use] 'We always speak
       Adyghe at home and in the village.'

Adverbial clauses seem to be more variable concerning their position than
complement clauses (or relative clauses), whereby we can notice that post-
predicate adverbial clauses repeatedly express cause (36) or purpose (14). For pur-
pose clauses, we can assume an explanation based on iconicity. The purpose of
an action resembles a spatial Goal and thus the linear order of the clauses reflects
the temporal or spatial order of the events. Causal or concessive adverbial clauses
refer to situations that do not necessarily occur prior to the situation expressed
in the main clause but rather provide the reason or cause for it (36) or signal
contrast or concession in relation to it (37). In both examples (36) and (37), the
adverbial clauses constitute separate intonational units, i.e. there is a short break
between the preceding main clauses and the sentence-final adverbial clauses.

(36)   Adyghe (courtesy of M. Höhlig) [H]
       adverbial clause expressing a cause or reason
       abʒexa   bze-č'e   t-jə-wənaʁʷe   parjə   gʷəš'əʔa-re-p
       Abdzakh language-INST 1PL-POSS-family nobody talk-DYN-NEG
       [č'emgʷe-me   t-a-xe-s-ŝə]
       Temirgoi-OBL.PL 1PL.ABS-3PL.IO-LOC-sit-CVB
       'In our family nobody speaks Abdzakh because we live among the
       Temirgoi.'

(37)   Adyghe (courtesy of M. Höhlig) [H]
       concessive conditional clause
       xet   š'əš'   qə-tje-wa-ʁe-m-jə   jə-pče   ʔʷə-jə-xə-š't.
       who part DIR-LOC-beat-PST-COND-ADD POSS-door LOC-3SG.A-open-FUT
       [wəč'aḳʷe qə-ʔʷə-ha-ʁe-m-jə   a-š'   fed]
       killer   DIR-speak-lat-PST-COND-ADD that-OBL similar
       '[You know the Adyghe traditions], whoever might knock, you will open
       your door, even if it were a killer.'

By contrast, in chaining constructions that also represent adverbial subordination, the dependent clauses do not follow the main clause because the chained clauses do not or only to a limited extent convey temporal reference but their temporal interpretation depends on the inflected predicate of the main clause. The temporal order of events in chaining constructions is therefore iconically reflected in the order of the chained clauses and the main clause (Forker In press).

The sentence in (38) illustrates a complement clause that follows the matrix clause. This order is not particularly common in the surveyed texts which contain only two instances (Table 3). As said above, I did not count all complement clauses in my data in order to assess the frequency of preposed vs. postposed ones and compare them with adverbial clauses. However, the paper by Serdobolskaya (2016) about complement clauses in Adyghe contains many examples of sentence-final complements which at least shows that in elicitation they are easily available (and probably triggered by Russian word order patterns).

(38)   Adyghe (Höhlig 1997: 251) [H]
       complement clause
       s-ŝe-re-p   *[qə-ze-re-s-ʔʷete-r]*
       1SG.A-know-DYN-NEG DIR-REL.IO-MNR-1SG.A-tell-ABS
       'I do not know how I should tell it.'

3.2 Summary of the quantitative analysis

Table 2 gives a summary of the texts ordered by place of recording. The texts from Adygea (Northern Caucasus, homeland [H]) are almost double in terms of number of words and clauses but also with respect to the number of post-predicate elements per clause when compared with the texts from Turkey (diaspora [D]). The texts from Turkey have been recorded between 1969 and 1990 and are thus older than the texts from Adygea, which have been recorded between 1990 and 2017. All texts represent three different Adyghe dialects, namely Shapsug (3 texts), Abzakh (15 texts) and Temirgoy (2 texts). The texts can be divided into two categories, namely into (A) traditional folklore narratives and folkloristic anecdotes and (B) personal, autobiographical narrations. The two genres also exhibit differences in style. In traditional narratives a formal style prevails (e.g. fewer loan words, long and structurally complicated sentences) whereas in personal accounts an informal style is found (e.g. many Russian loan words, shorter and structurally simpler sentences). Detailed information about the sources of the texts, the time and the place of the recording can be found in the Appendix.

Table 2: Comparing post-predicate elements in texts from Adygea and from Turkey

	Adygea	Turkey	Total
# texts	11	9	20
# words	4,083	2,063	6,146
# main clauses	770	384	1,154
# post-predicate elements	100	26	126
% post-predicate elements per clause	**12.99%**	**6.77%**	**10.91%**

If we compare the two genres — personal accounts and traditional narratives / anecdotes — with each other we find that personal accounts have more than twice as many post-predicate items than traditional narratives. This difference is more pronounced in Turkey than in Adygea, but it is observable in both places (Table 3).

Table 3: Comparing post-predicate elements according to genre

	Personal accounts		Traditional narratives	
	Adygea	**Turkey**	**Adygea**	**Turkey**
# main clauses	554	158	216	226
# post-pred	79	19	21	7
% post-pred per clause	**14.26%**	**12.02%**	**9.7%**	**3.1%**
% post-pred per clause	**13.76%**		**6.3%**	

Thus, it seems that place as well as genre correlates with the number of post-predicate elements. Traditional narratives from Turkey contain the lowest number of them (7 items), which are almost exclusively focal / modal and evidential particles, in particular the particle *nah*. The traditional narratives from Adygea also have post-predicate subjects in reported speech constructions (17) and some more post-predicate elements in various functions and of different formal types. On the other side of the spectrum, we find the personal accounts from Adygea that have as many as 79 post-predicate elements.

The difference seems to be due to genre in combination with style, but it might have been enlarged by language contact with two typologically and genealogically different contact languages, Turkish and Russian. Turkish is a head-final

language; the unmarked word order is SOV (e.g. Erguvanlı 1984: 43; Göksel & Kerslake 2005: 338). Post-predicate elements are restricted to informal style and mostly found in spoken language, but also in more informal writing. They are never stressed and generally backgrounded, i.e. they convey information that is shared by speaker and addressee; WH-words are not allowed to be placed after the predicate (Erguvanlı 1984: 43–63; Göksel & Kerslake 2005: 345–346). Not only S, DO and IO may occur in post-verbal position, but also various adverbials and subordinate clauses (complement clauses, relative clauses and adverbial clauses; Erguvanlı 1984: 63–66). Although these authors confirm that speakers accept and produce post-posed elements of various kinds (under particular information-structural conditions), the corpus data from colloquial standard Turkish (Iefremenko 2021) in WOWA indicate that such structures are relatively infrequent in actual usage: about 94% of the coded non-subject and non-pronominal constituents in this data set are pre-verbal.

By contrast, Russian has been characterized as having free or pragmatically governed word order with many features of SVO languages (Dryer 2022). A small corpus study by Billings (2015) (500 transitive clauses from the Russian National Corpus) has shown that SVO dominates (89.6%), SOV is the second most frequent order (4.4%) and all other orders are also attested but rather infrequently. It is possible but not necessary that the differences between the texts from Adygea and the texts from Turkey are due to contact with Russian and Turkish respectively. The detailed study by Höhlig (1997) that compares Adyghe in contact with Russian in the homeland and Turkish in the diaspora shows many examples of how both contact languages influence Adyghe in various parts of the grammar although it does not explicitly discuss word order. In order to test the contact hypothesis in the future, it is necessary to study the word order in the oldest written examples of Adyghe which have been produced by speakers who probably have not been exposed to Russian or Turkish to the same degree as today's speakers, which goes beyond the scope of this study. And when testing contact influence, we have to keep in mind that language contact can not only lead to a change towards another pattern, but also to an increased rigidity in the inherited pattern (e.g. Namboodiripad et al. 2019). Thus, the preference for verb-final utterances in Adyghe in Turkey could in principle also be due to such an effect and not to a preference for 'copying' the Turkish pattern.

4 Post-predicate elements in other Northwest Caucasian languages: Kabardian, Ubykh, Abkhaz and Abaza

In general, all Northwest Caucasian languages are described as head-final (Arkadiev & Lander 2021), but with flexible word order that permits all logically possible permutations. For Ubykh, the only Northwest Caucasian language that is no longer spoken, the latest grammar states that postverbal constituents are extremely rare in Ubykh texts (Fenwick 2011: 151–153), which must have been collected in Turkey after the forced exodus of the Northwest Caucasian people to the Ottoman Empire in the second half of the 19th century. In addition to SOV, only OSV is a relatively common alternative, but apparently marked order in Ubykh that "appears to provide a certain degree of emphasis to the fronted absolutive object" (Fenwick 2011: 151). Judging from the publication of traditional narratives from all Northwest Caucasian languages in Colarusso (1999), SOV is clearly dominant, which is in accordance with my Adyghe data presented in Section 3.

In elicitation, which for Ubykh cannot be done anymore, Kabardian, Abkhaz and Abaza are similar to Adyghe. Kumakhov & Vamling (2009: 112–130) provide elicited Kabardian examples of postverbal subjects and objects in nominal and pronominal form. They add that heaviness influences the position of arguments, i.e. heavy arguments are preferably placed at the beginning of sentences but can also occur at the end. Furthermore, they show that because proper nouns only facultatively take case markers for absolutive and oblique they require a strict S-DO-IO order but can still be placed in post-predicate position.

(39)   a. Kabardian (Kumakhov & Vamling 2009: 129)
            S-DO-V-IO
            Murat Nazir χʷ-i-še-nu-s'   **Asłen.**
            Murat Nazir VER-3SG.A-lead-FUT-ASSERT Aslan

        b. S-V-DO-IO
            Murat χʷišenus'   **Nazir Asłen.**
            Murat VER-3SG.A-lead-FUT-ASSERT Nazir Aslan

        c. V-S-DO-IO
            χʷišenus'   **Murat Nazir Asłen.**
            VER-3SG.A-lead-FUT-assert Murat Nazir   Aslan

            'Murat will lead Nazir to Aslan.'

Similarly, Abkhaz does not employ case marking for core arguments, such that in case of ambiguity the order of subject and object is fixed as S-O. Postverbal placement of arguments is allowed if this order is kept (Chirikba 2003a: 60)

Abkhaz allows for thetic utterances that introduce new referents (often in subject position) to occur after the verb, (40) as is also the case for Adyghe.

(40) Abkhaz (Chirikba 2003b: 259)
VS
jə-qʼa-n   *ažər-jə-pa-cʷa*   *hʷa*   *jʷə-žja*   *[a]-aj.šj-cʷa*
3PL-be-PST.FIN Adzhyr-3-son-PL QUOT two-HUM ART-brother-PL
'There lived two brothers (reportedly known as) Sons of Adzhyr.'

In the following example from Abaza, the direct object appears after the verb:

(41) Abaza (courtesy of P. Arkadiev)
IO-V-DO
mhamatgʼaráj   *j-ʕa-jə-r-t-ṭ*   *á-dgʼəl*
muhamat.girey 3SG.N.ABS-CISL-3SG.M.IO-3PL.ERG-give(AOR)-DCL DEF-land
'They gave land to Muhamat-Girey.'

In answers to WH-questions in Kabardian, the narrowly focused element can occur after the verb, but this order is marked compared to the neutral SOV order Kumakhov & Vamling (2009: 142).

In direct reported speech constructions with quotes preceding the matrix verb of speech subjects, or if the subject is omitted, other remaining constituents regularly follow the verb as shown in examples from Kabardian (e.g. Colarusso 1999: sentence 70 from the Kabardian texts), Abkhaz (Chirikba 2003a, sentence 15) and Abaza (42). This pattern was also noted for Adyghe (Section 3.1).

(42) Abaza (courtesy of P. Arkadiev)
QUOTE-V-S
wə-z-ʕa-j   *aĉə-ja?*   *j-hʷa-ṭ*   *a-ʁəč'*,
2M.ABS-REL.RSN-CISL-come what-QN 3M.ERG-say-DCL DEF-thief
zaḳ-gʼəj   *a-jə-m-rə-χ'-ʒa-ḳʷa*
one.CLN-ADD 3N.IO-3M.ERG-NEG-CAUS-cool.down-INTF-CVB.NEG
'"Why did you come?" said the thief, not calming down at all.'

Northwest Caucasian languages also allow certain types of adverbial clauses and complement clauses in post-predicate position (e.g. examples in Kumakhov & Vamling 2009: 192; Arkadiev 2020). For instance, according to Chirikba 2003a: 64), adverbial clauses with the present converb can either precede or follow the main clause.

5 Discussion

Summarizing, we can state that post-predicate elements are relatively rare in Adyghe texts, despite the possibility of eliciting them. There are no restrictions concerning grammatical functions and parts of speech of post-predicate elements in Adyghe — they can be arguments and adjuncts of various kinds, from short pronouns to more elaborate noun phrases, particles, etc. When comparing grammatical functions, I found more post-predicate subjects than objects and goals. However, I do not have data about the relative frequency. Thus, it might be the case that for functions such as subjects, locations or goals the probability of occurring in a position after the verb is higher than for objects, but this needs to be tested in future research. Subjects are probably overall more frequently represented in the texts than objects (or goals), since the vast majority of verbs have subject arguments, but only a sub-set of them allow for or even require objects or goals. Direct reported speech constructions in which the quote precedes the verb of speech regularly put the subject (or indirect object in the function of addressee if there is no subject) after the verb.

The majority of (non-clausal) post-predicate elements are topical, such that referents encode given information, but occasionally one also finds focal elements. In text-initial utterances, the newly introduced referent sometimes follows the verb. There is a correlation with genre and geographical origin of speakers. More traditional genres such as legends, fairy tales and anecdotes show a smaller amount of post-predicate elements than personal accounts and autobiographies; and texts recorded in Adygea contain more post-predicate elements than those recorded in Turkey, which might be due to language contact (mainly between Russian and Adyghe speakers in Adygea).

The other Northwest Caucasian languages seem to behave similarly to Adyghe. For Ubykh, which is no longer spoken, texts gathered in Turkey point to a very strong tendency for verb-final word order. This fits well to the observed difference between the Adyghe texts from Turkey and those from Adygea. It also matches with the observations on Laz, a Kartvelian outlier in Turkey (Forker 2024 [this volume]).

Finally, when comparing Northwest Caucasian languages to the other two indigenous language families in the Caucasus (see Forker 2024 [this volume]) we can safely state that Northwest Caucasian languages shows the biggest preference for verb-final order and thus the lowest number of post-predicate items.

Abbreviations

1	first person	INTF	intensifier
2	second person	IO	indirect object
3	third person	LAT	lative
A	agent	LNK	linking element
ABS	absolutive	LOC	locative
ADD	additive	M	male
ADJ	adjective	MAL	malefactive
ADV	adverbial	MOD	modal
ASSERT	assertive	N	neuter
AUX	auxiliary	NEG	negation
BEN	benefactive	OBL	oblique
CAUS	causative	PL	plural
CISL	cislocative	POSS	possession
CLN	non-human numeral classifier	PR	possessor series of personal prefixes
COM	comitative	PRT	particle
COND	conditional	PST	past tense
COORD	coordination	Q	question marker
CS	causal	QN	non-human question
CVB	converb	R	Russian loan
DAT	dative	RE	refactive
DCL	declarative	REC	reciprocal
DEF	definite	REL	relativizer
DIR	directional	RSN	reason
DYN	dynamic	SG	singular
EL	elative	T	Turkic loan
ERG	ergative	TEMP	temporal
FIN	finite	TRM	terminative
FUT	future	VER	version
INST	instrumental		

Digital corpora

Adyghe

- Catherine Paris. 1974. *La Princesse Kahraman: Contes d'anatolie en dialecte chapsough (tcherkesse occidental).* Paris: SELAF (Centre National de la Recherche Scientifique)

- The Pangloss Collection. 2023. *"Detchiyimko Pchipii and the prince Krmirgoy Aytekko", "La princesse Kahraman", "La guerre entre les Qabardes et les Abzakhs" (Shapsug Adyghe).* https://cocoon.huma-num.fr/exist/crdo/ meta2/cocoon-af3bd0fd-2b33-3b0b-a6f1-49a7fc551eb1 (11 April, 2023)

Ankara Turkish

- Kateryna Iefremenko. 2021. Oghuz (Ankara). In Geoffrey Haig et al. (eds.), *WOWA: Word order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* Bamberg: University of Bamberg. https://multicast.aspra.uni-bamberg.de/resources/wowa/ #turkic (11 April, 2023)

References

Arkadiev, Peter M. 2020. *Abaza sketch grammar.* Moscow: RGGU.

Arkadiev, Peter M. & Yury A. Lander. 2021. The Northwest Caucasian languages. In Maria Polinsky (ed.), *The Oxford handbook of languages of the Caucasus*, 369–446. Oxford: Oxford University Press.

Arkadiev, Peter M., Alexander B. Letuchiy, Nina Sumbatova & Yakov G. Testelets. 2009. *Aspekty polisintetizma: Očerki po grammatike Adygejskogo jazyka [Aspects of polysynthetism: Essays on the grammar of the adyghe language].* Moscow: Russian State University for the Humanities.

Arkadiev, Peter M. & Timur A. Maisak. 2018. Grammaticalization in the North Caucasian languages. In Heiko Narrog & Bernd Heine (eds.), *Grammaticalization from a typological perspective*, 116–145. Oxford: Oxford University Press.

Arkhangelskiy, Timofey, Irina Bagirokova, Yury Lander & Anna Lander. 2018–2023. *West Circassian (adyghe) corpus.* Online. http://adyghe.web-corpora.net/.

Billings, Stephanie Kay. 2015. *A corpus-based analysis of Russian word order patterns.* Brigham Young University. (MA thesis). https://scholarsarchive.byu. edu/etd/5624.

Chirikba, Viacheslav A. 2003a. *Abkhaz.* Munich: Lincom.

Chirikba, Viacheslav A. 2003b. Evidential category and evidential strategy in Abkhaz. In Alexandra Y. Aikhenvald & R. M. W. Dixon (eds.), *Studies in evidentiality*, 243–272. Amsterdam: Benjamins.

Colarusso, John. 1999. *A North West Caucasian reader.* Hamilton: McMaster University.

Dryer, Matthew S. 2022. One way in which Slavic languages really are SVO languages. *Theoretical Linguistics* 48(1-2). 49–55.

Erguvanlı, Eser. 1984. *The function of word order in Turkish grammar.* Berkeley: University of California Press.

Feer, Rainer. 2019. *The grammar, function and grammaticalization of Adyghe clause linking devices for cause and temporal succession.* Cheltenham: University of Gloucestershire. (MA thesis).

Fenwick, Rohan S. H. 2011. *A grammar of Ubykh.* Munich: Lincom.

Forker, Diana. 2021. Information structure in languages of the Caucasus. In Maria Polinsky (ed.), *The Oxford handbook of languages of the Caucasus*, 973–1002. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780190690694.013.28.

Forker, Diana. 2024. Post-predicate elements in Kartvelian and East Caucasian. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 283–307. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266349.

Forker, Diana. In press. Clause chaining in Adyghe (West Caucasian). In Hannah S. Sarvasy & Alexandra Y. Aikhenvald (eds.), *Clause chaining in the world's languages.* Oxford: Oxford University Press.

Göksel, Aslı & Celia Kerslake. 2005. *Turkish: A comprehensive grammar.* London/New York: Routledge.

Höhlig, Monika. 1997. *Kontaktbedingter Sprachwandel in der adygeischen Umgangssprache im Kaukasus und in der Türkei.* Munich: Lincom.

Iefremenko, Kateryna. 2021. Oghuz (Ankara). In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA: Word order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* Bamberg: University of Bamberg. https://multicast.aspra.uni-bamberg.de/resources/wowa/#turkic (11 April, 2023).

Jakovlev, Nikolaj F. & Daud A. Ashkhamaf. 1941. *Grammatika Adygejskogo literaturnogo jazyka [A grammar of the Adyghe literary language].* Moscow: Izdatel'stvo AN SSSR.

Kimmelman, Vadim I. 2011. Auxiliaries in Adyghe. Paper presented at the *conference on Caucasian languages,* Leipzig, May 2011.

Krifka, Manfred. 2007. Basic notions of information structure. *Interdisciplinary Studies on Information Structure* 6. 13–55.

Kumakhov, Mukhadin A. & Karina Vamling. 2009. *Circassian clause structure.* Malmö: Malmö University.

Lander, Yury A. 2014. *Reljativizacija v polisintetičeskom jazyke: Adygejskie otnositel'nye konstrukcii [Relativization in a polysynthetic language: Adyghe rel-*

ative clauses in typological perspective]. Moscow: University of Moscow. (Doctoral dissertation).

Lander, Yury A. 2017. Nominal complex in West Circassian: Between morphology and syntax. *Studies in Language* 41(1). 76–98.

Letuchiy, Alexander B. 2009. Affiksy benefaktiva i malefaktiva: Sintaksičeskije osobennosti i krug upotreblenij [Affixes of benefactive and malefactive: Syntactic properties and use]. In Peter M. Arkadiev, Alexander B. Letuchiy, Nina Sumbatova & Yakov G. Testelets (eds.), *Aspekty polisintetizma: Očerki po grammatike Adygejskogo jazyka*, 329–372. Moscow: Russian State University for the Humanities.

Namboodiripad, Savithry, Dayoung Kim & Gyeongnam Kim. 2019. English dominant Korean speakers show reduced flexibility in constituent order. In Daniel Edmiston, Marina Ermolaeva & Emre Hakgüder (eds.), *Proceedings of the fifty-third annual meeting of the Chicago linguistic society*, 247–260. Chicago: Chicago Linguistic Society.

Paris, Catherine. 1974. *La Princesse Kahraman: Contes d'anatolie en dialecte chapsough (tcherkesse occidental).* Paris: SELAF (Centre National de la Recherche Scientifique).

Rogava, Georgij V. & Zajnab I. Kerasheva. 1966. *Grammatika Adygejskogo jazyka* [A Grammar of Adyghe]. Krasnodar: Krasnodarskoe Knižnoe Izdatel'stvo.

Serdobolskaya, Natalia V. 2016. Semantics of complementation in Adyghe. In Kasper Boye & Petar Kehayov (eds.), *Complementizer semantics in European languages*, 291–744. Berlin: De Gruyter Mouton.

Serdobolskaya, Natalia V. & Alexander V. Motlokhov. 2009. Semantika konstrukcij s sentencial'nymi aktantami v Adygejskom jazyke [Semantics of complementation in Adyghe]. In Peter M. Arkadiev, Alexander B. Letuchiy, Nina Sumbatova & Yakov G. Testelets (eds.), *Aspekty polisintetizma: Očerki po grammatike Adygejskogo jazyka [Aspects of polysynthesis: Essays on Adyghe grammar]*, 498–559. Moscow: Russian State University for the Humanities.

Testelets, Yakov G. 2009. Nevyražennye aktanty v polipredikativnoj konstrukcii [Unexpressed arguments in bipredicative constructions]. In Peter M. Arkadiev, Alexander B. Letuchiy, Nina Sumbatova & Yakov G. Testelets (eds.), *Aspekty polisintetizma: Očerki po grammatike Adygejskogo jazyka [aspects of polysynthesis: Essays on Adyghe grammar]*, 654–711. Moscow: Russian State University for the Humanities.

Testelets, Yakov G. & Yury A. Lander. 2017. Adyghe. In Michael Fortescue, Marianne Mithun & Nicholas Evans (eds.), *The Oxford handbook of polysynthesis*, 948–970. Oxford: Oxford University Press.

The Pangloss Collection. 2023. *"Detchiyimko Pchipii and the prince Krmirgoy Aytekko", "La princesse Kahraman", "La guerre entre les Qabardes et les Abzakhs" (Shapsug Adyghe).* https://cocoon.huma‑num.fr/exist/crdo/meta2/cocoon‑af3bd0fd‑2b33‑3b0b‑a6f1‑49a7fc551eb1 (11 April, 2023).

Westbury, Josh. 2016. Left dislocation: A typological overview. *Stellenbosch Papers in Linguistics Plus* 50. 21–45.

Chapter 12

Word order variation in Romeyka

Laurentia Schreiber[a] & Mark Janse[b]

[a]University of Bamberg [b]University of Cambridge

The present chapter describes word order variation in Romeyka based on the multilingual spoken language dataset of the Word Order in Western Asia (WOWA) corpus. Descending from VO ancestry, Romeyka shows under contact from Turkish increasingly head-final syntax. While cross-linguistically relevant factors such as semantic role, flagging and phonological weight do not offer much explanation for the flexibility between pre- and post-verbal constituents in Romeyka, information structure and phrase type do seem to be relevant. In addition, inter-speaker variation has been found significant to account for word order variation in Romeyka, suggesting that in a setting of language shift, individual forms of bilingualism affect word order.

1 Introduction

Romeyka is a variety of Pontic Greek (henceforth PG) that is at present still spoken by Muslims in Trabzon Province in northeastern Turkey, although its status can be characterized as endangered (Schreiber 2016; Schreiber & Sitaridou 2017). Romeyka belongs to the Hellenic branch of Indo-European. Like most modern Greek dialects, it is descended from postclassical (Koine) Greek, with an undisputable Ionic substrate dating back to the Milesian colonization of the Black Sea coast in the 7th–6th c. BCE. The closest relatives of Pontic are Pharasiot and Cappadocian, collectively known as (East) Asia Minor Greek (henceforth AMG).[1]

---

[1]The designation "East Asia Minor Greek" is due to Janse (e.g., 2008: 191–192, 20202020: 202–203) and implied in Dawkins' "Greek of eastern Asia Minor" (1916: 213). The qualification "East" is necessary in light of Ralli's (2020) broader acceptation of the geographical designation "Asia Minor Greek" (Ralli 2020). For a different view on the internal relationship between Pontic, Pharasiot and Cappadocian see Karatsareas (2016: 40–55).

The location of Romeyka, which is nestled in the remote, mountainous area of the Pontic Alps, has probably contributed to the preservation of the language and of some archaic features and facilitated the development of three main dialect areas around the townships of Of/Çaykara (abbreviated ROf), Sürmene (abbreviated RSür), and Tonya (see Schreiber 2024). Romeyka has been in contact with Turkish varieties at least since the 17th century, although early contacts may date back till the 11th c. (Drettas 1997: 5–6) and it is difficult to assess the intensity of contact with Turkish throughout the Middle Ages and the Modern period. At least since intensified labour migration to larger cities in Turkey (as well as abroad) since the 1960s, the influence of Turkish has significantly increased, causing cultural assimilation and language shift at least in the urban speech communities (Schreiber & Sitaridou 2017). Today, the majority of Romeyka speakers are recessive bilinguals, with Turkish as dominant language, which affects the linguistic structure of Romeyka and facilitates contact-induced changes (Schreiber 2024).

The density of documentation and grammatical description of Romeyka is moderate, with increased interest in the field in the last years. After Mackridge's (1987) work on the Muslim Pontic Greek variety still spoken in Turkey, Sitaridou (2014b, 2013, 2014a, 2016, 2021) has contributed research mainly on the syntactic domain of Romeyka, followed by Neocleous' 2017 doctoral thesis on word order and information structure in Romeyka (Neocleous 2020, cf. Neocleous 2022). Recently, Schreiber (2024) has presented the first comprehensive grammatical description of Romeyka based on a naturalistic spoken language corpus. A considerably larger body of literature is available on Pontic Greek as spoken by Christian speakers in Turkey before the Greek-Turkish population exchange in 1923 (Deffner 1878, Parcharidis 1880, 1888) and in Greece after 1923 (Dawkins 1931, 1937, Papadopoulos 1933, 1955, 1958–1961, Tombaidis 1992, 1996, Drettas 1997, 1999, Revithiadou & Spyropoulos 2009) as well as on its closest relatives Cappadocian (Janse In press and references therein) and Pharasiot (Bağrıaçık 2018 and references therein).

The aim of the present chapter is to analyze the WOWA dataset of Romeyka (Schreiber 2021) with regard to the grammatical entities that occur in the postverbal domain and to word order in Romeyka in general. Since Romeyka is a shifting language under strong contact with Turkish (see Schreiber 2016 and Schreiber & Sitaridou 2017 for a sociolinguistic assessment of language vitality), which displays fundamentally different word orders, the data reveal a lot of variation which needs to be accounted for. Thereby, theoretically relevant topics from Greek linguistics such as word order properties of strong and weak pronouns and the question about a potential shift in Romeyka word order directionality will be

touched upon. Most importantly, it will be argued that basic word order properties (see Section 2 for a discussion of terminology) in Romeyka as a shifting language varies significantly between speakers and seems to be crucially determined by the individual multilingual profiles of the speakers.

The WOWA dataset of Romeyka has been compiled and coded according to the methodology of the WOWA corpus outlined in Haig et al. 2024 [this volume]; basic descriptive statistics have been carried out in R (logistic regression models). The data consist of five coherent texts produced by three speakers, extracted from the larger naturalistic spoken language corpus of Romeyka compiled by Schreiber (In prep).[2] The texts were recorded during fieldwork in Turkey in June/July 2019. Romeyka examples in the present chapter are referenced as follows: examples that stem from the WOWA dataset are referenced as Schreiber (2021: text ID, token ID), and examples from the Romeyka corpus by Schreiber (In prep) are referenced by the respective code in the corpus (as explained in Schreiber 2024).

In the following, Section 2 summarizes the present stage of research on word order in Romeyka. Section 3 sketches the general impact of information structure. In Section 4, the word order profile of Romeyka is characterized based on the WOWA data by focusing on different clause types (the NP in Section 4.1, the PP in Section 4.2), semantic and grammatical roles (Sections 4.3-4.5), auxiliaries in Section 4.6, and complex clauses in Section 4.7. Section 5 on areal aspects of language contact and Section 6 on the impact of recessive bilingualism discuss further important factors in the variability of Romeyka word order patterns.

2 Romeyka word order: Background and previous analyses

It is not straightforward to determine the present word order profile of Romeyka (a) due to a high pragmatically conditioned variability (see Section 3), and (b) due to ongoing language shift to Turkish and consequently high variation. It is generally agreed that Greek word order is determined by information structure. This was the case in Ancient Greek (van Emde Boas et al. 2019: 702–721), in Medieval Greek[3](Holton et al. 2019: 2022–2024) as well as in Standard Modern Greek (Holton et al. 2012: 518–520). As far as PG is concerned, it has been

---

[2]For an overview of speakers and data in the WOWA dataset of Romeyka, see Table 1 in Section 6. For a description of data collection and metadata see multicast.aspra.uni-bamberg.de/resources/wowa/data/hellenic/ponticgreek_romeyka/wowa_hell_ponticgreek_romeyka__metadata.pdf.

asserted that there is no basic worder (Drettas 1997: 277–280). On the other hand, Neocleous (2022, 2020), following Sitaridou (2016), investigated the diachronic development of Romeyka word order within the Minimalist framework and concluded that Romeyka is a "mixed directionality language" (Neocleous & Sitaridou 2022) with inherited VO order in pragmatically unmarked main clauses and OV order in subordinate clauses reinforced by contact with Turkish.

Neocleous (2022, 2020) claims that Romeyka has in main clauses an underlying VO word order, which is evident when all possible information-structural complications are set aside. Deviations from VO are explained through information-structural conditions holding in certain contexts (for a similar approach for Pharasiot, see Bağrıaçık 2018). In subordinate declarative clauses, unmarked word order is, according to Neocleous, OV with finite verbs and VO with infinitives. Although Neocleous seems not to distinguish between complement clauses preceding and following the matrix clause, most of his examples of subordinate clauses follow the matrix clause.

As has been addressed by Haig et al. (2024 [this volume]), it is by no means straightforward to determine unmarked or "basic" word order in a language and there are different accounts on how to establish this (see Dryer 2013 for a "rule of thumb" in determining dominant word order based on relative frequency, cf. also Dryer 1995). Neocleous (2022, 2020) defines a basic word order for both matrix and subordinate clauses in Romeyka based on the pragmatically unmarked word order, which he defines as an "'all-focus sentence', aka 'a presentational focus sentence', containing neither old information nor any presuppositions" (Neocleous 2020: 143). He elicits such clauses in response to the question 'What happened?', departing from the assumption that in response to this question all information is new and thus of equal discourse-pragmatic status.[4] However, departing from a naturalistic spoken language corpus of multilingual speakers as in Schreiber (2024), frequencies of word order patterns show a different picture, although certain limitations apply as well, such as the size of the corpus. While many observations of Neocleous are confirmed by the present data (e.g., the impact of information structure), the analysis cannot be adopted wholesale (see especially Section 6 on inter-speaker variation). Furthermore, the present analysis follows largely the WOWA approach (see Haig et al. 2024 [this volume])

---

[3]See also Horrocks (1990: 45) and Rafiyenko & Seržant (2020: 11) for Postclassical Greek. It should be noted that in later publications, Horrocks (2007) assumes VSO as the informationally most neutral word order in both Postclassical (Horrocks 2007: 623) and Late Medieval and Early Modern Greek (Horrocks 2019: 2022–2023).

[4]Certainly, Neocleous (2020) is not the first to use this approach; for a description of how neutral word order has been defined in the previous literature, see Bağrıaçık (2018: 146–151).

which is role-specific; the existing literature on word order in Romeyka does, to our knowledge, not include claims specific to Goals, Locations, etc. So, our prominent question should not be whether Greek had at any stage (Postclassical, Medieval, Modern) an unmarked VO word order, but rather whether OV has become more prominent in Romeyka — regardless of information structure — under contact influence from Turkish. Therefore, our analysis shall be based on frequencies of head-final orders. Note that this is still not straightforward to deal with, as Turkish allows for word order variation as well (Göksel & Kerslake 2005: 343–349), at least in informal spoken language — and certainly in the Trabzon Turkish dialect (see Schreiber Submitted).

3 The role of information structure

This section summarizes the role of information structure in Romeyka word order as proposed by Neocleous (2022, 2020), although the naturalistic corpus data do not always conform to these predictions and, for example, inter-speaker differences will need to be kept in mind (see also Janse & Schreiber In prep). Word order in Romeyka is largely determined by information structure, defined in terms of the concepts of topic and focus. A topic of a clause is defined here broadly as old/given information, that is, "an entity that has usually already been introduced into the discourse and is taken up again" (Bağrıaçık 2018: 114) and which is, if not already familiar to the hearer (or at least to the speaker), "agreed on by the speakers" (Soltic 2015: 48, following Gundel & Fretheim 2004). A constituent is in focus if it contains emphasized information which is generally assumed to be in this context new to the hearer (cf. "information focus" in Bağrıaçık 2018: 115). Both topic and focus can also yield contrastive information, thus called contrastive topic and contrastive focus. A contrastive topic is "an element that induces alternatives which have no impact on the focus value and creates oppositional pairs with respect to other topics" (Frascarelli & Hinterhölzl 2007: 88; cf. Bağrıaçık 2018: 267).

Neocleous (2020: 105) argues that there is a single subject position in Romeyka main and subordinate clauses and that all subjects (no differentiation regarding specificity) in pragmatically unmarked orders in Romeyka are left-dislocated topics,[5] appearing in the left-most clause position (1).

(1) Romeyka (Schreiber 2021: B, 0238)
   [*ena peðas*]$_{TOP}$ *ebidže*   *havus*
   a   boy   make.AOR.3SG pool
   'A boy made a pool.'

Both the pre-verbal and the post-verbal domain can host topics, but contrastive topics are only possible in pre-verbal position (example (2); Neocleous 2020: 186). Definite object topics, both pre-verbal (Neocleous 2020: 128) and post-verbal (Schreiber 2021), optionally trigger clitic doubling, that is, the coindexation of a (pro)nominal object topic by a clitic pronoun. Clitic doubling of definite object topics is an inherited Greek feature, which is obligatory in PG and other East Asia Minor Greek varieties such as Cappadocian and Pharasiot (Janse 2008), but not systematically attested in the Romeyka corpus. In PG (Drettas 1997: 276–280), clitic doubling occurs for nominal object topics in left-dislocated position (but not when the topicalizer *=ba(l)* occurs) by means of a referential resumptive pronoun to distinguish a topic from pre-verbal focus since focalization never triggers clitic doubling.[6]

Romeyka displays, like PG, an (arguably)[7] clitic topicalization particle *=ba(l)* (3a/b), which according to Neocleous (2020: 120) assigns contrastive topic to the marked constituent, although the manifold functions of this particle outlined in Schreiber (2024: 141–143) require more detailed investigation. The focus position in Romeyka (for both information and contrastive focus) is immediately to the left of the verb (4) and can be filled by several constituents (Neocleous 2020: 129); multiple focus is possible resulting in movement of all focused constituents to pre-verbal position (Neocleous 2020: 181).

In sum, according to Neocleous (2020, 2022), the pragmatically neutral basic word order in Romeyka main clauses is (S)VO, as illustrated in (1) above; if OV order occurs in main clauses, this is argued to be due to either focalization or topicalization.

---

[5]We are aware that the term left-dislocation is associated with specific mechanisms in the generative literature (Kaltsa & Sitaridou 2010, Neocleous 2020, 2022), but since we are not working within a generative framework, we prefer to use "topicalization" for preverbal topics which are still within the clause and "topic left-dislocation" for preverbal topics which are prosodically detached from the nuclear clause and constitute a separate intonation unit (Janse 2008), "backgrounding" for postverbal topics within the clause and "topic right-dislocation" for postverbal topics which constitute a separate intonation unit (in accordance with Janse 2008: 167–168).

[6]Compare Horrocks' 2019 rule [92] for the obligatory co-occurrence of clitic-doubling with dislocated and non-dislocated preverbal topics versus rule [93] for the absence of clitic-doubling with preverbal foci in Late Medieval and Early Modern Greek (Horrocks 2019: 2024–2025).

[7]On the enclitic status of *pa(l)* in PG, see Papadopoulos (1955: 119, 1958–1961: volume 2: 138), Setatos (1994), Drettas (1997: 46, 434), Janse (2002: 225–226), Ralli (2006: 131–132), Kaltsa & Sitaridou (2010: 263). Compare the use of *πάλιν* as an enclitic topic marker in Late Medieval Greek (Soltic 2013).

(2) Romeyka (Schreiber 2024: 247, ex. 75a, constructed example)
[avudo to saχan]ₜₒₚᵢ epero   *toᵢ*
this   the plate   take.PRS.1SG OPN.CL.3SG
'I take this plate.'

(3) a. Romeyka (Schreiber 2021: D, 0390)
      ulin efteme
      all   make.PRS.1PL
      'We make everything.'

    b. Romeyka (Schreiber 2021: D, 0391)
      pikniki ba efteme
      picnic TOP make.PRS.1PL
      'We make a picnic.'

(4) Romeyka (Schreiber 2021: B, 0243)
[ta is]_foc   *ebidže*
the footprints make.AOR.3SG
'He left footprints.'

4 Word order profile

4.1 Word order in the nominal phrase (NP)

Word order in the NP is head-final in Romeyka and nominal modification is pre-nominal with the exception of enclitic genitive pronouns. Attributive adjectives precede the head noun (5),[8] as do demonstratives (6) and numerals (7). Definite NPs trigger determiner spreading on attributive adjectives and in principle also on numerals,[9] that is, the definite article occurs before each modifying element as well as before the head noun (e.g., examples 6, 8, 10; for details on determiner spreading and nominal agreement in Romeyka see Schreiber 2024).

---

[8]While we apply in this chapter generally a very simplified glossing system that ignores some morphological information (and we also abstained from indicating word accent in the examples), we indicate case information only in Section 4.1 on NP word order, but ignore nominal number and gender. It has to be noted that case flagging vs. bare marking, as it is differentiated in the WOWA coding strategy, cannot be considered a reliable factor in Romeyka, since nominative and accusative(/oblique) case endings are often reduced and it can be partly only inferred from the syntactic context which case is expected in a certain example. In the coding of the WOWA dataset, only those tokens are tagged as 'case' which show a clear case ending, like MASC.SG.NOM. *-os* and MASC.SG.ACC. *-on*.

[9]Although the Romeyka corpus shows several deviations (see Schreiber 2024).

(5) Romeyka (Schreiber 2021: A, 0042)
     geniše ðromo   utš en
     broad road.NOM NEG be.PRS.3SG
     'There is no broad road.'

(6) Romeyka (Schreiber In prep: 08_04072019M_3; 161)
     hatšino d omorfo don   dobo
     this.ACC the nice.ACC the.ACC place.ACC
     'this nice place'

(7) Romeyka (Schreiber 2021: C, 0285)
     ðio nomade
     two persons.NOM
     'two people'

(8) Romeyka (Schreiber In prep: 04_01072019F_2; 087)
     mo   ta   dœrt tane ta   za
     with the.ACC four piece the.ACC cows.ACC
     'with the four cows'

Romeyka has pre-nominal nominal genitives (both nouns and NPs), i.e., the possessor precedes the possessed (9, 10).

(9) Romeyka (Schreiber 2021: C, 0303)
     tu   spid   i   arθob
     the.GEN house.GEN the.NOM people.NOM
     'the people of the house'

(10) Romeyka (Schreiber In prep: 08_04072019M_2; 090–092)
     du   dünja   olon   da   tehlikelija   da   dobe
     the.GEN world.GEN all.GEN the.NOM dangerous.NOM the.NOM places.NOM
     'the world's most dangerous places'

As for pronominal possession, Romeyka has pre-nominal full possessive pronouns (11) but also post-nominal weak/enclitic possessive(/genitive) pronouns (12).[10] As with other pre-nominal modifiers in definite NPs, the head noun keeps its definite article when combined with a full (pre-nominal) possessive pronoun (11), which historically includes an incorporated definite article as well.

(11) Romeyka (Schreiber 2021: A, 0129)
temetero do barχari
POSS.1PL the pasture
'our pasture'

(12) Romeyka (Schreiber 2021: E, 0564)
andras   ades
husband.NOM POSS.CL.3SG
'her husband'

Relative clauses in Romeyka are in principle pre-nominal (13), although post-nominal relative clauses exist as well (14). It is argued that pre-nominal relative clauses in Romeyka have evolved under contact influence from Turkish, while post-nominal relative clauses are a Hellenic relic (Neocleous 2020; Schreiber 2024).

(13) Romeyka (Gandon 2016: 222, ex. 517, glosses modified)
opse   iða   [alis   p   epiren]   ineka
yesterday see.AOR.1SG Alis.NOM REL take.AOR.3SG woman.ACC
'Yesterday I saw the woman who Ali married.'

(14) Romeyka (Neocleous 2020: 71, ex. 87, presentation/glosses modified)
o   peðas [op erθen   aso   cicenin] temon t   anepsin
the child REL come.AOR.3SG from.the grocery my   the nephew
en
be.PRS.3SG
'The child who came from the grocery's is my nephew.'

4.2 Adpositional phrases

Romeyka is a prepositional language. Prepositional phrases express in Romeyka semantic roles of location, Goal, source/origin, instrument or benefactive, as well as some temporal (and other) adjuncts. However, these semantic categories are not exclusively flagged with prepositions. The overall rate of prepositional marking in the WOWA dataset for Romeyka is 61% (percentage calculated based on

---

[10]Throughout this chapter, we indicate clitic status only for weak possessive and object pronouns, and refrain from doing so for all other parts of speech that are traditionally considered clitics in Greek linguistics, such as definite articles, prepositions, relativizer, auxiliaries, negation and modal particles, and the coordinating conjunction *tše* 'and'.

the occurrence of prepositions in the functions ablative, addressee, benefactive, comitative, (caused) Goal, instrumental, locative, recipient and beneficiary). For differences in the frequency with which adpositional phrases occur after the verb, see the sections on the respective semantic roles below.

Finally, it must be noted that Romeyka has a complex system of spatial orientation (see Schreiber 2024) and some spatial adverbs could be potentially considered circumpositions (see also Karatsareas 2016 work on circumpositions in Cappadocian), if they co-occur with a preposition, although their status as bound elements is not clear (examples 15, 16 vs. 17).

(15)   Romeyka (Schreiber In prep: 05_03072019M_3; 29)
       s   ena sergi   eban
       on a   blanket above
       'on (top of) a blanket'

(16)   Romeyka (Schreiber In prep: 01_28062019F_3; 42–43)
       s   oros   apes-merea
       at.the forest inside-somewhere
       'inside the forest'

(17)   Romeyka (Schreiber In prep: 01_04022016F_1; 052)
       eb-ebuka   asi   ɣorɣoran
       from-under from.the Gorgoras.ACC
       'from lower Gorgoras' [as differentiated from upper Gorgoras]

4.3 Ordering of spatial expressions relative to the verb

4.3.1 Locations

The preposition indicating location (as well as Goal/direction) is *s* 'to, at', which can merge with determiners and object pronouns, e.g., *s* 'to, at' + *to*.DET.ACC.SG > *so* 'to the' (18), *s* 'to, at' + *emasuna*.OPN.1PL > *semasuna* 'to us'. The overall frequency of post-verbal locations (including PPs and spatial adverbials) is 42%, as in (19), the majority of locations is pre-verbal. There is no significant statistical correlation found between the position and the independent variables animacy, weight or flagging (i.e., any overt phonological marking of case). However, there is some inter-speaker variation to be observed (see Table 3 in Section 6 below). The position of spatial adjuncts is sensitive to discourse. The immediate pre-verbal position is argued to be the (information) focus position (Neocleous 2020: 132, 148). In this vein, as exemplified in (19), multiple dislocation of (place) constituents is possible in Romeyka, as is the case for SMG (Alexiadou 1997: 58).

(18) Romeyka (Schreiber 2021: A, 0187)
*eskiden gœl*_{FOC} *ebiname   so   bodami*_{TOP}
formerly lake   make.IPF.1PL at.the valley
'In earlier times, we made a lake at the valley.'

(19) Romeyka (Schreiber 2021: A, 0001)
*emistine*_{TOP} [*aða sin   otšena*]_{FOC} *jašaevume*
we   here at.the Ogene   live.PRS.1PL
'We live here at Ogene.'

4.3.2 Goals

For the discussion of goals in Romeyka, only goals of motion verbs and verbs of caused motion are considered. Goals in the WOWA dataset are predominantly expressed by prepositional phrases headed by the prepositions *s* 'to' and *os* 'until', but also adverbially. 78% of goals (both prepositional and adverbial) are post-verbal (20), which shows that goals are much more likely to be post-verbal than locations. While information structure affects word order variation in spatial adjuncts in Romeyka, this is not likely to be the only factor determining word order in light of the significantly higher number of post-verbal Goals; see Rasekh-Mahand et al. 2024 [this volume] on spoken Persian for a similar observation. However, in light of an assumed VO/VX order in declarative clauses in Romeyka (Neocleous 2020) — as in other dialects of modern Greek — it is rather the high number of pre-verbal locations and sources (see Sections 4.3.1 and 4.3.3) that is striking and requires explanation. Comparing (20) with (21), information structure accounts for the pre-verbal position of the PP in (21). Otherwise, no statistically significant correlation is found between position and the dependent variables animacy, weight, and flagging, although adverbial goals have a higher likelihood to be post-verbal than PPs. Finally, it has to be noted that there is high inter-speaker variation (see Table 4 in Section 6 below).

(20) Romeyka (Schreiber 2021: A, 0175)
pao   so   raši
go.PRS.1SG to.the mountain
'I go to the mountain.'

(21) Romeyka (Schreiber 2021: A, 0136)
[*son   barχari muna*]_{TOP} [*direk araba*]_{FOC} *bai*
to.the pasture POSS.CL.1PL direct car   go.PRS.3SG
'A bus goes directly to our pasture.'

As for caused goals in the WOWA dataset, 100% are post-verbal (22), but note that due to very low token numbers (N=4) in the generally small WOWA dataset for Romeyka this information is only tentative and needs further investigation.

(22) Romeyka (Schreiber 2021: D, 0488)
      eferenam   son   barχari ksila
      bring.IPF.1PL to.the pasture wood
      'We brought wood to the pasture.'

4.3.3 Sources

The majority of sources in the WOWA dataset, i.e., 80%, are pre-verbal, including PPs with the preposition *as* 'from' and adverbials. This result needs to be taken with some caution, though, since nearly all examples stem from a single speaker. Information structure is likely to account for the position of the source, cf. ex. (23a) in assumed historically unmarked VX position vs. (23b) in pre-verbal contrastive focus position where the PP contrasts with the PP in (23a), i.e., the previous sentence in the same recording.

(23) a. Romeyka (Schreiber 2021: A, 0024–0025)
          bazen   para berename edroγame asa   bakale
          sometimes money take.IPF.1PL eat.IPF.1PL from.the shops
          'Sometimes we took money, we ate [food] from the shops.'
      b. Romeyka (Schreiber 2021: A, 0026)
          bazen   aso   spidi eberenam tš ebejname
          sometimes from.the house take.IPF.1PL and go.IPF.1PL
          'Sometimes we took [food] from the house.'

4.4 Ordering of direct objects relative to the verb

4.4.1 Nominal direct object

This section focuses on direct objects (DOs) of transitive verbs. In the WOWA dataset, 66% of nominal direct objects are post-verbal (arguments of 'have'/existentials, which are predominantly pre-verbal, not included in this count). The percentage appears to be the same for definite and indefinite nominal DOs, which suggests that definiteness does not play a role. Note, however, that in Cappadocian, word order is sensitive to definiteness, that is, indefinite object NPs tend to occur in post-verbal position (Janse 2006). The difference in word order between examples (24) and (25) is probably rather than in definiteness to

be found in information structure, with the object NP in (24) in focus position and the object NP in (25) in topic position. However, there may be also an inter-ference from the frequent occurrence of post-verbal nominal object NPs in Trab-zon Turkish (Schreiber Submitted), a phenomenon described for informal spoken Standard Turkish as "backgrounding", and which applies usually to definite NPs, although a non-definite NP can be placed in spoken Turkish in the post-verbal position "if it refers to an entity or category that has been mentioned (or implied) in the immediately preceding discourse" (Göksel & Kerslake 2005: 346).

Numerically, more complex NPs/PPs in the WOWA dataset, i.e., those with more than two words usually involving some kind of nominal modification or genitives, tend to be pre-verbal (26). Although this is contrary to the traditional assumption that very complex NPs/PPs preferably appear at the end of an utter-ance (see, e.g., Behaghel 1909 departing from German), it does align with similar findings from the spoken-language corpora investigated in this volume (see, e.g., Rasekh-Mahand et al. 2024 [this volume] and Leitner 2024 [this volume]). How-ever, it is likely that information structure is the more decisive factor, which might just happen to overlap with the factor of "weight" (see, e.g., 26).

A significant correlation was found between the dependent variable position and the independent variable animacy: human and animate nominal DOs are slightly more likely to be pre-verbal than inanimate nominal DOs. Interestingly, a reverse significant effect was found for pronominal human and animate DOs, which tend to be more likely post-verbal than inanimate pronominal DOs (see Section 4.4.2). As is the case for all semantic categories, the order of nominal DOs with regard to the verb is sensitive to information structure (see Section 3) and inter-speaker variation is relevant here as well (see Table 2 in Section 6). Finally, it also needs to be stated that the linear order of nominal objects with regard to the verb seems to be at times within the same speaker in free variation, even if other variables remain stable (e.g., 27 vs. 28), i.e., there is intra-speaker variation. This needs to be most likely attributed to ongoing language shift under strong influence from Turkish OV orders, as information structure cannot always be convincingly invoked as explanation for all OV orders.

(24) Romeyka (Schreiber 2021: E, 0564)
       andras ades   ba ksila gofdi
       husband POSS.CL.3SG TOP wood cut.PRS.3SG
       'Her husband cuts wood.'

(25) Romeyka (Schreiber 2021: E, 0572)
   efidže   so   ðormo ban   ena ksilo
   leave.AOR.3SG at.the road   above a   wood
   'He lost a log on the street.'

(26) Romeyka (Schreiber 2021: A, 0144−0145)
   ifedi   eɣo ebiɣa   išde muskaræ utš   dane muskar ebiɣa
   last_year I   buy.AOR.1SG DP calves   three piece calve   buy.AOR.1SG
   'Last year I bought calves. I bought three calves.'

(27) Romeyka (Schreiber 2021: A, 0073)
   efteme   fasulijas
   do.PRS.1PL beans
   'We do beans.' [i.e., 'We grow beans.']

(28) Romeyka (Schreiber 2021: A, 0099)
   opsar ebsame
   fish   catch.AOR.1PL
   'We fished.'

4.4.2 Pronominal direct objects

The percentage of post-verbal pronominal direct objects in the WOWA dataset
is with 58% slightly lower than that of nominal DOs. Importantly, (en)clitic ob-
ject pronouns are not coded in the corpus since they are bound. Romeyka has
both clitic object pronouns following the predicate and free object pronouns pre-
ceding the predicate. The two are mainly differentiated based on the criterion of
stress, whereby enclitic object pronouns have a reduced phonological form and
do not impact the stress pattern of the verb. However, in the Romeyka corpus
(Schreiber In prep) there appear to be also post-verbal full object pronouns, which
are sometimes difficult to differentiate from weak ones and which are probably
reinforced by contact with Turkish (see below; see also Schreiber 2024: 103). In
the WOWA dataset, the percentage of 58% post-predicate pronominal DOs only
refers to full pronominal forms, the vast majority being third person pronouns.
While none of the coded categories like weight or flagging is significant for the
position of pronominal DOs, there is a significant tendency for pronominal DOs
denoting human or animate entities to appear post-verbally as opposed to inani-
mate pronominal DOs (29 vs. 30). However, it is obvious from the examples, that
information structure plays a significant role here as well (see 29). Still, it is un-
clear whether 42% of pre-verbal pronominal DOs are largely due to information

structure since pronouns are usually given information which can occur in pre- and post-verbal position (see Section 3).

(29) Romeyka (Schreiber 2021: D, 0408–0409)
laɣo dune   ado utš eksero
how be.IPF.3SG this NEG know.PRS.1SG
'How it was? I don't know this.'

(30) Romeyka (Schreiber 2021: B, 0246)
utš iðan   adona
NEG see.AOR.3PL OPN.3SG
'They did not see him [=the bear].'

An important issue is the post-verbal placement of full pronominal objects (often third person pronouns) versus weak enclitic object pronouns. In the WOWA dataset, 60% of pronominal DOs are weak enclitic pronouns (thus not coded for position), 40% are full pronominal forms out of which 58% are post-verbal. In comparison, in the larger Romeyka corpus, 63% of all pronominal DOs are clitic pronouns (Schreiber 2024: 100, Table 12). In (30), a weak object pronoun would be likely to occur since 'the bear' is a topic which has been mentioned several times in the preceding context. Instead, the full pronominal form is placed in post-predicate (enclitic) position. As another example, in (31), the full pronominal form in post-verbal position is preferred over the weak pronominal form *=(a)ta*. It is possible that a contact explanation can account for the preference of post-posed full object pronouns in Romeyka, as (Trabzon) Turkish has no clitic object pronouns and object pronouns can appear post-verbally in informal spoken Turkish (also cf. only 13% post-verbal pronominal DOs in Hodgson 2024). According to Brendemoen (2005: 30), post-verbal pronominal DOs in Trabzon Turkish have arisen due to contact with PG. The potential mutual influence suggests a convergence-type of change between Romeyka and Trabzon Turkish.

(31) Romeyka (Schreiber 2021: B, 0234)
ama utš eboresane   dosin   adonusine
but NEG can.AOR.3PL give.INF OPN.3PL
'But they could not hit them.'

In sum, VO is assumed to be the unmarked order of pronominal DO and verb in Romeyka, whereby OV orders are motivated by information structure. The remaining question is what motivates the post-verbal position of full object pronouns. In terms of information structure, pronominal VO with full object pronouns can only be explained, if their information structural value is comparable

to that of weak clitic pronouns, that is "familiar topics". If the post-verbal position of full pronominal DOs in the Romeyka corpus is split up according to person and number, it is evident that third person singular full object pronouns are most likely to be post-verbal, although strikingly third person plural full object pronouns have the least chance to be post-predicate (unlike ex. 31).[11] So probably, gender could play a role in the sense that gender cannot be differentiated in weak third person singular object pronouns but does reflect in strong third person object pronouns. Furthermore, it is not clear whether phonological weight could also play a role, although 3SG strong pronominal forms are not significantly shorter (i.e., more "clitic-like") than 3PL object pronouns.

4.5 Ordering of other obliques relative to the verb

Other obliques refer to the semantic roles of recipients, addressees, comitatives, instruments, benefactives, and others. In the WOWA dataset, these categories do not figure prominently, so any quantitative analysis is pointless. Instead, the present section discusses some examples for each semantic category.

The WOWA dataset contains only two tokens for recipients, both of which are pronominal. Both pronominal recipients occur pre-verbally (32; but cf. 35). In general, the unmarked word order for nominal recipients is VO (33) with OV orders triggered by information structure, namely focus (34; see Schreiber 2024: 249–250). In ditransitive constructions, the unmarked word order is V–IO–DO (Schreiber 2024: 249). This applies also to pronominal recipients (see (35) and unlike (32) with a topicalized object pronoun).

(32) Romeyka (Schreiber 2021: A, 0182)
emenan ndona na   ðiɣune
OPN.1SG what   PRT give.PRS.3PL
'What do they give to me?'

(33) Romeyka (Schreiber 2024: 250, ex. 93, questionnaire data)
ta mila ðokan   ti mana   tuna
the apples give.AOR.3PL the mother POSS.CL.3PL
'They gave the apples to their mother.'

---

[11]Distribution of bound (weak) pronominal object pronouns in the Romeyka corpus (Schreiber In prep) per person and number: 1SG: 41 bound out of 67, i.e., 61% bound; 2SG: 18 out of 28, i.e., 64% bound; 3SG: 72 out of 133; i.e., 54% bound; 1PL: 17 out of 26, i.e., 65% bound; 2PL: 6 out of 10, i.e., 60% bound; 3PL: 63 out of 79, i.e., 80% bound.

(34) Romeyka (Schreiber 2024: 250, ex. 96, questionnaire data)
din batsi eðotše   ena ido ..sturatši
the girl   give.AOR.3SG a   DEM ..stick
'He gave a stick to the girl.'

(35) Romeyka (Schreiber 2024: 251, ex. 102, questionnaire data)
etšine bal ðotš   emena   milo
he   TOP give.AOR.3SG OPN.1SG apple
'He gave me an apple.'

In the WOWA dataset, all three tokens of addressees are pronominal and appear post-verbally (36, the stress pattern *ípen ádona* reveals that the pronoun must be indeed the post-posed strong form and not a clitic). In general, like with other obliques, nominal addressees are expected to follow the verb in unmarked word order, although they can move to pre-verbal focus position (note that the Romeyka corpus does not feature an example of a nominal addressee, but cf. ex. (37) with *ta patsiðes* 'the girls' as contrastive topic). Pronominal addressees, seem to occur nearly in all cases (of the Romeyka corpus) post-verbally.

(36) Romeyka (Schreiber 2021: E, 0582)
iben   adona
say.AOR.3SG OPN.3SG
'He told her […].'

(37) Romeyka (Schreiber In prep: 02_02022015F_1; 073–074)
eleɣane   ištera ta peðia kopela ta patsiðes eleɣane   ɣospiðes
say.IPF.3PL later the boys girl   the girls   say.IPF.3PL prostitutes
'Then they said to the boys girl, to the girls they said prostitutes.'

Comitatives, i.e., referents denoting accompanying persons (or at least animate entities), occur in the WOWA dataset predominantly in post-verbal position (38a), although pre-verbal placement is possible as well (see the PP as pre-posed given topic in (38b), although stylistic variation seems to play a role here as well).

(38) a. Romeyka (Schreiber 2021: A, 0060–0061)
jaja   ebejnane me ta za
on_foot go.IPF.3PL with the cows

 b. *me ta za jaja   ebejnane*
with the cows on_foot go.IPF.3PL
'They went on foot with the cows.'

Instruments appear in the WOWA dataset in the vast majority pre-verbally, which may be explained by focalization of new information (39) or contrastive focus. However, based on the data of the Romeyka corpus, it seems that instruments do not differ from other obliques in terms of unmarked VO order (40). Ex. (41) shows that for information structural reasons both nominal obliques and objects can occur pre-verbally, i.e., Romeyka allows for multiple focus, which is argued to be "order-preserving" (Neocleous 2020: 181–182). However, then it is not clear why the oblique precedes the DO in (29).

(39) Romeyka (Schreiber 2021: E, 0570)
  mo d aksinari ešgise da ksila
  with the axe split.AOR.3SG the woods
  'He split the wood with the axe.'

(40) Romeyka (Schreiber In prep: 09_04072019_7; 11)
  ekoftame me ti kerenti
  cut.IPF.1PL with the scythe
  'We cut with the scythe.'

(41) Romeyka (Schreiber 2021: A, 0036)
  ula me ta rašes ta yomare ekovalename
  all with the pannier the loads carry.IPF.1PL
  'We always carried the loads with the panniers.'

Benefactives refer to situations where X does something in the interest of Y, generally implying that Y is a sentient being. Although no benefactives appear in the WOWA dataset, they are realized in Romeyka by means of the preposition *ja(t)* 'for' (42). Their position seems not to differ from that of comitatives, which are prepositional phrases as well, and which appear often in the pre-verbal domain for information structural reasons, although their unmarked word order is post-verbal.

(42) Romeyka (Schreiber 2024: 147, ex. 565, questionnaire data)
  sade jad emena faji utš eθelisa pseθinimo
  only for OPN.1SG food NEG want.AOR.1SG cook.NMZ
  'I did not want to cook just for myself.'

Among other obliques in the WOWA dataset, i.e., those which are none of the semantic roles above, 57% are in post-verbal position, which equals roughly the overall percentage of post-verbal placement in the WOWA dataset which is 55%. Since nearly all of these other obliques are prepositional phrases, they follow the constraints outlined for other PPs above.

4.6 Auxiliaries

Neocleous & Sitaridou (2022) discuss the following auxiliaries in Romeyka: (i) *ime* (+ particle), (ii) *iχa* (+ infinitive), (iii) *eš(i)* (+ finite verb in present tense or imperfective past). They always precede the main verb, irrespective of whether it is finite, infinitive or particle. Additionally, there is a periphrastic progressive construction with *steko/stekome* 'stand' or *kahome* 'sit.'

The invariable form *eš(i)* plus finite verb is used as a periphrastic progressive denoting processes that are close to completion and goes most likely back to the 3sg present tense form of the verb *eχo > eš(i)* (ex. 43; cf. Drettas 1997: 334 on a related progressive form in PG).[12] The inflected imperfective forms of the verb *eχo* 'have' are used as an auxiliary in the formation of counterfactual conditional clauses (44), which are formed by one of the modal particles *na/an/as* + inflected imperfective of *eχo* + non-finite verb/inflected infinitive (Schreiber 2024: 298, Table 37; see also Sitaridou 2014a).

(43) Romeyka (Schreiber 2021: B, 0241)
    eš erde   argo
    AUX come.AOR.3SG bear
    'The bear is/was? coming.'

(44) Romeyka (Schreiber In prep: 04_01072019F_13; 53)
    eγo na m iχa škisen   da[13]   *da ksila [...]*
    I   PRT NEG AUX split.INF? OPN.CL.3PL the woods
    'If I had not chopped the wood, [...].'

With the limited use of the auxiliary *iχa* (morphologically homonymous with imperfective 1sg form of *eχo* suggesting two functions of *eχo* as lexical verb and auxiliary), Romeyka stands out from other varieties of modern Greek like SMG. Its closest relative Cappadocian has a pluperfect with impersonal *iton/itan* (homonymous with the 3rd person aorist form of *ime*) preceded by a finite main verb. Diachronically AMG also uses the auxiliary *ime*, morphologically homonymous with the copula *ime* 'be' (i.e., *ime* can thus function both as existential and auxiliary). For the word order properties of copula and 'become' complements in Romeyka, see Section 4.7.2.

---

[12]According to M. Bagriacik (p.c.), the form *eš* goes probably back to a homonymous existential auxiliary already existing in older stages of Greek, rather than to the 'have'-auxiliary, as the former presents a more likely grammaticalization pathway for a progressive form.

[13]The analysis of the verb form *škis-en=da* is not clear, we analyse *=ta* here as a weak object pronoun but it also resembles the suffix *-ta* marking gerunds like *jelaχ-ta* 'laughing'. The form *škisen-* could be potentially a reduced infinitive, cf. *škisini*.INF.

In both cases where forms of *eχo* are used as an auxiliary in Romeyka, the auxiliary precedes the main verb, although importantly, in the counterfactual conditionals the verb following the auxiliary is not (i.e., in the case of the infinitive) or only partly (i.e., in the case of the "inflected infinitive", see Schreiber 2024: 229–233 going back to Sitaridou 2014a) inflected. Neocleous (2020: 269) confirms AuxV order in Romeyka main and subordinate clauses. The construction Vfin + *iton* in pluperfects which exists in Cappadocian is not attested for Romeyka.

4.7 Complementation

4.7.1 Complement and adjunct clauses

In Romeyka both finite and non-finite complementation exist. Depending on the type of predicate, a complementizer is used to introduce the complement clause (CC), while other clause types do not require a complementizer. In Schreiber (2024: 278), it has been argued that more than one complementation strategy is available for some clause types due to contact influence from Turkish, resulting in an increase of non-finite complementation strategies. While the non-finite strategy of using infinitives in some CCs is an archaic trait of Romeyka, non-finite deverbal nouns as a complementation strategy have increased under contact with Turkish. Within the finite complementation strategies, complementation by means of the complementizer *na* is more restricted in Romeyka compared to SMG, while juxtaposition with paratactic syntax and without complementizer is widespread (45), especially with verbs of saying and in (in)direct speech.

(45) Romeyka (Schreiber In prep: 02_2906019F_1; 02)
egusame [o jaja evren arkon]
hear.AOR.1PL the Yahya find.AOR.3SG bear
'We heard that Yahya has found a bear.'

In Romeyka CCs, the complement clause predominantly follows the matrix verb (45). For some predicate types, among which are verbs of saying, the reverse order is possible as well, see for example the preverbal headless relative clause in (46); even circum-positions exist (47).

(46) Romeyka (Schreiber In prep: 01_28062019F_3; 24)
[to leγo] utš eγrigo
what say.PRS.1SG NEG understand.PRS.1SG
'I don't understand what I say.'

(47)   Romeyka (Schreiber In prep: 08_04072019M_2; 066)
       [t   aleɣo] kseris   [dohna e]
       the horse know.PRS.2SG what   be.PRS.3SG
       'Do you know what "aleɣo" is?'

Na-clauses and infinitives obligatorily follow the main verb (48), (49). How-
ever, deverbal nouns calquing Turkish complement clauses appear before the
main verb but they are strictly speaking NPs and thus no actual CCs (50). Nomi-
nalizations selected by some aspectual verbs like *bašlaevo* 'start' which requires
a PP follow the main verb (51). If a complementizer is used, it appears at the
beginning of the CC.

(48)   Romeyka (Schreiber In prep: 02_02022015F_1; 014)
       utš eθelena   [n   andriza]
       NEG want.IPF.1SG PRT marry.IPF.1SG
       'I didn't want to marry.'

(49)   Romeyka (Schreiber 2024: 285, ex. 376, questionnaire data)
       utš eboresa   [tšimeθina]
       NEG could.AOR.1SG sleep.INF.1SG
       'I could not sleep.'

(50)   Romeyka (Schreiber 2021: A, 0193)
       [hab-aðadžega to   panimo] eɣo utš eɣabo
       from-here   the go.NMZ I   NEG like.PRS.1SG
       'I don't want to go from here.'

(51)   Romeyka (Schreiber In prep: 04_01072019F_13; 30)
       ebašlaepse   [so   borbatima]
       start.AOR.3SG at.the walking
       'She started to walk.'

Subordination of adjunct and relative clauses exhibits pre-dominantly head-
final syntax, although there is a lot of variation only partly dependent on the
clause type. Romeyka adjunct and relative clauses are predominantly finite, al-
though some non-finite strategies exist. The dependent clause is predominantly
pre-posed and generally introduced by pre-verbal adverbial subordinators or rela-
tive markers (ex. 52; for the syntax of relative clauses, see also Section 4.1. above).

(52) Romeyka (Schreiber In prep: 04_01072019F_13; 45)
      [omon d   eruise   s   ormi] ejendune   natsurula
      when REL fall.AOR.3SG to.the river become.IPF.3SG wet
      'When she fell into the river, she got wet.'

With regard to head-order directionality within the complement clause, according to Neocleous (2020: 118), the unmarked word order in subordinate/complement clauses is head-final (53) as a consequence of contact with Turkish head-final syntax. However, the example in (45) above seems to form a counter-example to this generalization, since the CC in (45) is a pragmatically neutral, information-structurally unmarked kind of statement (the object is new, not topical nor contrastive) and following Neocleous (2020), we would expect OV order here. Thus, both OV and VO orders seem to be possible in unmarked CCs; potential restrictions with regard to predicate types need further research.

(53) Romeyka (Schreiber 2021: C, 0332)
      eterezen   [o   argo erθen]
      look.AOR.3SG the bear come.AOR.3SG
      'He saw that the bear came.'

4.7.2 Copula and 'become'-complements

In the WOWA dataset, only 8% of copula complements appear after the copula, which means that the copula *ime* 'be' appears predominantly in clause-final position. Neocleous (2020: 117) also confirms that copula clauses with the 3rd person form *en/ine* are always head-final. However, when it comes to an explanation of the very low number of copula complements that appear post-verbally in the WOWA dataset (N=2), it is not straightforward to determine any decisive factor. In any case, predicate nominals (54) and predicate adjectives (55) seem to behave alike. Information structure can also account for some of the pre-predicate copula complements (56), while interrogative copula clauses are head-final due to the focus position of the *wh*-element (57). Finally, since the overwhelming majority of clause-final copulas occur in the text of a single speaker (see Table 5 in Section 6 below), the speaker variable may have an effect as well. Still, it is not clear what determines post-predicate copula complements as in (58), although information structure may serve as an explanation here as well.

(54) Romeyka (Schreiber 2021: A, 0128)
      ta mandria muna   boš   en
      the stables   POSS.CL.1PL empty be.PRS.3SG
      'Our stables are empty.'

(55) Romeyka (Schreiber 2021: A, 0020)
džumartesi tatili   en
saturday   holiday be.PRS.3SG
'Saturday is a free day.'

(56) a. Romeyka (Schreiber 2021: D, 0504–0506)
ama do   vutero poli
but   the butter much
'But the butter [was] plentiful.'

   b. *do   vutero bolin   adone*
the butter much be.IPF.3SG
'The butter was plentiful.'

   c. *eliɣo utš   en   do diri*
little NEG be.PRS.3SG the cheese
'The cheese was not scarce.'

(57) Romeyka (Schreiber 2021: D, 0408)
laɣo dune
how be.IPF.3SG
'How was it?'

(58) Romeyka (Schreiber 2021: A, 0049)
en   ja ta ðorma muna   patikas
be.PRS.3SG DP the roads POSS.CL.1PL unpaved
'Our roads are unpaved (as you know).'

Finally, it has to be noted that the copula — and especially its 3SG present tense form *en* — is often omitted (59, also 56a above; see also the last column in Table 5 in Section 6). Due to clause-final null-copulas in Turkish, it is compelling to assume a contact influence here, although an internal explanation may also play a role due to the particular nature of the verb *ime*. In Cappadocian, the full verb *ime* 'be' is used in existentials, where it is never dropped; besides, a copula *ime* 'be' exists which behaves as a clitic (=*me*) since MedGr times and thus results always in verb-final orders, although it is never left out in third person singular present tense.

(59) Romeyka (Schreiber In prep: 02_9062019F_1; 24)
etšinos χaremenos
he   happy
'He is happy.'

Interestingly, 'become'-complements in the WOWA dataset are 69% post-verbal, suggesting a different behaviour compared to copula clauses. However, it must be noted that nearly all 'become'-complements are produced by the same speaker (i.e., Speaker 2, see Table 1 in Section 6), so this result is to be treated with caution. Although the number of coded tokens is too small for any statistical analysis, there is in the dataset a tendency for adjectives as 'become'-complements to be more likely post-predicate than nominal complements (but cf. *ejendune gedže* 'it became night' (Schreiber 2021: B, 0262; E, 0529), also *inete akšemis* 'it became evening' (Schreiber 2021: D, 0437) vs. *gedže ejendune* (Schreiber 2021: B, 0242)). Again, information structure has an influence here; see (60) vs. (61) in topic position.

(60) Romeyka (Schreiber 2021: D, 0445)
      ula inumunesten   annera
      all become.IPF.1PL soakingly_wet
      'We became allover soakingly wet.'

(61) Romeyka (Schreiber 2021: A, 0210)
      emeklis ba   na inese   na stetšis   aðatšeka
      retired TOP PRT become.PRS.2SG PRT stay.PRS.2SG here
      'When you become retired, you will stay here.'

5 Areal issues & language contact

Similarly to the well-documented influences of Anatolian Turkish on AMG described by Dawkins (1916), who inspired much of the literature on language contact, long-standing language contact with Turkish in the Trabzon area has evidently led to contact-induced changes in Romeyka in several domains of the language including the lexicon and grammar (for a tentative overview see Chapter 6 in Schreiber 2024; see also Brendemoen 1998, 1999, 2002, 2005, 2006, 2010, 2019 for Greek influences on Turkish). Importantly, similarly strong influences of Turkish are reported from Laz (a Kartvelian language, see Kutscher 2008; Lacroix 2019, 2009, Öztürk & Pöchtrager 2011) and Homshetsma/Hemshinli (Western Armenian; Vaux 2007) which are both minority languages in the neighbouring provinces of Rize and Artvin in northeastern Turkey and which share a sociolinguistically and historically similar contact setting with Turkish as dominant language of the area. Although a comparison of contact-induced influences from Turkish on the morphosyntax of Laz and Hemshinli is beyond the scope of this chapter, Neocleous (2020: 282–284) reports a clear restructuring of Laz

word order patterns based on Turkish to pragmatically unmarked word order OV (see also the extremely low percentage of post-predicate elements in the WOWA dataset of Arhavi Laz, Stilo & Lacroix 2021) and unlike its Kartvelian relative Georgian (Neocleous 2020: 243–247). As far as what has been deducible for Hemshinli, it also shares similarities with Turkish, for example in non-finite subordination which is otherwise less common in related varieties outside the specific contact setting (Gandon 2016: 210–212). Finally, comparing the word order properties of PG as spoken in Armenia (Hodgson 2024) with that of Romeyka, Armenian PG has only an overall score of 32% post-posed elements, as opposed to 55% in Romeyka.

Another aspect of language contact in the area is the contact influence PG exerted on the Turkish Eastern Black Sea dialect in several domains including word order (Brendemoen 1998, 1999, 2005, 2006, 2019). For example, post-verbal pronominal DOs in Trabzon Turkish (62) are argued by Brendemoen (2005: 30) to arise due to contact with PG (cf. the percentage of post-verbal pronominal DOs in colloquial spoken Turkish of Ankara (Iefremenko 2021) at 7%). When considering the areal picture of contact-induced changes and convergence, the influence of indigenous minority languages on the majority language should not be neglected, although this seems to apply to a lesser degree to the influence of Armenian (and Laz?) on regional Trabzon Turkish (Brendemoen 2005: 29). In any case, potential mutual contact influences highlight the point that for the purpose of inter-language comparison, not primarily the respective standard variety (such as Istanbul Turkish) should be considered, but rather regional varieties (as is done by Neocleous 2020).

(62)   Trabzon Turkish (Brendemoen 2005: 30, ex. 2, presentation adapted)
    yedi   *oni*
    eat.PST.3SG OPN.3SG
    'He ate it.'

6 The role of inter-speaker variation

In order to explain (some of) the variability in word order patterns in Romeyka that appear in the quantitative data and can only partly be explained by linguistic factors, it is crucial to consider the nature of Romeyka which is currently spoken in Turkey as shifting variety, and the composition of the present sample. As Schreiber (2024) has shown, the character of Romeyka as shifting language is not (yet) characterized by language attrition as defined by Thomason (2001: 12)

as including structural simplification and loss without compensation but rather with a high inter- (and intra-)speaker variation. Indeed, it is the idiolectal variation based on the individual multilingual profiles of the speakers which explains some of the variability in word order patterns. In other words, the overall figure of 55% post-verbal placement (including 66% post-verbal placement of nominal DOs) in the WOWA dataset of Romeyka does not reflect a stable norm in "the speech community", but a mean value aggregating over very different individuals. As Craevschi (2022) has shown, the Romeyka data set exhibits the greatest degree of inter-speaker variation among the twenty-four WOWA data sets analysed by him. In fact, the influence of the independent variable "Speaker" outweighs significantly the influence of all other variables which were controlled for in the analysis, even when the imbalanced contribution of the three speakers to the overall data set is taken into consideration (Speaker 3 only contributes around 10% of the total tokens). This is shown in Table 1.

Table 1: Overview of speakers and their corresponding texts in the WOWA dataset

Speaker	Speaker characteristics	Texts provided	Total tokens
1	male, middle-aged, ROf, Karaçam	A	198
2	female, middle-aged, ROf, Karaçam	B, D, E	251
3	male, middle-aged, RSür, Beşköy	C	52

The influence of the speaker variable is not the same for all constituent types, and due to the low absolute numbers of tokens for some constituent types, cannot be readily statistically validated for all roles. It does turn out to be significant for predicting the placement of nominal DOs (see also Craevschi 2022). As shown in Table 2, Speaker 1 has with nominal DOs predominantly OV, while Speaker 2 has dominantly OV order. Speaker 3 appears to be largely balanced but note the smaller absolute number of tokens in the data of Speaker 3. A similar inter-

Table 2: Percentage of post-verbal nominal direct objects per speaker

	Total nominal DOs ('do'+'do-def')	Total VO	% VO
Speaker 1 (=text A)	51	18	35%
Speaker 2 (=text B, D, E)	103	86	83%
Speaker 3 (=text C)	21	12	57%

speaker difference is visible for free pronominal DOs, although the percentages for post-verbal placement are there lower in general (see also Section 4.4.2).

A similar picture of inter-speaker variation arises for the semantic roles of locations (Table 3) and goals (Table 4). While Speaker 1 shows predominantly pre-verbal locations (and more or less balanced pre- and post-verbal Goals), Speaker 2 uses predominantly post-verbal locations and especially goals. Speaker 3 shows clearly preverbal locations — other than with nominal DOs — although the amount of data provided by Speaker 3 is too little to get a clear picture.

Table 3: Percentage of post-verbal locations per speaker

	Total locations	Total VX	%VX
Speaker 1 (=text A)	32	11	34%
Speaker 2 (=text B, D, E)	35	23	66%
Speaker 3 (=text C)	16	1	6%

Table 4: Percentage of post-verbal Goals per speaker

	Total goals (no pronouns)	Total VX	%VX
Speaker 1 (=text A)	32	17	53%
Speaker 2 (=text B, D, E)	44	43	98%
Speaker 3 (=text C)	1	1	nc

Ignoring for a moment Speaker 3 (due to low absolute token numbers), it is evident that Speakers 1 and 2 differ rather consistently: Speaker 2 postposes these three argument types approximately twice as often as Speaker 1. With regard to nominal direct objects, locations, and goals, then, an account of word order purely in terms of language-internal features such as information structure, animacy, or weight, is clearly missing a very important dimension of variation.

To account for the inter-speaker variation just described, one would need to consider the individual multilingual profiles and language competences of the speakers. Speaker 1 and 2 stem from the same municipality, which is located remotely and is reported to have high language vitality (Schreiber 2016); both speakers are roughly of the same age group. Still, Speaker 1 shows more head-final syntax than Speaker 2. This can be explained by interference from Turkish head-final word orders due to a multilingual profile where Turkish is (at least at

the time of data collection) the dominant language. Gender is likely to account for the dominance of Turkish in the multilingual repertoire of (male) Speaker 1 as opposed to (female) Speaker 2. Usually, men are more exposed to Turkish than woman due to higher mobility for various reasons (like labour and military service, but also differences in education, see Schreiber 2016). However, it should be noted that individual language biographies can easily override this gender bias. Since unfortunately no detailed biographical information is available for Speaker 1 and 2, gender is initially assumed here as the decisive variable causing the different word order patterns.

As for the individual multilingual profile of (male) Speaker 3 who shows a more balanced word order with regard to pre-predicate and post-predicate elements, it has to be mentioned that apart from Turkish and Romeyka he is also competent in Modern Greek, which is likely to have in turn an influence on his data.

In sum, inter-speaker differences in the individual multilingual repertoires of the speakers are able to account for at least some of the variation in word order patterns found in Romeyka.

However, when it comes to copula complements (Table 5), the picture of inter-speaker variation is different with overall low numbers of post-predicate complements irrespective of the percentage of post-predicate other semantic roles. This suggests that there is indeed a change in word-order patterns in copula complements, which is reflected in the data by all three speakers. Moreover, all three speakers tend to omit some copulas; the explanation of this phenomenon, however, requires further research and could be potentially affected by the method of data elicitation for the WOWA dataset (see below).

Table 5: Copula complements in the WOWA dataset per speaker

	Total copula complements	Total VX	% VX	Omitted copulas
Speaker 1 (=text A)	19	2	10%	2
Speaker 2 (=text B, D, E)	4	0	0%	6
Speaker 3 (=text C)	3	1	33%	4

In order to be able to integrate the findings on word order variation in the present WOWA dataset correctly, some critical notes on the dataset and methodology of data collection are in order. Firstly, with 500 tokens, the present dataset is very small and on the lower edge of what can meaningfully be analysed by

quantitative means. It has also to be noted that the WOWA Romeyka dataset has a relatively high number of tokens (N=98) that could not be coded due to non-canonical constructions such as mixing with Turkish or elliptical constructions which also add to the picture of Romeyka as a shifting variety. Secondly, part of the five texts in the WOWA dataset were elicited by means of a storytelling task prompted by picture cards which could have had an influence on the data as well, especially with regard to the omission of copulas which was striking in the narratives elicited with the help of the picture cards. Finally, the fact that the five texts stem from three different speakers which are not directly comparable in terms of gender, age, speech community and their multilingual competences further complicates the analysis. It has to be mentioned, though, that with regard to diatopic variation, the three main dialect areas (Schreiber 2019) are not expected to show significant differences in the domain of word order, except for the potentially different amount of exposure to and use of Turkish, which is considered to be higher in the villages closer to the sea and lower in the more remote mountain villages (see Schreiber 2019).

7 Conclusion

The aim of the present chapter was to analyze the WOWA dataset of Romeyka (Schreiber 2021) with regard to word order in Romeyka in general and the grammatical entities that occur in the post-verbal domain. Romeyka has been described as having inherited VO word order, which has developed mixed directionality under contact with Turkish, which is visible especially in unmarked OV order in subordinate clauses (Neocleous 2020, 2022) but also in ongoing change in certain domains such as copula complements, although the diachronic picture is complex here. Furthermore, the analysis of the present WOWA dataset has revealed that inter-speaker variation mirrors ongoing language shift to Turkish, which complicates the attempt to define default word order patterns for the language and requires a nuanced methodology in assessing word order as if to allow for any meaningful conclusion. In general, information structure accounts for much of the variability in word order patterns that has been described above. Furthermore, PPs seem to behave differently than other obliques/objects, as well in locations and goals as in benefactives and comitatives.

To conclude, the picture of word order in Romeyka is by no means clearly deducible from quantitative data, as is indicated by the number of 55% overall frequency for post-verbal placement. There are several aspects accounting for the significant variation found in the present WOWA dataset: (i) the role of pragmatics, namely information structure, is highly relevant in Romeyka with at the

same time hardly a significant correlation of other factors such as semantic role, flagging or weight, although goals show a tendency to be post-predicate which seems to qualify as an areal (or otherwise for certain reasons universal?) pattern in other languages of the area as well; (ii) the high amount of inter-speaker variation which can be traced back to different levels of Turkish influence in the individual multilingual profiles of the speakers and reflects the status of Romeyka as a shifting variety; but also (iii) considerable intra-speaker variation as a characteristic of language shift; and (iv) ongoing language change as in the case of copula complements, be it internally caused and reinforced by language contact, which requires a very fine-grained and domain-specific investigation not only on the synchronic feature but involving diachronic developments as well as potential contact influences.

Abbreviations

ACC	accusative	NOM	nominative
AOR	aorist	OPN	object pronoun
AUX	auxiliary	PL	plural
CL	clitic	POSS	possessive pronoun
DEM	demonstrative	PRS	present tense
DP	discourse particle	PRT	aspectual particle
GEN	genitive	PST	past tense
INF	infinitive	REL	relativizer
IPF	imperfective	SG	singular
NEG	negation	TOP	topicalizer
NMZ	nominalization	1, 2, 3	1st, 2nd, 3rd person

Acknowledgements

References

Alexiadou, Artemis. 1997. *Adverb placement. A case study in antisymmetric syntax.* Amsterdam/Philadelphia: Benjamins. DOI: 10.1075/la.18.

Bağrıaçık, Metin. 2018. *Pharasiot Greek: Word order and clause structure.* Ghent: Ghent University. (Doctoral dissertation).

Behaghel, Otto. 1909. Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern. *Indogermanische Forschungen* 25. 110–142.

Brendemoen, Bernt. 1998. Einige Bemerkungen über die Wortstellung in anatolischen Dialekten. In Nurettin Demir & Erika Taube (eds.), *Turkologie heute – Tradition und Perspektive: Materialien der 3. Deutschen Turkologen-Konferenz*, 27–46. Wiesbaden: Harrassowitz.

Brendemoen, Bernt. 1999. Greek and Turkish language encounters in Anatolia. In Bernt Brendemoen, Elizabeth Lanza & Else Ryen (eds.), *Language encounters across time and space: Studies in language contact*, 353–378. Oslo: Novus.

Brendemoen, Bernt. 2002. *The Turkish dialects of Trabzon: Their phonology and historical development.* Wiesbaden: Harrassowitz.

Brendemoen, Bernt. 2005. Sentence subordination in the Turkish Eastern Black Sea dialect: Another example of copying from Greek? In Dag Haug & Eirik Welo (eds.), *Haptačahaptāitiš: Festschrift for Fridrik Thordarson on the occasion of his 77th birthday*, 29–47. Oslo: Novus.

Brendemoen, Bernt. 2006. Aspects of Greek-Turkish language contact in Trabzon. In Hendrik Boeschoten & Lars Johanson (eds.), *Turkic languages in contact*, 63–73. Wiesbaden: Harrassowitz.

Brendemoen, Bernt. 2010. Some remarks on the copula in the Eastern Black Sea dialects. In Hendrik Boeschoten & Julian Rentzsch (eds.), *Turkology in Mainz/Turkologie in Mainz*, 47–54. Wiesbaden: Harrassowitz.

Brendemoen, Bernt. 2019. The Eastern Black Sea dialects of Turkey from the point of view of language contact. In Christos Tzitzilis & Georgos Papanastassiou (eds.), *Language contact in the Balkans and Asia Minor*, 248–261. Thessaloniki: Aristotle University of Thessaloniki.

Craevschi, Alexandru. 2022. *Historical contingency and typological tendencies in languages of Western Asia: A quantitative study of word order of non-subject constituents.* Bamberg: University of Bamberg. (MA thesis).

Dawkins, Richard M. 1916. *Modern Greek in Asia Minor.* Cambridge: Cambridge University Press.

Dawkins, Richard M. 1931. Notes on the study of the Modern Greek of Pontos. *Byzantion* 6. 389–400.

Dawkins, Richard M. 1937. The Pontic dialect of Modern Greek in Asia Minor and Russia. *Transactions of the Philological Society* 36. 15–52.

Deffner, Michael. 1878. Die Infinitive in den pontischen Dialekten und die zusammengesetzten Zeiten im Neugriechischen. *Monatsberichte der Königlich Preussischen Akademie der Wissenschaften zu Berlin* 1878. 191–230.

Drettas, Georges. 1997. *Aspects pontiques*. Paris: Association de Recherches Pluridisciplinaires.

Drettas, Georges. 1999. Το ελληνο-ποντιακό διαλεκτικό σύνολο [The Greek-Pontic dialectal set]. In Anastasios-Foivos Christidis, Maria Arapopoulou & Giannoula Giannoulopoulou (eds.), *Διαλεκτικοί θύλακοι της ελληνικής γλώσσας*, 91–100. Athens: Ministry of National Education & Religious Affairs/Centre for the Greek Language.

Dryer, Matthew S. 1995. Frequency and pragmatically unmarked word order. In Pamela A. Downing & Michael Noonan (eds.), *Word order in discourse*, 105–135. Amsterdam: Benjamins.

Dryer, Matthew S. 2013. Determining dominant word order. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info/chapter/s6 (15 November, 2022).

Frascarelli, Mara & Roland Hinterhölzl. 2007. Types of topics in German and Italian. In Kerstin Schwabe & Susanne Winkler (eds.), *On information structure, meaning and form: Generalizations across languages*, 87–116. Amsterdam: Benjamins.

Gandon, Ophélie. 2016. *La relativisation dans une perspective aréale: l'aire Caucase – Anatolie de l'Est – Iran de l'Ouest*. Université Sorbonne Nouvelle. (Doctoral dissertation).

Göksel, Aslı & Celia Kerslake. 2005. *Turkish: A comprehensive grammar*. London/New York: Routledge.

Gundel, Jeanette K. & Thorstein Fretheim. 2004. Topic and focus. In Laurence Horn & Gregory Ward (eds.), *The handbook of pragmatics*, 175–196. Oxford: Blackwell.

Haig, Geoffrey, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils Schiborr. 2024. Post-predicate elements in the Western Asian Transition Zone: Data, theory, and methods. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 3–54. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266331.

Hodgson, Katherine. 2024. Pontic Greek (Armenia). In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* Bamberg: University of Bamberg. https://multicast.aspra.uni-bamberg.de/resources/wowa/.

Holton, David, Geoffrey Horrocks, Marjolijne Janssen, Tina Lendari, Io Manolessou & Notis Toufexis. 2019. *The Cambridge grammar of medieval and early Modern Greek.* Cambridge: Cambridge University Press.

Holton, David, Peter Mackridge, Irene Phillipaki-Warburton & Vassilios Spyropoulos. 2012. *Greek: A comprehensive grammar of the modern language.* London/New York: Routledge.

Horrocks, Geoffrey. 1990. Clitics in Greek: A diachronic review. In Maria Roussou & Stavros Panteli (eds.), *Greek outside Greece, vol. 2*, 35–52. Athens: Diaspora Books.

Horrocks, Geoffrey. 2007. Syntax: From Classical Greek to the Koiné. In Anastasios F. Christidis & Maria Arapopoulou (eds.), *A history of Ancient Greek: From the beginnings to Late Antiquity*, 618–631. Cambridge: Cambridge University Press.

Horrocks, Geoffrey. 2019. Syntax. In David Holton, Geoffrey C. Horrocks, Marjolijne Janssen, Tina Lendari, Io Manolessou & Notis Toufexis (eds.), *The Cambridge grammar of medieval and early Modern Greek, vol. 4*, 1861–2040. Cambridge: Cambridge University Press.

Iefremenko, Kateryna. 2021. Oghuz (Ankara). In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/#turkic.

Janse, Mark. 2002. Aspects of Pontic grammar. *Journal of Greek Linguistics* 3. 203–231.

Janse, Mark. 2006. Object position in Cappadocian and other Asia Minor Greek dialects. *Modern Greek Dialects and Linguistic Theory* 2. 115–129.

Janse, Mark. 2008. Clitic doubling from ancient to Asia Minor Greek. In Dalina Kallulli & Leon Tasmowski (eds.), *Clitic doubling in the Balkan languages*, 165–202. Amsterdam: Benjamins.

Janse, Mark. 2020. Back to the future: Akritic light on diachronic variation in Cappadocian (East Asia Minor Greek). In Klaas Bentein & Mark Janse (eds.), *Varieties of Postclassical and Byzantine Greek*, 201–239. Berlin: De Gruyter Mouton.

Janse, Mark. In press. Η καππαδοκική διάλεκτος [The Cappadocian dialect]. In Christos Tzitzilis (ed.), *Neoellinikes ðialekti [The Modern Greek dialects]*. Thessaloniki: Instituto Neoellinikon Spuðon.

Janse, Mark & Laurentia Schreiber. In prep. Word order and information structure in Romeyka.

Kaltsa, Maria & Ioanna Sitaridou. 2010. Topicalisation in Pontic Greek. *Modern Greek Dialects and Linguistic Theory* 4. 262–279.

Karatsareas, Petros. 2016. The Asia Minor Greek adpositional cycle. A tale of multiple causation. *Journal of Greek Linguistics* 16. 47–86.

Kutscher, Silvia. 2008. The language of the Laz in Turkey: Contact-induced change or gradual language loss. *Turkic Languages* 12. 82–102.

Lacroix, René. 2009. *Description du dialecte Laze d'Arhavi (Caucasique du Sud, Turquie)*. Lyon: Université Lumière Lyon 2. (Doctoral dissertation).

Lacroix, René. 2019. Laz. In Geoffrey Haig & Geoffrey Khan (eds.), *The languages and linguistics of Western Asia: An areal perspective*, 830–860. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110421682-021.

Leitner, Bettina. 2024. Khuzestani Arabic. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Postpredicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 413–430. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266357.

Mackridge, Peter. 1987. Greek-speaking Moslems of North-East Turkey. Prolegomena to a study of the Ophitic sub-dialect of Pontic. *Byzantine and Modern Greek Studies* 11. 115–137.

Neocleous, Nicolaos. 2020. *Word order and information structure in Romeyka: A syntax and semantics interface account of order in a minimalist system*. Cambridge: University of Cambridge. (Doctoral dissertation).

Neocleous, Nicolaos. 2022. The evolution of VO and OV alternation in Romeyka. In Hiwa Asadpour & Thomas Jügel (eds.), *Word order variation: Semitic, Turkic and Indo-European languages in contact*, 141–161. Berlin: De Gruyter Mouton.

Neocleous, Nicolaos & Ioanna Sitaridou. 2022. Never just contact: The rise of final auxiliaries in Asia Minor Greek. *Diachronica* 39. 369–408.

Öztürk, Balkız & Markus A. Pöchtrager. 2011. *Pazar Laz*. München: Lincom Europa.

Papadopoulos, Anthimos A. 1933. Συμβολὴ εἰς τὴν ἔρευναν τῆς ποντικῆς διαλέκτου (ἤτοι γραμματικὰ καὶ συντακτικὰ ζητήματα) [A contribution to the study of the Pontic dialect: Grammatical and syntactic issues]. *Ἀθηνᾶ* 45. 15–48.

Papadopoulos, Anthimos A. 1955. Ἱστορικὴ γραμματικὴ τῆς ποντικῆς διαλέκτου *[Historical grammar of the Pontic dialect]*. Athens: Ἐπιτροπὴ Ποντιακῶν Μελετῶν [Commission of Pontic Studies].

Papadopoulos, Anthimos A. 1958–1961. Ἱστορικὸν λεξικὸν τῆς ποντικῆς διαλέκτου *[Historical dictionary of the Pontic dialect]*, vol. 1–2. Athens: Myrtidis.

Parcharidis, Ioannis. 1880. *Γραμματική της διαλέκτου Τραπεζοῦντος. Περί επιθέτων, αντωνυμιών και βαρυτόνων ρημάτων [Grammar of the Trabzon dialect: On adjectives, pronouns and baritone verbs]*, Manuscript No. 335. Athens: Κέντρο Ερεύνης των Νεοελληνικών Διαλέκτων και Ιδιωμάτων [Research Centre of Modern Greek Dialects & Idioms], Academy of Athens.

Parcharidis, Ioannis. 1888. Συλλογή ζώντων μνημείων της αρχαίας ελληνικής γλώσσας εν Ὄφει' [Collection of living monuments of the ancient Greek language in Ofi']. *Ο εν Κωνσταντινουπόλει Ελληνικός Φιλολογικός Σύλλογος [The Greek Philological Association in Constantinople]* 18. 120–178.

Rafiyenko, Dariya & Ilja A. Seržant. 2020. Postclassical Greek: An overview. In Dariya Rafiyenko & Ilja A. Seržant (eds.), *Postclassical Greek: Contemporary approaches to philology and linguistics*, 1–15. Berlin: De Gruyter Mouton.

Ralli, Angela. 2006. Syntactic and morphosyntactic phenomena in Modern Greek dialects: The state of the art. *Journal of Greek Linguistics* 7. 121–159.

Ralli, Angela. 2020. Selected topics of the morphology of Asia Minor Greek: An introduction. In Angela Ralli (ed.), *The morphology of Asia Minor Greek*, 1–19. Leiden: Brill.

Rasekh-Mahand, Mohammad, Elham Izadi, Mehdi Parizadeh, Geoffrey Haig & Nils Schiborr. 2024. Post-predicate elements in modern colloquial Persian: A multifactorial analysis. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 197–230. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266343.

Revithiadou, Anthoula & Vasilis Spyropoulos. 2009. *Οφίτικη ποντιακή: έρευνα γλωσσικής καταγραφής με έμφαση στη διαχρονία και συγχρονία της διαλέκτου [Ophitic Pontic: Language inventory survey with emphasis on the diachrony and synchrony of the dialect]*. Athens: John S. Latsis Public Benefit Foundation.

Schreiber, Laurentia. 2016. *Assessing sociolinguistic vitality: An attitudinal study of Rumca (Romeyka)*. Heike Wiese, Heiko F. Marten & Oliver Bunk (eds.) (Arbeitspapiere "Sprache, Variation und Migration": Studentische Arbeiten 4). Potsdam: Universität Potsdam. http://www.uni-potsdam.de/fileadmin01/projects/svm/Arbeitspapiere/MA_Schreiber_May.pdf (17 November, 2017).

Schreiber, Laurentia. 2019. Romeyka. In Geoffrey Haig & Geoffrey Khan (eds.), *The languages and linguistics of Western Asia: An areal perspective*, 892–933. Berlin: De Gruyter Mouton.

Schreiber, Laurentia. 2021. Pontic Greek (Romeyka). In Geoffrey Haig, Donald Stilo, Nils N. Schiborr & Mahîr Dogan (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

Schreiber, Laurentia. 2024. *A (Contact-)Grammar of Romeyka.* Bamberg/Ghent: University of Bamberg / Ghent University. (Doctoral dissertation). DOI: 10 . 20378/irb-95115.

Schreiber, Laurentia. Submitted. Muslim Pontic Greek in Turkey: Language shift and bilingual language use. In Konstantinos Sampanis, Matthias Kappler & Eleni Karantzola (eds.), *Sociocultural aspects of Greek-Turkish language contacts.* Wiesbaden: Harrassowitz.

Schreiber, Laurentia. In prep. The Romeyka corpus.

Schreiber, Laurentia & Ioanna Sitaridou. 2017. Assessing the sociolinguistic vitality of Istanbulite Romeyka: An attitudinal study. *Journal of Multilingual and Multicultural Development* 39. 1–16.

Setatos, Michalis. 1994. Το ποντιακό πα [The Pontic pa]. *Ελληνική Διαλεκτολογία* 4. 39–48.

Sitaridou, Ioanna. 2013. Greek-speaking enclaves in Pontus today: The documentation and revitalization of Romeyka. In Mari Jones & Sarah Ogilvie (eds.), *Keeping languages alive: Documentation, pedagogy, and revitalization*, 98–112. Cambridge: Cambridge University Press.

Sitaridou, Ioanna. 2014a. Modality, antiveridicality, and complementation: The Romeyka infinitive as a negative polarity item. *Lingua* 148. 118–146.

Sitaridou, Ioanna. 2014b. The Romeyka infinitive: Continuity, contact and change in the Hellenic varieties of Pontus. *Diachronica* 31. 23–73.

Sitaridou, Ioanna. 2016. Reframing the phylogeny of Asia Minor Greek: The view from Pontic Greek. *CHS Research Bulletin* 4. 1–17.

Sitaridou, Ioanna. 2021. (in)Vulnerable inflected infinitives as complements to volitionals and modals: Evidence from Galician and Romeyka. In Thórhallur Eythórsson & Jóhannes Gísli Jónsson (eds.), *Syntactic features and the limits of syntactic change*, 161–177. Oxford: Oxford University Press.

Soltic, Jorie. 2013. Late Medieval Greek πάλιν: A discourse marker signalling topic switch. *Greek, Roman & Byzantine Studies* 53. 390–419.

Soltic, Jorie. 2015. *The Late Medieval Greek πολιτικὸς στίχος poetry. Language, metre and discourse.* Ghent: University of Ghent. (Doctoral dissertation).

Stilo, Donald & René Lacroix. 2021. Laz (Arhavi). In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation*. Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/ (27 July, 2023).

Thomason, Sarah G. 2001. *Language contact*. Edinburgh: Edinburgh University Press.

Tombaidis, Dimitrios E. 1992. Η τύχη των μικρασιατικών ιδιωμάτων στον ελληνικό χώρο [The fate of Asia Minor idioms in the Greek area]. *Centre for Asia Minor Studies Bulletin* 9. 241–250.

Tombaidis, Dimitrios E. 1996. Η ποντιακή διάλεκτος [The Pontic dialect]. In Dimitrios Tombaidis (ed.), *Μελετήματα ποντιακής διαλέκτου*, 222–233. Thessaloniki: Kodhikas.

van Emde Boas, Evert, Albert Rijksbaron, Luuk Huitink & Mathieu de Bakker. 2019. *The Cambridge grammar of Classical Greek*. Cambridge: Cambridge University Press.

Vaux, Bert. 2007. Homshetsma: The language of the Armenians of Hamshen. In Hovann Simonian (ed.), *The Hemshin: History, society and identity in the highlands of Northeast Turkey*, 257–278. London/New York: Routledge.

Chapter 13

Post-predicate arguments in Modern Eastern Armenian

⊙Katherine Hodgson[a], Victoria Khurshudyan[b] & Pollet Samvelian[c]

[a]University of Cambridge [b]Institut National des Langues et Civilisations Orientales (INALCO) [c]Université Paris III - Sorbonne Nouvelle

This study uses a corpus of oral narratives to investigate the positioning of DOs, as well as other arguments, including goals, in Modern Eastern Armenian (MEA). We find a preference for OV, very strong for indefinite DO (96.8% OV), weaker for definite DO (67% OV). Animacy and weight appear to have a slight effect favouring VO, but the numbers are not statistically significant. Definiteness, animacy, and weight also appear to have a slight effect favouring post-predicate position for other roles. Other roles articulated as pronouns occur less frequently in post-predicate position than lexical NPs. There is no clear evidence that any of the other factors investigated (givenness, topicality, crowding effect, verb type, clause type) affects the position of arguments. The (generally informal spoken) MEA data from this study exhibit similarities to the "OVX" pattern that characterizes comparable registers in languages of the Mesopotamia region, with goals showing a preference for post-predicate position (68.9%), in contrast to other arguments.

1 Introduction

1.1 Overview

Modern Eastern Armenian (MEA) is generally grouped with SOV languages based on the branching direction of its various constituents (quite consistently left-branching), but also some syntactic properties of its VP (e.g. the preverbal position of the focus and bare/indefinite objects), which are assumed to be characteristic of OV languages. The quantitative study of Stilo (2018) also shows a

clear preference for OV order. However, according to some studies (Dum-Tragut 2009, Badikyan 1976, among others), definite DOs display a clear preference for the postverbal position, and grammars of literary MEA generally consider VO to be the unmarked word order. In a recent study, Samvelian et al. (2023) conducted a quantitative investigation of word order in MEA. Based on a small task-oriented corpus and two experiments, they conclude that (S)VO order cannot be qualified as a marked option in MEA, since definite DOs display a strong preference for the VO order in "out-of-the-blue" sentences. Their results confirm, on the other hand, the preference of non-definite DOs for the OV order, indicating furthermore that indefinite (with the indefinite article) and bare DOs are equally likely to favor the OV order. Apart from definiteness, length is also mentioned as a relevant factor, since long indefinite and bare DOs are more likely to behave like definite DOs by appearing in postverbal position.

Apart from DOs, other arguments, especially goals of verbs of motion and caused motion, may also occur in post-predicate position in MEA. This phenomenon has been shown to have an areal dimension, being characteristic of many languages of Western Asia; see contributions to this volume. The purpose of the present study is to investigate factors affecting the position (pre- vs. post-predicate) of DOs, as well as certain other types of arguments that may occur variably in pre- or post-predicate position. Specifically, we aim to check whether the same word order preferences for the placement of DOs reported in Samvelian et al. (2023) also hold in a corpus of unscripted spoken language, as well as to investigate the word order preferences of goals and other types of arguments that show variable position. Besides definiteness and length, we also investigate the relevance of some other factors in the placement of DOs and other arguments:

- Topicality: It has been claimed that word order in MEA is closely linked to information structure (Hodgson 2019). Given that our corpus is annotated for topic persistence and referential distance, as well as other factors connected to topicality, such as givenness and animacy, we can check whether there is a correlation between topicality and postverbal placement.

- Type of referential element (lexical NP vs. pronoun).

- Crowding effect: Hayrapetyan (1981) mentions the presence (or not) of the subject as a relevant feature in the pre- or postverbal placement of the DO in Classical Armenian. This factor may also be relevant in MEA.

- We also investigate the possibility that clause type (main vs. subordinate) and verb type (simple vs. periphrastic) could have an effect on the position of objects.

1.2 Typological preliminaries and background

The position of direct objects in Eastern Armenian is somewhat controversial. Although Armenian is often assumed to be an OV language, Eastern Armenian in particular shows considerable flexibility in the position of objects, especially definite objects, and in WALS is described as having no dominant order (Dryer 2013). The placement of DOs has been the focus of a few studies, including corpus-based and experimental ones. Samvelian et al. (2023) provide a thorough survey of the literature on the issue. They note that despite the fact that typological and theoretical syntactic studies have generally grouped Armenian with SOV languages (Der-Houssikian 1978: 227–228, Dryer 1998: 286, 310, Dum-Tragut 2002, Giorgi & Haroutyunian 2016: 190, Hawkins 1979: 625, 1983: 286, Hodgson 2013: 6, Kahnemuyipour & Megerdoomian 2011, 2017: 81, Kozintseva 1995: 8, Minassian 1980: 263, Tamrazian 1991: 101, 1994: 7, among others), Armenian grammars and descriptive studies, on the other hand, generally refer to the postverbal position as the "natural" position of the object. This is either overtly claimed based on small scale quantitative studies (Badikyan 1976), or is induced by the fact that in most examples illustrating transitive sentences and the DO, the latter is placed in the postverbal position.

Given the fact that OV-VO variation is not trivial in MEA, Samvelian et al. (2023) build on quantitative methods in order to provide a reliable picture of word order distribution in MEA, as well as different factors that may be involved in the choice of OV versus VO order. The main factor investigated by Samvelian et al. (2023) is definiteness. Indeed, several studies have mentioned the relevance of this factor in the placement of the DO (Badikyan 1976, Dum-Tragut 2009, Stilo 2018), with definite DOs occurring more frequently in postverbal position. In a preliminary corpus investigation based on a sample of 570 transitive sentences from the EANC (Eastern Armenian National Corpus, see below), Faghiri & Samvelian (2020) report that definite DOs are overwhelmingly postverbal (79.1% VO vs. 20.9% OV).

The results of Samvelian et al. (2023) are based on a small-scale corpus study and 2 experiments. Their task-oriented spoken corpus includes recordings of picture-based storytelling from 10 native speakers of MEA (see Khurshudyan 2006 and Samvelian et al. 2023 for a detailed description of the corpus and its annotation). Their corpus contains 231 finite declarative transitive sentences with SOV or SVO orders. The results confirm that definite DOs tend to appear in the postverbal position. Samvelian et al. (2023: 476) also show the relevance of length in the placement of the DO: "long and/or heavy bare and indefinite DOs appear frequently in the SVO order, while simple (that is short or minimal) bare and indefinite DOs are more likely to appear in the SOV order."

Samvelian et al. (2023) also provide two sentence production experiments. The first experiment, which compares definite and indefinite DOs, uses a cued sentence recall protocol, and the second one, which compares indefinite and bare DOs, uses constrained sentence production protocol. Furthermore, Samvelian et al. (2023) study the effect of animacy and length.

In the first experiment, 82.7% of definite DOs occur in the postverbal position, while the percentage of postverbal indefinite DOs is 38.3%. As for animacy, although its main effect is not significant, it nevertheless has an effect on the order with definite DOs: animate definite DOs favor SVO more than inanimate definite DOs do. The second experiment shows that both indefinite and bare objects are overwhelmingly preverbal: 74.4% vs. 75.7% respectively. The experiment also confirms the effect of length: non-simple (or long) DOs are more likely to appear in the SVO order.

To sum up, both Faghiri & Samvelian (2020) and Samvelian et al. (2023) report a clear-cut divide between definite DOs on the one hand and indefinite and bare DOs on the other hand. In both corpus-based and experimental studies, definite DOs are overwhelmingly postverbal. Therefore, not only does definiteness seem to be the best predictor for the pre- or postverbal placement of the DO, but also postverbal position seems to be the default placement for definite DOs.

The results of Samvelian et al. (2023) are in sharp contrast with the findings of Stilo (2018). In his data from Colloquial Yerevan Armenian, Stilo (2018) finds that both definite and indefinite DOs show a strong preference for preverbal position: indefinite DOs are 95% preverbal, while definite DOs show a slightly weaker, but still strong preference for preverbal position (86.1%). The other dialects of Armenian included in the study of Stilo (2018) also show a general strong preference for preverbal position in the case of both indefinite and definite DOs: indefinite DOs are preverbal in 100% of cases in the dialects of Erzurum and Stepanakert, and 81% preverbal in Lori, while definite DOs are 98% preverbal in Stepanakert, 91.9% preverbal in Erzurum, and 83.2% preverbal in Lori.

The contrast between the results of Stilo (2018) and those of Samvelian et al. (2023) can be accounted for by several facts: dialectal variation, the language register, and the "genre" of the corpus (discourse, oral, written, etc.), as well as the fact that the data in the experimental studies of Samvelian et al. (2023) include only "out-of-the-blue" declarative non-embedded sentences, which may be considered unmarked, while the present study includes all types of sentences. The data from colloquial Yerevan used by Stilo (2018), which have not been available to us, might also reveal other reasons for the discrepancies. A similar pattern has been observed in Romeyka, where those studies based on elicitation of out-of-the-blue sentences, with overt subjects and objects, yield predominantly VO

structures, while the data from connected spontaneous spoken discourse show a higher rate of OV (Schreiber & Janse 2024 [this volume]). As Samvelian et al. (2023) note in their conclusion, it is very likely that the rate of SOV is higher in spontaneous oral discourse. It is also likely that geographical variation, perhaps associated with different language contact patterns, is relevant, as Stilo (2018) presents data from various dialects spoken in different parts of the Armenian-speaking world. However, even his data from Yerevan show a far lower percentage of post-predicate arguments than those of Samvelian et al. (2023).

In this context, it is important to note that formal literary MEA, as used in writing and formal speech, shows significant differences in morphology, phonology, and syntax[1] from any form of colloquial spoken Armenian, including the colloquial MEA spoken in Yerevan from which Stilo's data are taken. As stated in the introduction, VO order is promoted as the unmarked word order in grammars of the formal literary language.[2] It is likely that the difference between the results of Stilo (2018), even for speakers from Yerevan, and those of Samvelian et al. (2023) reflects the fact that the former deals with colloquial language, and the latter includes data with characteristics of formal literary language,[3] indicating another potential syntactic difference between literary and colloquial MEA. The present study provides further evidence in support of the proposal that register is a key factor in the difference between the results of Stilo (2018) and Samvelian et al. (2023), since the "outlier" speaker with the highest percentage of VO (speaker 3, with 44% VO, as opposed to an average of 17.4% VO for all the other speakers) is

---

[1]Morphological differences between formal literary EA and colloquial EA as spoken in Yerevan include the form of the 3sg. present auxiliary (formal *e* vs. colloquial *a*) and certain forms of the "emphatic" pronoun (formal nom.pl. *irenkʿ* vs colloquial *irankʿ*, formal gen.sg. *ir*, dat.sg. *iren* vs. colloquial gen.sg. *ira*, dat.sg. *iran*) among many others. Phonological differences include the change of the diphthong /ay/ > [e], as in the demonstratives (distal), which have the forms *es*, *ed*, *en* in the colloquial language; see also the formal/colloquial correspondences *hayr/her* 'father,' *mayr/mer* 'mother,' *dzayn/dzen* 'voice,' *layn/len* 'wide,' etc. Syntactic differences include the use of different cases with certain adpositions, for example the use of the dative of 1st and 2nd person pronouns with postpositions such as *het* 'with,' *mot* 'close to, by,' *hamar* 'for' etc. in the formal literary language, as opposed to the genitive in the colloquial language: formal *indz het* 'with me,' *indz mot* 'close to me,' *indz hamar* 'for me' vs. colloquial *im mot, im het, im hamar*, as well as differences in relativization strategies (notably use of indeclinable complementizer instead of declined relative pronoun) described in Hodgson (2019).

[2]Donabédian (p.c.) even reports hearing a teacher of formal literary Western Armenian state that OV order is "Turkish," and therefore incorrect.

[3]For example, examples (18–21) in Samvelian et al. (2023), taken from their corpus, show features characteristic of formal literary language (3sg. aux. *e*, genitive of "emphatic" pronoun *ir* etc.), and while the words used in the experiments could be described as register-neutral, the instructions that appear on the screen as shown in Samvelian et al. (2023) Fig. 1 are in formal literary EA, which is likely to have prompted the use of this form of language in the responses.

also the only one who uses certain word forms associated with the formal literary language (see Table 3 and associated discussion). These findings echo similar findings from Persian, where different grades of formality (and the difference between spoken and written modalities) have a significant impact on certain aspects of word order, see Rasekh-Mahand et al. (2024 [this volume])) for details.

The discussion of the placement of DOs in Stilo (2018) takes place in context of the wider issue of post-predicate constituents in many OV languages of Western Asia. These languages differ from "rigid" OV languages in that they regularly allow at least certain constituents to appear in post-predicate position, giving (O)VX word order. Stilo (2018) and subsequent studies show that the type of constituent that is most common and widespread in post-predicate position is Goal of verbs of motion or caused motion (see Haig et al. 2024 [this volume], for an overview). In some languages, similar behaviour is shown by other constituents which, like goals, can be considered to have the semantic property of "endpoint," namely recipient, benefactive, adddressee, and object of change of state predicates such as 'become'. Arguments without this semantic property, such as ablative, instrumental, locative, and comitative, do not typically appear in post-predicate position in these languages (see Haig & Khan 2019, Stilo 2018). The data in Stilo (2018), which include four different dialects of Armenian (Erzurum, Lori, Stepanakert, and colloquial Yerevan), indicate that in all of these, goals show a preference for post-predicate position. In some dialects, recipient and benefactive also show a tendency to appear in post-predicate position, though not as frequently as goals. Addressee arguments do not seem to be affected by this tendency in Armenian, showing 90–100% pre-predicate position in all the dialects investigated. In fact, most of the historically OV languages in this volume exhibit the same split (see Haig et al. 2024 [this volume], Section 4), distinguishing addressees from spatial goals. Instrumental, ablative, comitative, and locative are predominantly pre-predicate across the whole area.

This pattern is associated with a particular geographic location, with the epicentre in the Mesopotamian region (modern Northern Iraq, Western Iran, and southeastern Turkey) (Haig & Khan 2019, Haig et al. 2024 [this volume]). As for the typological profile of languages showing OVX word order, these have been characterised as OV languages showing some properties typical of VO languages, such as prepositions and/or initial complementizers. The phenomenon reflects the area's status as a "transition zone" from Turkic-type head-final, through Iranian mixed typologies, to Semitic head-initial (Stilo 2006, Haig & Khan 2019, Haig et al. 2024 [this volume]). Armenian, too, could be described as a language with mixed typology, having undergone a change from mainly head-initial Classical Armenian (prepositions, probable VO preference) to mainly head-final modern

Armenian, which does, however, retain some properties associated with head-initial languages, such as initial complementizers. Various proposals have been made regarding the causes of the OVX phenomenon, including iconicity (arguments with endpoint semantics appearing in final position), and contact with languages such as Aramaic (or Russian) that typically have post-predicate arguments (Haig & Khan 2019, Haig et al. 2024 [this volume]). However, in the case of Armenian, it is possible that post-predicate arguments represent a conservative feature retained from Classical Armenian. This could apply to the presence of VO orders as well as OVX, as suggested by both Samvelian et al. (2023) and Stilo (2018), although the fact that post-predicate Goals, but not objects, seem to appear less frequently in subordinate clauses, which have been claimed to show a general tendency for more conservative syntax, perhaps suggests the possibility that OVX may be a more recent phenomenon.

2 The Methodology and the corpus

2.1 Overview

The corpus is composed of 7 oral narratives by 7 participants (6 women and a man), who narrate their favorite movies. It was compiled within the framework of the Eastern Armenian National Corpus[4] (henceforth EANC ArmFilmNarr). The recording was done in Yerevan in 2007–2008, and the total recording time is 1.35h. The corpus contains 11,832 tokens, divided into 2,241 clauses. The corpus was originally created for the study of topic accessibility and continuity of subjects and DO in MEA (in a distinct research project) based on a framework inspired by Givón (1975), Givón (1983) and Du Bois (1987) (for more information on this study, see Hodgson et al. In press). The purpose of the present study is to investigate factors affecting the position (pre- vs. post-predicate) of DOs, as well as certain other types of arguments that may occur variably in pre- or post-predicate position. Thus, we note the position (pre-or post-predicate) of mono-transitive DO, Goal, addressee, recipient, benefactive, endpoint of 'become,' ablative, instrumental, locative, and comitative arguments. For each of these, we note other factors that could potentially affect the position of the argument: weight (number of words/intonation words), type of anaphoric element (e.g pronoun vs.

---

[4]Eastern Armenian National Corpus is a comprehensive corpus of Modern Eastern Armenian comprising approximately 110 million tokens. It encompasses written and oral data starting from the mid 19th century to the present. The corpus is provided with full morphological annotation, offering robust search functionalities, and is openly accessible at www.eanc.net. For more details on Eastern Armenian National Corpus, see Khurshudyan et al. (2022).

definite NP vs. indefinite NP), animacy, and givenness. The presence and position of an overt subject argument is also noted, in order to test for crowding effects, i.e. the tendency to avoid having more than one overt argument on the same side of the predicate. The type of verb form is also noted, as the distinction between periphrastic and simple verbs could potentially affect the position of arguments, as could clause type (main vs. subordinate clause). For DO, figures are also given for topic persistence (number of mentions of the referent in the following 10 clauses) and referential distance (distance in clauses to previous mention of the referent, noted as 1, 2, or 3, with 3 indicating 3 or more). Figure 1 provides an illustrative example.

#	text	Predicate type	Voice	Order	Arg.O	Arg.A	Pers.O	Pers.A	H.O	H.A	NG.O	NG.A	Pre/PostV O	Pre/PostV A	Weight O	Weight A	Other role	Weight	Type	Animacy	Given/New	Pre/post	Others	Verb form	Clause type	Force, Negation	RD.O	RD.A	TP.O	TP.A
59	2_F	TR		LOCOVC	NP.DEF	V.AGR	3	3	NH	H	N	G	Pre		1		Loc	1	PRO.LOC	I	N	Pre		IPT-C			3	1	0	9

/ əndey / duɾə bacʿum a /
'there she opens the door'

Figure 1: Example of annotation of a sentence

In the first column (clause-arm), we find the clause (in Figure 1, the clause and its translation are shown below the columns). The second column shows the type of predicate (here "tr," i.e. monotransitive). There is a column for voice, which is left blank when active. The word order of the clause as a whole is shown (LocOVC, i.e. Locative, Object, (lexical) Verb, Copula/auxiliary). Then, the properties of O are noted in the grey columns, and A in the green columns. These include type of argument (O is NP.DEF, i.e. definite NP, and A is V.agr, i.e. verb agreement), person (3sg for both), animacy (H = human, NH = non-human), givenness (G = given, N = new), the position of overt arguments (Pre, i.e. pre-predicate, for O, A is left blank, because it is expressed by verb agreement alone), and weight of overt arguments (blank for A, 1, i.e. one word, for O). The next column shows that the clause also contains a locative (Loc). In the following columns, its values for weight (1), type of argument, i.e. type of anaphoric element (PRO, i.e. pro-form), animacy, givenness, and position are given. Other columns show the verb type (IPT-C, i.e. imperfective participle + copula/auxiliary), which is of interest particularly in terms of simple (monolectic) vs. complex (participle + auxiliary) verb forms, and the clause type (left blank for main clauses). Referential distance (RD) and topic persistence (TP) are noted in separate columns for subjects and objects of main clauses in which both are 3rd person. Examples of the main categories of anaphoric elements (examples 1-10) and verb types (examples 11-17) are given below:

2.2 Anaphoric elements

(1) **Verb agreement** (for subject only)
Eastern Armenian
*čʿ-git-**em***
NEG-know-1SG.PRS
'**I** don't know.'

(2) **Zero anaphora** (especially for DO, but also some other object-like
elements: example (2) shows zero anaphora for both DO and recipient)
Eastern Armenian

dra	*hamar*	*a*		*nvir-um*
DEM2.GEN	for	be.3SG.PRS		give-IPFV

'He gives **it to her** because of that.'

(3) **Agreement marker** (effectively genitive clitic, can be used for objects of
some adpositions)
Eastern Armenian

aɣjik-ə	*het-**ə***	*xos-um*	*a*
girl-DEF	with-AGR3	speak-IPFV	be.3SG.PRS

'The girl talks with **him**.'

(4) **Personal pronoun**
Eastern Armenian

kʿez	*šat*	*em*	*sir-um*
2SG.DAT	much	be.1SG.PRS	love-IPFV

'I love **you** very much.'

(5) **Demonstrative**
Eastern Armenian

ed	*bacʿ-um*	*en*
DEM2	open-IPFV	be.3PL.PRS

'They open **that**.'

(6) **"Emphatic"**[5].
Eastern Armenian

menkʿ	***iran***	*k-gtn-enkʿ*
1PL.NOM	EMP.DAT	FUT-find-1PL.PRS

'We will find **him**.'

---

[5]For a discussion of this element and its behaviour in this corpus, see Hodgson et al. (In press).
See also Donabedian-Demopoulos (2007), Sigler (2001)

(7) **Pro-adverb**
Eastern Armenian
***əndey** duŕ-ə*   *bacʿ-um*   *a*
there   door-DEF open-IPFV be.3SG.PRS
'**There** she opens the door.'

(8) **Bare NP**
Eastern Armenian
***takʿsi** a*   *gal-is*
taxi   be.3SG.PRS come-IPFV
'**A taxi** comes.'

(9) **NP with indefinite article**
Eastern Armenian
*tencʿ **mi takʿsi** gal-is*   *a*
thus   IA taxi   come-IPFV be.3SG.PRS
'So **a taxi** comes.'

(10) **Definite NP**
Eastern Armenian
***takʿsi-n** kangn-acʿn-um*   *en*
taxi-DEF stand-CAUS-IPFV be.PL.PRS
'They stop **the taxi**.'

2.3 Verb types

2.3.1 Simple

(11) **Monolectic present of some verbs**
Eastern Armenian
un-i   *ir*   *ŕistaran-ə*
have-3SG.PRS EMP.GEN restaurant-DEF
'He has his restaurant.'

(12) **Aorist**
Eastern Armenian
dukʿ   *mard*   ***span-ecʿikʿ***
2PL.NOM person kill-2PL.AOR
'You killed a person.'

(13) **Subjunctive**
Eastern Armenian
*vor   iran   dur **ga***
COMP EMP.DAT like come.3SG.SUB
'so that she likes it'

(14) **Future/conditional forms with the prefix k-** are also classed as "simple"
here
Eastern Armenian
k-ogn-enkʽ
FUT-help-1PL
'We will help.'

2.3.2 Complex (participle + auxiliary)

(15) **Present**
Eastern Armenian
*heto taŕ-er-ə   **gr-um   a***
after letter-PL-DEF write-IPFV be.3SG.PRS
'After (that) it writes the letters.'

(16) **Perfect**
Eastern Armenian
*heto arden bolševik-ner-ə **mt-el   en**   Hayastan*
after already bolshevik-PL-DEF enter-PFV be.3PL.PRS Armenia
'Then the Bolsheviks have already entered Armenia.'

(17) **Future**
Eastern Armenian
*Yes   **gn-alu em***
1SG.NOM go-FPT   be.1SG.PRS
'I will go.'

3 Results and analyses

3.1 Word order variation

Our corpus of oral narratives of favorite films comprising 2,241 clauses shows, overall, twelve possible word order combinations (see Table 1).

Table 1: The distribution of word order configurations in EANC Arm-FilmNarr corpus

#	WO	#	%
1.	SV	415	41.8%
2.	OV	275	27.7%
3.	AOV	94	9.5%
4.	VO	69	6.9%
5.	VS	52	5.2%
6.	AVO	32	3.2%
7.	AV	30	3.0%
8.	OVA	10	1.0%
9.	OAV	10	1.0%
10.	VA	3	0.3%
11.	VAO	2	0.2%
12.	VOA	1	0.1%
	Total	993[a]	100%

[a]The difference between the total in this table (993) and the total number of clauses in the corpus (2,241) is due to the fact that the table only counts clauses that contain overt subjects and/or objects, and a verb. Since Armenian makes frequent use of zero anaphora for both subjects and objects, many clauses have no overt subject or object. There are also some incomplete or syntactically anomalous clauses that have not been counted here.

According to the corpus results, the most frequent word order is SV, with around 42% (415) of all occurrences, which could be accounted for by the abundant number of intransitive verb constructions typical of the narrative genre.

3.2 Direct object

3.2.1 General overview

The second most frequent word order is OV. The absence of the agent (A) can be accounted for by MEA's pro-drop character, which is particularly frequent in oral discourse. OV word order also indicates the prevalence of preverbal position for objects in general. Since almost all clausal objects in MEA are postverbal, it is more informative to limit ourselves to non-clausal objects. Therefore, we concentrate on non-clausal objects, and every mention of objects will be understood as referring to monotransitive, non-clausal objects, unless otherwise stated. The overall distribution of non-clausal objects is 78.7% preverbal vs. 21.3% postverbal

(see Table 2). The data do not show a significant difference between the behaviour of lexical NP and pronominal objects, with the former showing 21.9% VO, the latter 18.9% VO (see Table 2). Armenian does not possess clitic pronouns, so the pronouns in question are "strong," independent pronouns, with the equivalent of "weak" unstressed pronouns being zero anaphora.

Table 2: The distribution of all overt monotransitive non-clausal DOs in EANC ArmFilmNarr corpus

DO type	Total	VO	% VO
Lexical NP	374	82	21.9%
Pronominal	106	20	18.9%
Total	480	102	21.3%

The net preference for preverbal position (21.3% VO, i.e. 78.7% OV) shown in these data is contrary to the study by Samvelian et al. (2023), but roughly consistent with the data in Stilo (2018). Potential factors that could account for this difference are discussed above in Section 2: discourse mode (oral vs. written), genre (oral narratives vs. out-of-the-blue sentences), and register (colloquial vs. formal). The current oral corpus of favorite film narratives potentially includes all sentence modalities which can be part of a structured narrative, and not simply declarative, out-of-the-blue sentences, as was the case for the data used in the study by Samvelian et al. (2023).

Another important difference is register, since Samvelian et al. (2023) include language with characteristics of formal register, whereas the current corpus covers generally colloquial register, although with some semi-formal elements (the context of the recordings was generally informal, though the speakers were aware that they were being recorded, which may have prompted them to use some elements of formal/literary EA). In Samvelian et al. (2023), it has already been proposed that EA word order variation could also be correlated with register variation, as also discussed for Persian in Rasekh-Mahand et al. 2024 [this volume]; the significant morphological, phonological, and syntactic differences between colloquial and formal/literary EA are discussed in Section 2.[1]

Evidence that register is indeed a significant factor in this context is provided by the fact that Speaker 3[6] (see Table 3) is the only speaker to use formal/literary

---

[6]Speaker 3 is the "outlier" with a significantly higher percentage of postverbal DOs than the other speakers (44%, compared to an average of 17.4% for all other speakers, and 21.3% in the corpus as a whole, see Table 3)

forms of demonstratives. For example, her speech contains two examples of medial demonstrative *ayd* as opposed to colloquial *ed/et*) and of the 3rd person "emphatic" pronoun nominative plural *irenkʿ* as opposed to colloquial *irankʿ* (gen.sg. *ir* is also used once by Speaker 1 as well as by Speaker 3).

Table 3: The distribution of postverbal DOs according to speaker in EANC ArmFilmNarr corpus

Speaker	Total O	VO	% VO
1	23	5	21.7%
2	79	13	16.5%
3	45	20	44.4%
4	51	12	23.5%
5	34	6	17.6%
6	43	7	16.3%
7	63	8	12.7%

3.2.2 Impact factors

The following set of impact factors that could potentially be relevant for EA word order variation have been included in our analysis:

1. definiteness

2. givenness

3. animacy

4. topic persistence

5. referential distance

6. weight (heavy NP shift)

7. lexical vs. pronominal Os

8. crowding / null subject (pro-dropping) effects /overt A

9. main vs. subordinate clauses

10. simple vs. complex verb forms

3.2.2.1 Definiteness

"Definiteness" here refers to the presence of the definite article (which is enclitic on the noun). According to Samvelian et al. (2023), EA word order variation is directly correlated with definiteness, with definite DOs being mainly postverbal, and indefinite/bare DOs preverbal. We checked the distribution of preverbal and postverbal bare, indefinite and definite DOs in the current study, and the results confirm the correlation between the definiteness of DOs and postverbal position (see Table 4).

Table 4: The distribution of preverbal and postverbal bare, indefinite (with article) and definite DOs in EANC ArmFilmNarr corpus

Object type	Total	VO	% VO
Definite NP	232	76	33%
Indefinite NP (with indefinite article)	21	2	10%
Bare NP (indefinite without article)	103	2	2%
Total	356[7]	275	mean: 15%

As Table 4 shows, the distribution of definite DOs is 67% preverbal vs. 33% postverbal, whereas that of bare DOs is 98% preverbal vs. 2% postverbal. The indefinite DOs have an intermediate position with 90% preverbal vs. 10% postverbal. Therefore, there is a certain hierarchy of postverbal word order possibility depending on definiteness, in which definite DOs are the most postverbal, followed by indefinite (with article) DOs, and bare DOs are the least postverbal of all. This hierarchy corresponds to the grammatical semantics of definite $\Rightarrow$ specific $\Rightarrow$ non-specific. The peculiarity of EA indefiniteness is that it is bipartite with indefinite (specific) and bare (non-specific) semantics, and the "classical" indefinite noun would tend to be bare, i.e., non-specific, in EA, rather than with an indefinite article, i.e., specific. As the indefinite article is mainly used for marking (specific) indefiniteness, its usage is fairly rare. Hence, the results of the current study show general tendencies that are consistent with Samvelian et al. (2023) in that EA postverbal word order is associated with definiteness and that the more definite/specific the DO is, the more likely it is to be postverbal.

---

[7]This total is smaller that of Table 4, because it only includes lexical NPs, while Table 4 also includes pronominal objects.

3.2.2.2 Givenness

"Given," as opposed to "new," is used here to indicate referents that have been previously mentioned in the discourse. Given Os are considerably more likely to appear in post-predicate position than Os which represent new information. However, this effect is weaker than the effect of grammatical definiteness (definite NP O = 32.9% post-predicate, indefinite (including bare) NP O = 3.3% post-predicate), although the net preference of post-predicate Os for givenness is still considerable (over 80% of post-predicate Os are given) (see Table 5). Thus, it is likely that the givenness effect is merely a reflection of the effect of definiteness (note that in Armenian, certain categories of NPs may be grammatically definite even if not given, e.g. possessives, partitives, and some nominalized non-nominal constituents):

Table 5: The distribution of given and new DOs in EANC ArmFilmNarr corpus

	Total	VO	% VO
O given	289	86	29.8%
O new	190	17	8.9%

3.2.2.3 Animacy

We observe that animate Os are more likely to appear in post-predicate position than inanimate ones. Note, however, that animate referents are considerably more likely to be definite than inanimate ones (animate DO = 79.7% definite, inanimate DO = 59.5% definite), so in order to understand whether the effect of animacy is significant, we need to investigate it in combination with definiteness. The results are shown in Table 6.

We find that among definite Os, animate referents appear slightly more frequently in post-predicate position than inanimate ones, although a Chi-square 2x2 contingency table shows that the difference between definite animate and definite inanimate DOs is not statistically significant. This is similar to the results of Samvelian et al. (2023: 481), who find that the main effect of animacy is not significant, but "there is a marginally significant interaction … indicating that animacy has an effect on the order of definite DOs". The number of indefinite DOs in post-predicate position is, as we have seen, very small, both for animate and inanimate referents.

Table 6: The distribution of animate and inanimate DOs in EANC Arm-FilmNarr corpus

	Total	VO	% VO
Animate O definite NP	63	27	42.8%
Inanimate O definite NP	170	51	30%
Animate O indefinite NP	16	0	0%
Inanimate O indefinite NP	113	4	3.5%

3.2.2.4 Topic persistence

Based on the methodology of Givón (1983), topic persistence was measured as the number of occurrences of that referent in the following 10 clauses (for more details on the methodology and on MEA data see Hodgson et al. In press). The average topic persistence of preverbal Os is equal to 1.6, whereas that of postverbal ones is 1.9. Therefore, postverbal Os have higher topic persistence than preverbal Os. Taking into account the previous impact factors, it could be argued that prototypically postverbal Os are mostly definite, given, and animate, hence have higher topic persistence. The average topic persistence of postverbal Os is higher than that of preverbal definite Os, which show an average topic persistence of 1.5 (the same as that of definite Os as a whole). Human Os have on average a considerably higher topic persistence (3.7) than non-human Os (1), so the fact that postverbal position is possibly associated with animacy/humanity could be a factor in the higher average topic persistence of postverbal as opposed to preverbal Os. Indeed, if we look at objects that are both definite and human, there is no significant difference in average topic persistence between preverbal and postverbal examples (preverbal: 3.08, postverbal: 3.13).

Table 7: Average topic persistence of pre- and post-predicate DOs in EANC ArmFilmNarr corpus

	Average topic persistence
Pre-predicate O	1.6
Post-predicate O	1.9

3.2.2.5 Referential distance

There is only a very small difference between the average referential distance (distance in clauses to previous mention of the referent) of pre-predicate and post-predicate Os, with post-predicate Os showing a slightly smaller average referential distance (2.3) than pre-predicate Os (2.5) (see Table 8).

Table 8: Average referential distance of pre- and post-predicate DOs in EANC ArmFilmNarr corpus

	Average referential distance
Pre-predicate O	2.5
Post-predicate O	2.3

3.2.2.6 Weight (heavy NP shift)

In order to check the correlation between the weight of DOs and their word order variation, the length of DOs in words was measured, more particularly for 1-word, 2-word and 3+ word Os (see Table 9). The results showed a higher frequency of postverbal position for 2-word Os (35% postverbal) vs. 1-word Os (15.4% postverbal). The fact that non-specific indefinites, lacking the indefinite article, which, as we have seen, overwhelmingly appear in preverbal position, are often one-word phrases is a probable factor in the higher frequency of preverbal position for 1-word Os. When O is composed of more than three words, the frequency of postverbal position is actually lower than that for 2-word DOs (23% postverbal), though still higher than that for 1-word DOs.

Subsequently, the measurements were refined in order to explain the distribution of 3+word Os as well as the difference in behaviour between 2-word and 3+word Os. To do this, supplementary factors of definiteness and the pronominal character of Os were added to the analysis of 2-word and 3+ word Os (see Table 9).

According to the results, 2-word definite Os are 43.4% postverbal (cf. 21.3% postverbal for all Os[8]) and 3+ word definite Os are 31.6% postverbal. We then zoomed in on indefinite 3+word Os, distinguishing those with the simple indefinite article *mi* 'a' and the indefinite expression composed of the indefinite article *mi* 'a' and the quantifier *hat* 'unit' which is often in reality the equivalent

---

[8]This figure excludes those Os that appear in both pre- and post-predicate position.

Table 9: The distribution of 1-word, 2-word, and 3+ word Os correlated with O type in EANC ArmFilmNarr corpus

	1-word O			2-word O			3+-word O		
	Total	VO	%VO	Total	VO	%VO	Total	VO	%VO
Bare	73	0	0%	22	2	9.1%	5	0	0%
Indefinite	0	0	0%	5	1	20%	14	1	7.1%
Definite	105	27	25.7%	83	36	43.4%	38	12	31.6%
Pronominal	88	14	16.7%	10	3	30%	4	1	25%
Total	266	41	15.4%	120	42	35%	61	14	23%

of the indefinite article in colloquial register (see Table 11). 21% of all 3+ word Os include the simple indefinite article *mi* 'a,' whereas 79% include the indefinite article with a quantifier (see Table 10). The indefinite article in EA being principally unstressed, the expression of the indefinite article with a quantifier *mi hat* 'a unit' could be considered as one prosodic word, hence it could be included in our measurements as one word rather than two due to its "real weight." However, even when *mi hat* is counted as one word rather than two, 2-word objects still show a higher percentage of OV order (36.2%) than 3+-word objects (26.9%), something which is unexpected from the point of view of heavy NP shift. The generally inconclusive evidence for an effect of weight are consistent with the findings from other corpora of spontaneous spoken language investigated in this volume, which report only marginal effects of weight on object placement (see Haig et al. 2024 [this volume], Section 5).

Table 10: The distribution of 3+ word indefinite Os in EANC ArmFilm-Narr corpus

3+-word indefinite O	Total	VO	% VO
mi	3	1	33.3%
mi hat	11	0	0%

3.2.2.7 Lexical vs. pronominal O

The linguistic character of DOs showed little impact on the word order distribution, with lexical DOs being 21.9% postverbal, and pronominal DOs being 18.9%

postverbal (see Table 11). Overall, preverbal DOs are largely dominant in both cases and the distribution proportion in line with that of all DOs in the corpus (21.3% postverbal, Table 2).

Table 11: The distribution of preverbal and postverbal lexical and pronominal DOs in EANC ArmFilmNarr corpus

Type of O	Total	VO	% VO
Lexical NP	374	82	21.9%
Pronoun	106	20	18.9%

3.2.2.8 Crowding / Null Subject Effect

One of the hypotheses concerning the impact factors was that argument crowding or its opposite, null subject, could affect word order variation so that the presence of an overt subject could induce postverbal DOs to avoid crowding, or the presence of a preverbal DOs could be correlated to null subject effect.

Table 12: The distribution of DOs with and without other overt arguments in EANC ArmFilmNarr corpus

	Total overt DO	VO	% VO
Overt S	148	35	23.6%
No overt S	339	69	20.4%
No overt S or other role[a]	267	51	19.1%
Overt preverbal S or other role	175	39	22.3%

---

[a]"Other role" refers to those elements discussed in section 3.3.

To check this, we first observed the distribution of overt/no overt subjects (S) which, however, manifested insignificant impact on the position of DO (see Table 12). Overall, the distribution is almost identical, therefore, no correlation is observed between the crowding/null subject effect and the word order variation. The same is true if we take into account the presence of elements with other roles as well as subject; the presence of a preverbal argument does not seem to be associated with any increase in postverbal position for DO.

3.2.2.9 Main vs. subordinate clauses

In some languages, object position differs depending on whether the object is in a main clause or a subordinate clause. Our MEA data do not show any significant difference in the position of objects between main clauses and subordinate clauses:

Table 13: The distribution of DOs in main and subordinate clauses in EANC ArmFilmNarr corpus

Clause type	Total O	VO	% VO
Main clause	391	80	20.5%
Subordinate clause	87	20	22.9%

3.2.2.10 Simple vs. complex verb forms

In some languages, different types of verb forms are associated with different positions of arguments. For example, some verb forms of nominal origin may show argument positions analogous to those of noun modifiers, which in modern Armenian almost invariably precede the element they modify. MEA has complex (periphrastic) verb forms involving participles, which have some nominal characteristics, as well as simple verb forms inherited from Classical Armenian.

Thus, we might expect a stronger preference for pre-predicate arguments with periphrastic verb forms. However, direct objects show no evidence for such a pattern, with little difference between different types of verb forms (see Table 14). In fact, periphrastic verb forms show a slightly higher percentage of post-predicate Os than simple verb forms (21.3% vs. 18.3%).

Table 14: The distribution of DOs with complex and simple verbs in EANC ArmFilmNarr corpus

Verb type	Total O[9]	VO	% VO
Complex verb	314	67	21.3%
Simple verb	115	21	18.3%

---

[9]Here, only clauses with a single participle + auxiliary or a single simple verb are counted.

3.3 Other post-predicate arguments

3.3.1 General overview

In addition to objects, other types of arguments also show variable position in Armenian and neighbouring languages of Western Asia. As discussed in Haig & Khan (2019), previous studies find that across the area, there is a tendency for Goal arguments of verbs of motion and caused motion to appear in post-predicate position. In some languages, this tendency is extended to other elements that could be considered to share "endpoint" semantics, namely recipient, benefactive, endpoint of change of state verbs such as 'become,' and addressee; see Haig et al. (2024 [this volume]: Section 4) for an updated overview.

Stilo (2018) finds that in his Armenian data, there is indeed a tendency for goals to appear in post-predicate position. He finds a weaker tendency for benefactive and recipient, but no such tendency for addressee, which strongly prefers pre-predicate position. He finds that instrumental, ablative, locative, and comitative arguments show a preference for pre-predicate position across the area, being generally unaffected by the tendency for post-predicate position associated with goals and goal-like elements. We investigate all these types of arguments, and find that goals do indeed show a preference for post-predicate position (approximately 70% post-predicate). This tendency also seems to apply to benefactive, though the numbers involved are small. The numbers for recipient of 'give,' and of verbs with similar meanings, are also small, but unlike Stilo's (2018) data, show few examples in post-predicate position. Like Stilo (2018), we find that addressee arguments, which are rarely expressed overtly, show no tendency to appear in post-predicate position in our Armenian data. In contrast, the name in sentences such as 'they call him/her/it X,' which are also infrequent, shows a marked preference for post-predicate position (see Table 15).

If we consider goals separately from the other roles in Table 15, we see that they have a much higher instance of post-predicate position (for the other categories with very high post-predicate figures, i.e. name and benefactive, the numbers are too small to draw firm conclusions):

(18)   Goal: 68.9% post-predicate
       Other roles: 29.1% post-predicate

The relatively high percentage of post-predicate position (40%) for ablative is unexpected in the light of proposals that interpret the tendency for post-predicate position as an iconic expression of "endpoint" semantics. It is possible that a different type of analogy is at work here, with ablative equated with Goal

Table 15: Frequency of post-predicate placement, other roles

Other roles	Total	Post-predicate	% Post-predicate
Name ('they call him/her/it X')	5	4	80%
Benefactive	11	8	72.7%
Goal	122	84	68.9%
Ablative	50	20	40%
Comitative	44	13	29.5%
Location	116	30	25.8%
Become	9	2	22.2%
Instrumental	71	15	21.1%
Recipient (transfer of possession in general)	11	2	18%
Recipient (of verb *tal* 'give')	6	1	16.7%
Addressee	6	0	0%
Total other roles[a]	445	178	40%

[a]This figure represents the total of 11 roles, minus Recipient (of verb *tal* 'give'), as this is already included in Recipient (transfer of possession in general).

as both are typical arguments of verbs of motion (ablative as starting point, Goal as endpoint). However, it is more likely that the relatively high post-predicate percentage of ablative in these texts simply reflects the fact that it is more likely than average to be given (76% vs. 65% total other roles) and animate (32% vs. 27% total other roles), factors which seem to favour post-predicate position (grammatical definiteness is not relevant here, as the ablative case ending in MEA cannot co-occur with the definite article).

3.3.2 Impact factors

3.3.2.1 Definiteness

As in the case of DOs, "definiteness" is used here to indicate the presence of the definite article. Note that many of the roles investigated here (those that take ablative, instrumental, locative, or genitive case, and most nominative goals and locations) grammatically exclude the definite article. The fact that goals, which show a marked preference for postverbal position, are often grammatically indefinite for reasons of morphology rather than semantics gives a large number

of postverbal indefinites, masking the effect of definiteness in itself. Therefore, goals are shown separately from other roles in Table 16 below. However, even when goals are excluded, we still find that definites are more likely than indefinites to appear in post-predicate position (38.1% vs. 26.9%). Indeed, goals themselves also show a higher percentage of post-predicate position when grammatically definite (75% vs. 64.9% for indefinite). Nonetheless, the effect of definiteness on roles other than DO is not statistically significant.

Table 16: The distribution of definite and indefinite NP other roles in EANC ArmFilmNarr corpus

	Total	Post-predicate	% Post-predicate
Indefinite NP Goal	57	37	64.9%
Definite NP Goal	52	39	75%
Indefinite NP other	119	32	26.9%
Definite NP other	105	40	38.1%

3.3.2.2 Givenness

In the light of the fact that many other roles cannot take the definite article for grammatical reasons, it might be expected that givenness could show a stronger correlation than definiteness with post-predicate position. However, for roles other than Goal, we find no apparent effect of givenness at all. Note that the correlation between givenness and post-predicate position is much weaker for other roles than for direct object, with new other roles showing 30.1% post-predicate position, compared to 8.9% for new direct objects (see Table 17).

Table 17: The distribution of given and new other roles in EANC Arm-FilmNarr corpus

	Total	Post-predicate	% Post-predicate
Given Goal	77	56	72.7%
New Goal	39	24	61.5%
Given other	221	64	29.0%
New other	103	31	30.1%

3.3.2.3 Animacy

As with direct objects, we see that animate referents appear more frequently in post-predicate position than inanimate ones, although, once again, the effect does not reach statistical significance. However, the difference is smaller here than that found with direct objects, where animates are 28% post-predicate and inanimates 18%. A possible reason for this is that many of the inanimate other role referents are goals, which differ from all other roles in showing a preference for post-predicate position (goals are almost exclusively inanimate). Thus, in Table 18, we present the data for other roles excluding goals. It can be seen that if we discount goals, the effect of animacy becomes more apparent.

Table 18: The distribution of animate and inanimate other roles in EANC ArmFilmNarr corpus

	Total	Post-predicate	% Post-predicate
Total other roles animate	122	45	36.9%
Total other roles inanimate	331	135	40.7%
Other roles animate -Goal	113	38	33.6%
Other roles inanimate -Goal	218	58	26.6%

As we have seen, animate referents are more likely to be definite than inanimate ones, so in order to accurately gauge the effect of animacy, definiteness must also be taken into account, as shown in Table 19. As with DOs, we find that for definite NPs animate referents do show a higher proportion of post-predicate position than inanimates, although again, this does not reach statistical significance, while for indefinite NPs, which show lower frequency of post-predicate position, there does not seem to be any effect of animacy. For the reasons discussed above, these figures do not include goals.

3.3.2.4 Weight

As in the case of DOs, we find that other roles with weight 1 (composed of one word) show a lower frequency of post-predicate position than those which are longer, indicating that weight may be a factor promoting post-predicate position. This is the case even if we exclude pronouns, which are typically composed of one word and show a stronger preference for pre-predicate position than lexical NPs (see following Section e) "Lexical vs. pronominal"). Very heavy elements (weight 4+) show the highest percentage of post-predicate position, indicating

Table 19: The distribution of animate and inanimate other roles in EANC ArmFilmNarr corpus divided by definiteness

Definite NP						Indefinite NP					
Animate			Inanimate			Animate			Inanimate		
N	VX	%VX	N	VX	%VX	N	VX	%VX	N	VX	%VX
41	19	46.3%	64	21	32.8%	25	7	28%	94	25	26.6%

Total definite NP: 105	Total indefinite NP: 119
Total definite VX: 40	Total indefinite VX: 32
% definite VX: 38.1%	% indefinite VX: 26.9%

the possibility of heavy NP shift. The effect of weight is not particularly strong, but other roles present somewhat stronger evidence than DOs for its relevance as a factor favouring post-predicate position (see Table 20).

Table 20: The distribution of other roles according to weight in EANC ArmFilmNarr corpus

Weight	Total	Post-predicate	% Post-predicate
1	239	87	36.4%
1 (NP only)	19	7	36.8%
2	139	59	42.4%
3	51	21	41.4%
3+	74	32	43.2%
4+	23	11	47.8%

3.3.2.5 Lexical vs. pronominal

A comparison of lexical NPs and pronouns shows that the former appear more frequently in post-predicate position in the EANC ArmFilmNarr corpus. However, note that certain types of pronouns, such as interrogative and relative pronouns, show particular syntactic behaviour that places them in pre-predicate position, so it will be more informative to investigate those types of pronouns which can appear in either pre- or post-predicate position. For this reason, we look at demonstrative pronouns, personal pronouns, and the "emphatic" pronoun *inkʿə*

(for a discussion of this element and its behaviour in this corpus, see Hodgson et al. In press; see also Donabedian-Demopoulos 2007). Since all these elements are inherently definite, it is also informative to compare them with definite NPs. We still find that lexical NPs show a clearly higher percentage of post-predicate occurrences than any of these pronouns, and if we compare these pronouns with definite NPs, the difference is greater still. It is interesting that other roles seem to show a clearer difference between the behaviour of lexical and pronominal elements than DOs, which show only a small difference in the frequency of post-predicate position (see Table 21).

Table 21: The distribution of lexical and pronominal other roles in EANC ArmFilmNarr corpus

Type of element	Total	Post-predicate	% Post-predicate
Total pronouns	116	31	26.7%
Total lexical NP	337	149	44.2%
Demonstratives	12	4	33.3%
"Emphatic" pronoun	10	2	20%
Personal pronouns	17	6	35.3%
Definite NP	157	79	50.3%

3.3.2.6 Crowding effect

As discussed in the corresponding section on DOs, it has been proposed that the presence of another pre-predicate argument could favour post-predicate position, in order to avoid "crowding" of more than one argument on the same side of the predicate. However, our data do not provide evidence for this, as in fact the percentage of post-predicate arguments is higher (51.3%) when there is no other overt argument than when there is an overt pre-predicate subject and/or object (36.8%). When both subject and object are pre-predicate, other arguments are less frequent still in post-predicate position (33.3%). In the presence of a post-predicate subject, other roles appear more infrequently in post-predicate position (20%), implying that there may be a tendency to avoid more than one post-predicate argument. However, when there is a postverbal object, the figures are very close to the average for other roles as a whole (37.9%, as compared to ≈ 40% for other roles in general). Thus, these data do not provide conclusive evidence for any type of crowding effect (see Table 22).

Table 22: The distribution of other roles according to the presence of other overt arguments in EANC ArmFilmNarr corpus

	Total	Post-predicate	% Post-predicate
No overt S, A, or O	146	75	51.3%
Overt preverbal S	149	51	34.2%
Overt preverbal A	7	3	42.9%
Overt preverbal O	46	21	45.6%
Overt preverbal A and O	21	7	33.3%
Total overt preverbal only	223	82	36.8%
Overt preverbal A, postverbal O	8	2	25%
Overt preverbal O, postverbal A	0	0	0%
Overt postverbal S	25	5	20%
Overt postverbal A	0	0	0%
Overt postverbal O	29	11	37.9%
Overt postverbal A and O	0	0	0%
Total overt postverbal only	54	16	29.6%

3.3.2.7 Main vs. subordinate clause

In the section on direct objects, we saw that in these data, there is no significant difference between the position of direct objects in main vs. subordinate clauses. However, other roles, and especially goals, show a higher frequency of post-predicate position in main clauses as compared to subordinate clauses. Subordinate clauses have been observed to show more conservative word order patterns than main clauses, for example the persistence of OV in subordinate clauses in Germanic languages such as German. Thus a possible explanation for these findings is that the tendency for post-predicate Goals is a relatively recent phenomenon that has spread by contact from other languages of the area, such as Iranian languages and Neo-Aramaic, and has not spread fully to subordinate clauses. In this context, it is interesting that direct objects do not show such a difference, suggesting that their variable position could indeed be a conservative feature inherited from Classical Armenian, rather than a more recent contact-induced phenomenon. However, note that the percentage of postverbal goals (48%, see Table 23) is still much higher than that of postverbal DOs (22.9%, see Table 13) in subordinate clauses, so we can say that the tendency for post-predicate Goals is still present. In addition, we only have a small number of subordinate clauses with Goal arguments, so it is not possible to draw firm conclusions on this issue.

Table 23: The distribution of other roles in main and subordinate clauses in EANC ArmFilmNarr corpus

	Total	Post-predicate	% Post-predicate
MC total other roles	341	144	42.2%
SC total other roles	84	27	32.1%
MC goals only	94	70	74.5%
SC goals only	25	12	48%

3.3.2.8 Simple vs. complex verb form

We observe that complex (participle + auxiliary) verb forms show a lower percentage of post-verbal arguments, including goals, than simple verb forms. A possible explanation for this phenomenon is that arguments of participial (nominalized) verb forms tend to show positional characteristics of noun modifiers, i.e. preceding the element they modify. It is also possible that the phenomenon is linked to properties of focus marking in Eastern Armenian, as the auxiliary in complex verb forms can mark focus when the focused element precedes the lexical verb, but not when it follows. However, neither of these explanations is particularly convincing given the fact that direct objects do not seem to show this pattern (as seen in Section 3.2.2, in this corpus, complex verb forms in fact show a slightly higher percentage of post-predicate DOs than simple verb forms). Further research is clearly needed to clarify the interaction between verb type, word order, and information structure in Armenian.

Table 24: The distribution of other roles with complex and simple verbs in EANC ArmFilmNarr corpus

	Total	Post-predicate	% Post-predicate
Complex verb total other roles	288	110	38.2%
Complex verb goals only	77	53	68.8%
Simple verb total other roles	53	27	50.9%
Simple verb goals only	18	14	77.8%

3.3.3 Summary of other roles

The data from our MEA corpus confirms that in MEA, as in other languages of the wider area, goals of verbs of motion and caused motion show a preference for post-predicate position (approximately 70%). However, in contrast to some languages of Western Asia (see Haig et al. 2024 [this volume]), there is no such tendency observed for other constituents sharing the semantics of "endpoint," such as recipient or addressee. An apparent exception is benefactive, which shows an even stronger preference for post-predicate position (72.7), although the small number of examples makes this less reliable. It is also worth noting that the number of overt recipient and addressee referents in this corpus is very low. Apart from Goal, virtually all the other roles investigated which have more than 20 examples (comitative, location, instrumental) show similar figures, of 20–30% post-predicate position, similar to those for direct objects overall. Ablative has a slightly higher figure (40% post-predicate), but this may be because the ablative referents in this corpus show higher than average figures for givenness and animacy, which may have some effect favouring post-predicate position.

Definiteness, too, is shown to favour post-predicate position, but indefinite other roles do not show the extreme preference for pre-predicate position that is characteristic of indefinite direct objects. Even bare indefinite other roles show 38.7% post-predicate position, while indefinite other roles in general (including those with the indefinite article or an indefinite proform) show 39.2%, virtually identical to the average of other roles as a whole (40%). Weight appears to have an effect on position, with heavier elements appearing somewhat more frequently in post-predicate position. Other roles show more evidence of an effect of weight than direct objects, although this is still not particularly strong. Other roles also show more clearly than direct objects the differences in behaviour between pronouns and lexical NPs, with the latter being more likely to appear in post-predicate position (44.2%, vs. 26.7% for pronouns). As with direct objects, there is no evidence for a crowding effect, whereby the presence of other pre-predicate arguments could promote post-predicate position in order to avoid "crowding" of arguments on one side of the predicate. Unlike direct objects, other roles present possible evidence that post-predicate constituents may be more common in main than subordinate clauses, and with simple rather than complex verb forms. A possible explanation for the former could be that the post-predicate Goal phenomenon is a recent contact-induced development that has spread more slowly to subordinate clauses, although the numbers are too small to draw any firm conclusions. The link between argument position and verb form is a topic for further research.

4 Conclusions

As regards direct objects, definiteness proved to be a key impact factor for the postverbal position: 33% (def) vs. 10% (indef) vs. 2% (bare). This is consistent with the previous study by Samvelian et al. (2023), with the difference that in the present study, the percentage of post-predicate definite Os is considerably lower than those reported by Samvelian et al. (2023), who find 82.7% of definite Os in post-predicate position in their first experiment. The percentage of post-predicate definite DOs in the present study is intermediate between the very high figures found by Samvelian et al. (2023), and the very low figures (around 14% for colloquial Yerevan) reported by Stilo (2018). One probable factor behind the difference is that the data in the experimental studies of Samvelian et al. (2023) include only out-of-the-blue sentences; note that a similar pattern has been observed in Romeyka, where those studies based on elicitation of out-of-the-blue sentences yield predominantly VO structures, while the data from connected spontaneous spoken discourse show a much higher rate of OV (Schreiber & Janse 2024 [this volume]). As Samvelian et al. (2023) note in their conclusion, it is very likely that the rate of SOV is higher in spontaneous oral discourse.

Another relevant factor, also proposed by Samvelian et al. (2023), is register. Samvelian et al. (2023) includes data with characteristics of formal literary language, which show significant morphological, phonological, and syntactic differences from colloquial Yerevan EA, such that the two should be considered different forms of language. The association of post-predicate Os with formal registers is supported by the fact that the speaker who uses by far the highest percentage of postverbal Os in the present study (44%, as opposed to an average of 17.4% for all the other speakers, and 12.7% for the speaker with the lowest percentage) is also the only one who uses certain word forms associated with the formal literary language. Thus, if we discount this one speaker, who uses a more formal register, the percentage of postverbal Os in this study is not so different from that recorded by Stilo (2018) for colloquial Yerevan. In addition, a similar figure (approx. 90% OV) is obtained for nominal direct objects in the Agulis corpus of spoken vernacular Armenian (Hodgson In press).

The effect of grammatical definiteness is stronger than that of the pragmatic property of givenness (29.8% of given Os appear in post-predicate position, vs. 8.9% of new information Os). Among definite Os, animate referents show a slightly higher percentage of post-predicate position (42.8% for animates vs. 30% for inanimates), although this does not reach statistical significance. This, too, is consistent with the findings of Samvelian et al. (2023). Given that definiteness and animacy are characteristics associated with topical referents, it is un-

surprising that topic persistence is higher for postverbal Os (1.9 (postverbal) vs. 1.6 (preverbal)). Post-predicate Os are also associated with slightly lower average referential distance than pre-predicate ones (2.3 vs. 2.5), which is also to be expected given the association of post-predicate position with topicality in general. Some effect of heavy NP shift effect was observed, with longer NPs being more frequent in post-predicate position. (13% (1-word Os) vs. 31% (2-word Os) vs. 19% (3+word Os)). This is also broadly consistent with the findings of Samvelian et al. (2023). Neither the object type (lexical or pronominal) nor the crowding / null subject effect had any evident impact on the position of DOs. Overall, the present corpus study of oral narratives showed that OV word order is more frequent in MEA than VO (79% preverbal vs. 21% postverbal). A prototypical postverbal O in this corpus is definite, given (91%), human (41% vs. 27% of preverbal objects), with higher topic persistence.

As regards other post-predicate constituents, the same factors that have been found to be associated with post-predicate position for DOs (definiteness, animacy, and weight) seem to have a slight effect in the case of other roles too, although the numbers involved cannot be considered statistically significant. The effect of definiteness is less pronounced than for DOs, with 50.3% of definite NP other roles appearing in post-predicate position, vs. 39.2% of indefinites. As we can see, indefinite other roles do not show such a strong tendency to avoid post-predicate position as indefinite DOs. As with DOs, animate definite other roles (excluding goals) are more likely to appear in post-predicate position than inanimate definite ones (46.3% vs. 32.8%). As with DOs, animacy does not seem to affect the position of indefinites. Other roles present slightly stronger evidence of heavy NP shift to post-predicate position than DOs, with those comprised of 4+ words showing the highest percentage of post-predicate position (weight 1 = 36.4% post-predicate, weight 2 = 42.4%, weight 3 = 41.4%, weight 4+ = 47.8%), although the observed effect is still fairly weak. Other roles present evidence that pronominal arguments are less likely to appear in post-predicate position than lexical NPs (26.7% post-predicate, vs. 44.2% for lexical NPs), while for DOs there is no apparent difference. As for direct objects, no evidence is found of a crowding effect.

The preference for goals to appear in post-predicate position is a separate issue, that has been shown to have an areal dimension. This preference is confirmed by the data in this study, where 68.9% of goals appear in post-predicate position. This study presents no evidence that the preference for post-predicate position is extended to other constituents with "endpoint" semantics, such as recipients or addressees, with the possible exception of benefactive. However, the numbers

of overt examples of all these types of argument (recipient, addressee, benefactive) are very small, so we cannot draw a firm conclusion here. The fact that the tendency for goals to appear in post-predicate position is more pronounced in main than subordinate clauses is possible evidence that it is a relatively recent, contact-induced phenomenon. The position of direct objects shows no significant difference between main and subordinate clauses, and it is possible that the existence of postverbal objects, and perhaps other arguments, too, is a conservative characteristic inherited from Classical Armenian (see Samvelian et al. 2023, Stilo 2018), which may perhaps also explain its apparent association with formal register. This is a topic for future research. In any case, Modern Eastern Armenian can be said to fit the typological profile of an "OVX" language, in that despite showing mainly head-final characteristics, it also has some characteristics associated with typically head-initial languages, such as initial complementizers.

Abbreviations

1	first person	FUT	future
2	second person	GEN	genitive
3	third person	IA	indefinite article
AGR3	third-person agreement marker	IPFV	imperfective
AOR	aorist	NEG	negative
CAUS	causative	NOM	nominative
DAT	dative	PFV	perfective
DEM2	medial demonstrative	PRS	present
DEF	definite	PL	plural
EMP	emphatic	SG	singular
FPT	future participle	SUB	subjunctive

References

Badikyan, Khachik. 1976. *Žamanakakicʿ hayereni parz naxadasutʿyan šaradasutʿyunə.* Yerevan: Yerevani hamalsarani hratarakchutyun.

Der-Houssikian, Haig. 1978. The semantic significance of S, O, V sequences in Armenian. In Mohammad A. Jazayery, Edgar C. Polomé & Werner Winter (eds.), *Linguistic and literary studies, vol. 2: Descriptive linguistics*, 227–236. Lisse: Peter de Ridder Press.

Donabedian-Demopoulos, Anaïd. 2007. A la recherche de la logophoricité en arménien. In Jocelyne Fernandez-Vest (ed.), *Combat pour les langues du monde/Fighting for the world's languages, hommage à Claude Hagège*, 165–176. Paris: Editions L'Harmattan.

Dryer, Matthew S. 1998. *Aspects of word order in the languages of Europe*. Berlin: De Gruyter Mouton.

Dryer, Matthew S. 2013. Order of object and verb. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info/chapter/83.

Du Bois, John. 1987. The discourse basis of ergativity. *Language* 63(4). 805–855. DOI: 10.2307/415719.

Dum-Tragut, Jasmine. 2002. *Word order correlations and word order change: An "applied-typological" study on literary Armenian variants*. Munich: Lincom.

Dum-Tragut, Jasmine. 2009. *Armenian: Modern Eastern Armenian*. Amsterdam: Benjamins. DOI: 10.1075/loall.14.26.

Faghiri, Pegah & Pollet Samvelian. 2020. Word order preferences and the effect of phrasal length in SOV languages: Evidence from sentence production in Persian. *Glossa: A journal of general linguistics* 5(1). DOI: 10.5334/GJGL.1078.

Giorgi, Alessandra & Sona Haroutyunian. 2016. Word order and information structure in Modern Eastern Armenian. *Journal of the Society for Armenian Studies* 25. 185–201.

Givón, Talmy. 1975. Focus and the scope of assertion: Some Bantu evidence. *Studies in African Linguistics* 6. 185–205.

Givón, Talmy. 1983. Topic continuity in discourse: An introduction. In Talmy Givón (ed.), *Topic continuity in discourse: A quantitative cross-language study*, 1–42. Amsterdam: Benjamins Publishing Company.

Haig, Geoffrey & Geoffrey Khan. 2019. Introduction. In Geoffrey Haig & Geoffrey Khan (eds.), *The languages and linguistics of Western Asia: An areal perspective*, 1–29. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110421682-001.

Haig, Geoffrey, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils Schiborr. 2024. Post-predicate elements in the Western Asian Transition Zone: Data, theory, and methods. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 3–54. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266331.

Hawkins, John A. 1979. Implicational universals as predictors of word order change. *Language* 55. 618–648.

Hawkins, John A. 1983. *Word order universals*. New York: Academic Press.

Hayrapetyan, Sergo A. 1981. Uγiγ xndri šaradasutʻunǝ dasakan grabarum. *Herald of Social Sciences* 9. 71–84.

Hodgson, Katherine. 2013. *Discourse configurationality in Eastern Armenian: The nominal projection.* Cambridge: University of Cambridge. (MA).

Hodgson, Katherine. 2019. *Relative clauses in colloquial Armenian: Syntax and typology.* Paris: Institut national des langues et civilisations orientales. (Doctoral dissertation).

Hodgson, Katherine. In press. Armenian, dialect of Agulis. In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* Bamberg: University of Bamberg.

Hodgson, Katherine, Victoria Khurshudyan & Anaïd Donabedian. In press. Word order, anaphora, and topicality in Eastern Armenian. In Claudine Chamoreau, Enrique Palancar & Anaïd Donabedian (eds.), *Topicality and the shaping of grammar: New perspectives from lesser-known languages.* Amsterdam: Benjamins.

Kahnemuyipour, Arsalan & Karine Megerdoomian. 2011. Second position clitics and the vP phase: The case of the Armenian auxiliary. *Linguistic Inquiry* 42(1). 152–162. DOI: 10.1162/ling_a_00033.

Kahnemuyipour, Arsalan & Karine Megerdoomian. 2017. On the positional distribution of an Armenian auxiliary: Second-position clisis, focus, and phases. *Syntax* 20(1). 77–97.

Khurshudyan, Victoria. 2006. *Средства выражения хезитации в устном армянском дискурсе в типологической перспективе: Hesitation means in oral Armenian discourse in typological perspective.* Moscow: Russian State University for the Humanities. (Doctoral dissertation).

Khurshudyan, Victoria, Timofey Arkhangelskiy, Misha Daniel, Vladimir Plungian, Dmitri Levonian, Alex Polyakov & Sergei Rubakov. 2022. Eastern Armenian national corpus: State of the art and perspectives. In Victoria Khurshudyan, Nadi Tomeh, Damien Nouvel, Anaid Donabedian & Chahan Vidal-Gorene (eds.), *Proceedings of the workshop on processing language variation: Digital Armenian (DigitAm) within the 13th Language Resources and Evaluation Conference*, 28–37. Marseille: European Language Resources Association. https://aclanthology.org/2022.digitam-1.5.

Kozintseva, Natalia. 1995. *Modern Eastern Armenian.* Munich: Lincom Europa.

Minassian, Martiros. 1980. *Grammaire d'Arménien Oriental.* New York: Delmar.

Rasekh-Mahand, Mohammad, Elham Izadi, Mehdi Parizadeh, Geoffrey Haig & Nils Schiborr. 2024. Post-predicate elements in modern colloquial Persian: A multifactorial analysis. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald

Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 197–230. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266343.

Samvelian, Pollet, Pegah Faghiri & Victoria Khurshudyan. 2023. On the persistence of SVO: The case of Modern Eastern Armenian. *Linguistics* 61(3). 459–497. DOI: 10.1515/ling-2021-0013.

Schreiber, Laurentia & Mark Janse. 2024. Word order variation in Romeyka. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 337–373. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266353.

Sigler, Michele. 2001. A logophoric pronoun in Western Armenian. *Annual of Armenian Linguistics* 21. 13–30.

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

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

Tamrazian, Armine. 1991. Focus and WH-movement in Armenian. *UCL Working Papers in Linguistics* 3. 102–121.

Tamrazian, Armine. 1994. *The syntax of Armenian: Chains and the auxiliary*. London: University College London. (Doctoral dissertation).

Part IV

Case studies from Semitic languages

Chapter 14

Khuzestani Arabic

🅞 Bettina Leitner

Universität Wien

This chapter describes the basic word order profile of Khuzestani Arabic and dis-
cusses possible reasons for deviations from the default word order VX (X repre-
senting non-subject arguments). This discussion includes an analysis of where the
change from VX to XV may be triggered by language contact or by language in-
ternal reasons related to information structure. The description is mainly based on
data from the WOWA-corpus (https://multicast.aspra.uni-bamberg.de/resources/
wowa/#semitic) and supplemented by the author's own corpus data.

1 Introduction and data

The Arabic variety spoken in the southwestern Iranian region of Khuzestan be-
longs to the Mesopotamian dialectal area and the subgroup of *gələt* dialects[1]. The
grammar of Khuzestani Arabic (KhA) has been described mainly by Bruce Ing-
ham and the author of this chapter (among the most important contributions are
Ingham 1973, 1976, 2008, and Leitner 2022b). This chapter presents the first anal-
ysis of word order structures in KhA since Ingham (1991) and has the advantage
of being based on a comparatively large text corpus.

Arab settlement in southern and western Iran (i.e. Khuzestan and Fars) is al-
ready documented for Sasanian times (226–651 AD), and thus precedes the arrival
of the Arab Muslim armies (Zarrinkūb 1975: 27). However, the real Arab disper-
sal into Iran began after the initial Islamic victories when many tribes from the

---

[1]The term *gələt*, first used by Blanc (1964) to classify the Iraqi Arabic dialects, is based on the 1SG
PFV verb for 'to say': *gələt* versus *qəltu*—the latter being the other group of dialects spoken in
Iraq and southern Anatolia. *gələt*-dialects are associated with Bedouin and rural Arabic, even
though nowadays of course almost no speakers live as nomads any longer and many have
moved to cities, cf. Leitner (2021b).

vicinities of Kufa and Basra entered Iranian soil following the conquest (cf. Leitner 2022b: 6–7 and the references mentioned there). Many of the Arab tribes who immigrated into Khuzestan had originated in Arabia (cf. Savory 1986: 81; cf. Nadjmabadi 2009: 132, Fn. 28–29; Field 1939: 604) and first settled in southern Iraq. Their subsequent immigration to Khuzestan led to an extensive Arabization of the province, parts of which were officially called Arabistan from the 16th/17th century until 1923 (Oppenheim 1967: 3, 10; cf. Ingham 1997: ix). Most Arab tribes, such as the Kaʕab, adopted Shiism after their settlement in Iran, but some remained Sunnis, e.g. the Muntafiq, who migrated to Hoveyzeh in 1812 (Savory 1986: 81).

The Iran-Iraq War (1980–-1988) forced many families to flee their hometowns and thus led to considerable demographic changes. Both the city of Khorramshahr/Muḥammara (and its port) and the city of Abadan (and its refineries) were completely destroyed in the course of the Iran-Iraq War by Iraqi artillery and aerial bombardments.

The region's capital city Ahvaz in turn has witnessed an immense growth in the past decades (according to Nejatian 2015 the number of inhabitants in Ahvaz grew from 334,399 in 1976 to 724,653 in 1991, and to 1,112,021 in 2011).

Persian, being the majority language as well as the only official language and language of education and administration in Iran, plays a crucial role for the people in Khuzestan. The majority of the Khuzestani Arabs are bilingual, but there are still monolingual Arabic speakers, especially among the older generation. KhA is insulated from influence by MSA, but sharing a long geographically open border with Iraq, Khuzestan is not totally isolated from the Arabic-speaking world. The linguistic influence Persian has had on KhA is strongest in lexicon, but it is also evident in some aspects of its phonology and syntax. This paper will focus on possible language contact influences in the domain of syntax with a focus on word order and sentence structure. For this purpose, we will characterize the word order profile of KhA relying primarily on the WOWA corpus data[2]. For a more general evaluation of language contact phenomena in Khuzestani Arabic, cf. Gazsi (2011), Matras & Shabibi (2007); Leitner (2020).

The data for this contribution was gathered in field studies in Khuzestan in 2016 and comprises 6 texts (about 9,600 transcribed words) from 8 different speakers (5 female, 3 male, aged between 30 and 65). Table 1 provides an overview of this data, which contains three narrative interviews, one conversation, one traditional tale, and one procedural description (recipe). Examples taken from the author's data other than that of the WOWA-corpus data will be labelled as "(own data)".

---

[2]https://multicast.aspra.uni-bamberg.de/resources/wowa/#semitic.

Table 1: Metadata WOWA corpus

	Text name	Genre	Gender	Age
A	Mōze - Past times	Narrative interview	F	65
B	Shepherd	Narrative interview	M	30
C	Hamidiye women	Conversation	F, F	60, 60
D	Ghazawiyya Palm farmer	Narrative interview	M, M	35, 30
E	Umm Saʕad - Ḥamda	Traditional tale	F	45
F	Amine - recipes	Procedural text	F	30

The main body of this chapter contains general information on KhA sentence structure and word order (Section 2). This is followed by a discussion on the factors that may trigger pre-predicate position of constituents in KhA and the likelihood of an explanation of these changes via Persian influence or, otherwise, due to language-internal reasons related to information structure and focus-fronting (Section 3).

2 Word order profile

Out of the dataset's 546 analyzable tokens[3], 479 were found in post-predicate position and only 67 in pre-predicate position, which confirms that Khuzestani Arabic, like most Arabic varieties, is a VO language. This suggests that the influence of the contact language Persian, an OV language, on KhA word order is not very strong. In the following, a brief word order profile of KhA will be given.

2.1 Adjective/noun

Adjectives generally follow nouns in KhA as in most other Arabic varieties (1).

(1)   Khuzestani Arabic (own data)
   baḷḷa xall   nšərrb-a   māy fāyər yiġsil
   DM   HORT make_drink.IPFV.1PL-3SG.M water boiling wash.IPFV.3SG.M
   ṣadr-a
   breast-3SG.M
   'Let us make him drink hot [lit. boiling] water that makes him feel good [lit. cleans his breast].'

---

[3]Tokens are defined as non-subject-constituents with one of the following roles: direct object; Goal; recipient; addressee; location; instrument; comitative; copula complement noun; possessed NP in a possessive expression; and complement of a change-of-state verb.

2.2 Possessor/possessed

Possession is expressed either via synthetic nominal attribution or via the analytic genitive. The basic syntagm for synthetic nominal attribution constructions (Arabic *ʔiḍāfa*) is NOUN (in construct state) + NOUN/DEF-NOUN. The second noun is usually a (definite or indefinite) substantive, as in (2) and (3).

(2) Khuzestani Arabic (Leitner 2021a: D, 0591)
ḥalīb əl-hōš
milk DEF-cow.COLL
'cow's milk'

(3) Khuzestani Arabic (Leitner 2021a: A, 0166)
ṭāsa-t   rōba
bowl-CS yoghurt
'a bowl [full] of yoghurt'

The two default types of the analytic genitive syntagm are: NOUN (POS-SESSED) + MĀL + NOUN (POSSESSOR) (4) and NOUN + MĀL-PRONOMINAL SUFFIX (5). There are also examples in which the element after *māl* is an adverb. This linker for nominal attribution is usually labeled in Arabic dialectology as a "genitive exponent" or "genitive marker" (see Leitner 2022b: 176–189 and the references mentioned there for a more detailed elaboration of such constructions in KhA).

(4) Khuzestani Arabic (Leitner 2021a: E, 0652)
ġaṣər māl   malək
castle GL.SG.M king
'a castle of a king'

(5) Khuzestani Arabic (Leitner 2021a: D, 0537)
əl-xūṣ   māl-a
DEF-palm_fronds GL.SG.M-3SG.M
'its [the palm's] fronds'

(6) Khuzestani Arabic (Leitner 2021a: A, 0217)
maṯal   ʕad-na xamsīn ḥōliyye   ʕad-na sittīn
for_example at-1PL fifty   young_female_buffalo at-1PL sixty
ḥōliyye
young_female_buffalo
'We had like fifty [young female] buffaloes or sixty buffaloes.'

For possessive constructions with the preposition *ʕad*, lit. 'at', the default syntagm is *ʕad*-PRO (POSSESSOR) + POSSESSED (6).

While *māl* can be used both predicatively and attributively, *ʕad* can only be used predicatively. In the WOWA corpus, there are 4 out of 32 instances of the possessive construction with *ʕad* that have the POSSESSED in pre-predicate position, i.e. preceding *ʕad*-PRO, as in (7): *ʔamān ma ʕad-na* 'we don't have safety'. This marked pre-predicative word order is often used in combination with negation to stress non-possession of a certain item and often co-occurs with a repetition of an already mentioned noun (here: 'safety'), as illustrated in the following example.

(7) Khuzestani Arabic (Leitner 2021a: B, 0224)
     əs-surūḥ   w   ʕalaf   w   taʕb   w   nṭāra   w
     DEF-grazing and fodder and exhaustion and watching and
     b-əl-ʔaham   ham ha-l-ayyām   masʔala masʔalt əl-ʔamān
     in-DEF-most_important also DEM-DEF-days question question DEF-safety
     ʔamān ma   ʕad-na əl-ḥalāl   b-īd-ak   w   yəmbāg
     safety NEG at-1PL DEF-cattle in-hand-2SG.M and be_stolen.IPFV.3SG.M
     'The grazing and the fodder, the exhaustion, and the guarding, and the important thing these days is the question of safety, we don't have safety, the cattle that is in your hand might be stolen [any minute].'

kəllšāy 'everything' often precedes a negative possessive-construction *mā ʕad*-PRO, as in the following example (8), to express 'to really have nothing', i.e. for emphasizing the fact that one really does not own anything (and with that implicitly contrasting her with others who have more). This structure might be a calque on Persian *hičči na-dāšt* 'she had nothing', but is also found in Iraqi Arabic (see e.g. Leitner et al. 2021: 172 "*kull wakit ma ʕindi.* 'Ich habe gar keine Zeit.'" and "*kull šī māku.* 'Es gibt gar nichts.'"; cf. Section 3).

(8) Khuzestani Arabic (Leitner 2021a: E, 0614)
     w   hāy l-əbnayya bass əhəya təsraḥ   b-əl-ġanam
     and DEM DEF-girl   but 3SG.F graze.IPFV.3SG.F with-DEF-sheep
     kəllšāy   mā ʕad-ha
     everything NEG at-3SG.F
     'And this girl was always just grazing the sheep, she had nothing [else].'

2.3 Demonstrative/noun

Demonstratives usually come before the noun (9).

(9)   Khuzestani Arabic (Leitner 2021a: C, 0502)
      hāḏann   lə-hədūm
      DEM.PL.F DEF-clothes
      'these clothes'

   Though their position before the noun prevails, demonstratives can also follow the head. In such constructions, however, the noun is often emphasized (10).

(10)   Khuzestani Arabic (own data)
       əl-walad hāḏa   rabbō-(h)
       DEF-boy DEM.SG.M raise.PFV.3PL.M-3SG.M
       'They raised this boy.'

   Also, the noun can be both preceded and followed by a demonstrative (11; such constructions are usually limited to the SG proximal demonstratives M *hāḏ*, F *hāḏi, hāy*).

(11)   Khuzestani Arabic (own data)
       əḥna hāy əl-ʕarab   hāy
       1PL   DEM DEF-arabs DEM
       'we Arabs [COLL]'

2.4 Numeral/noun

Numerals generally precede nouns (12, 13).

(12)   Khuzestani Arabic (Leitner 2021a: A, 0217)
       sittīn ḥōliyye
       sixty young_female_water_buffalo
       'sixty [young female] water buffaloes'

(13)   Khuzestani Arabic (Leitner 2021a: B, 0226)
       sitt əšhur
       six month.PL
       'six months'

2.5 Adpositions

Prepositional phrases usually follow the verbal predicate, as in (14) and (15). A counterexample is e.g. provided by (18), where *mən zuġur* 'from childhood (on)' precedes the predicate.

(14) Khuzestani Arabic (Leitner 2021a: E, 0654)

gālat   ʔāna ʔaḏəllan   bə-hāḏa   ġaṣər māl   əl-malək
say.PFV.3SG.F 1SG   stay.IPFV.1SG in-DEM.SG.M castle GL.SG.M DEF-king

'She said: "I will stay in this, the king's castle."'

(15) Khuzestani Arabic (Leitner 2021a: F, 0751)

w   ənnōb tāli   nrawwi   əb-ṣīniyye
and then   next form_balls.IPFV.1PL on-tablet

'...and then we form balls [of dough for baking bread] on the tablet.'

2.6 Auxiliary/main verb

The default position for auxiliary verbs is before the main verb (16, 17).

(16) Khuzestani Arabic (Leitner 2021a: F, 0785)

gabul   ma čānu   ystəfādūn   ləbləbi
formerly NEG AUX.3PL.M use.IPFV.3PL.M chick_peas

'In former times they didn't use chick peas [for cooking].'

(17) Khuzestani Arabic (Leitner 2021a: A, 0187)

ġarafna   l-əl-hōr   gəmna ənḥušš
row.PFV.1PL to-DEF-marshland AUX.1PL cut_grass.IPFV.1PL

'We rowed to the marshland [*hōr*], we started to cut grass.'

However, as has been suggested in Leitner (2020, 2022a), there seems to be an ongoing change probably triggered by contact with Persian that yields clause-final position of the auxiliary, cf. also the following example (18). This development is paralleled by the tendency towards postpredicate position of copulas (cf. Section 3 below; Leitner 2022a). Of course, it may never be entirely ruled out that it is rather pragmatic reasons that cause some of the postpositions of the auxiliary (e.g. as a time frame setter, cf. Brustad 2000), but the comparative numbers presented in Leitner (2022a) speak rather for an explanation as a contact feature.

(18) a. Khuzestani Arabic (own data)

hāde   ham mən zuġur   yəštəġəl   čān
DEM.SG.M also from childhood work.IPFV.3SG.M AUX

b. New Persian (own data)

in   ham az   kudeki   kār   mi-kard
DEM.SG also from childhood work PROG-do.PST.3SG

'This one has also been working from childhood on.'

2.7 Complement clause/matrix verb

Complement clauses follow the matrix verb and a complementizer *əlli* (19) or *ənnu* 'that' (20). In general, however, the complementizers are often omitted and asyndetic constructions are preferred as in (21).

(19) Khuzestani Arabic (own data)
      ətgūl   *əlli lyōm mā ətrūḥ*   *l-əš-šəġəl*
      say.IPFV.3SG.F that today NEG go.IPFV.3SG.F to-DEF-work
      'She says that today she won't go to work.'

(20) Khuzestani Arabic (own data)
      w   maʕrūf ənnu məṭəl
      and known that for_example
      'And [it is] known that for example…'

(21) Khuzestani Arabic (own data)
      gāl   *əlyōm māku*   *ṭalʕa*
      say.PFV.3SG.M today EXIST.NEG going_out
      'He said that today there is no going out.'

2.8 Nominal direct object/verb

The default or unmarked word order is VO as in (22). This order appears 274 times out of a total of 317 direct objects in the WOWA-KhA-corpus. 43 tokens are found in pre-predicate position (OV). Of these 43 OV-constructions, 27 show a resumptive pronoun (co-referential with the object) after the verb as in (23) and only 16 had no resumptive pronoun after the verb, e.g. (24). The 27 examples of OV + resumptive pronoun thus are not plain OV constructions, but instead cases of topicalization (as further discussed in Section 3).

(22) Khuzestani Arabic (Leitner 2021a: C, 0434)
      šəfət-l-i   *maʕǧiza*
      see.PFV.1SG-for-1SG miracle
      'I saw a miracle.'

(23) Khuzestani Arabic (Leitner 2021a: C, 0392)
      əl-ḥaywāna   nəḥlib-ha
      DEF-animal.SG milk.IPFV.1PL-3SG.F
      'We milk the cattle.'

(24)   Khuzestani Arabic (Leitner 2021a: A, 0100)

ləbasne   əxwīəṣāt-ne   ləbasne   əḥžīəlāt-ne   yaʕni
wear.PFV.1PL ring.DIM.PL-1PL wear.PFV.1PL bracelet.DIM.PL-1PL DM

šīəla-t   balbūl ləbasne   yaʕni əṭ-ṭōg   u-māṣxa
shawl.DIM-CS balbūl wear.PFV.1PL DM   DEF-necklace and-māsxa

ləbasne
wear.PFV.1PL

'… we put on our rings, we put our bracelets, … we wore the *balbūl* shawl[4], [and] the necklace and *māsxa* [kind of jewelry].'

2.9 Pronominal direct object/verb

Pronominal direct objects are generally suffixed to the verb and thus inherently postverbal as in (25) and (26). Only in cases where the speaker wants to express additional emphasis and/or mark it as the topic of an utterance (cf. Brustad 2000: 331, 333 on comparable examples of independent pronouns that are sentence-initial and the topic but not subject of a sentence), an independent pronoun may additionally be mentioned preceding the verb with the suffixed pronoun, see example (27) (and the discussion of such examples in Section 3).

(25)   Khuzestani Arabic (Leitner 2021a: C, 0356)

təsmaʕ-ni
hear.IPFV.2SG.M-1SG

'You hear me.'

(26)   Khuzestani Arabic (own data)

hāy ətfəhm-əč   səʔli-ha   suʔāl
DEM understand.IPFV.3SG.F-2SG.F ask.IMP.SG.F-3SG.F question

'She understands you. Ask her a question!'

(27)   Khuzestani Arabic (Leitner 2021a: C, 0412)

waḷḷa   āna iyā-ni   hād   əl-bīəhdāš[5]   māl
by_god 1SG come.PFV.3SG.M-1SG DEM.SG.M DEF-healthcare_center GL.SG.M

salf-i
district-1SG

'And I – He came to me, [from] this healthcare center of my district…'

---

[4]A thin shawl, lit. "made of (the material) *balbūl*", cf. Steingass (2001: 179) on the Persian term *bulbul čašm* 'a sort of silk'.

[5]< Pers. *behdāšt* 'hygiene, healthcare' (Junker & Alawi 2002: 108).

2.10 Goal/verb

The default position of goals is post-predicate (28).

(28) Khuzestani Arabic (Leitner 2021a: C, 0464)
ham rəḥna   ən-naxal   rəḥna   l-əš-šilib
also go.PFV.1PL DEF-palm_groves go.PFV.1PL to-DEF-rice_fields
'We also went to the palm groves, we went to the rice fields...'

From the 83 goals in the WOWA-corpus only five were in pre-predicate position: three times the adverb *hnā* 'here' as in (29), once *əb-baṭn-a* 'in its belly' (30) and once the 1SG pronoun *āna*, which is however indicated as well by a pronominal affix on the verb and added sentence-initially for emphasis and marking it as a sentence topic (see example (38) below and the discussion on whether it should be considered a pre-predicate token in Section 3).

(29) Khuzestani Arabic (Leitner 2021a: A, 0091)
wa hnā yō ḥaṭṭan   warde   yō əzmām
and here or put.PFV.3PL.F nose_ring or nose_ring
'And here they put a *warde* or a *zmām* [two types of nose rings].'

(30) Khuzestani Arabic (Leitner 2021a: F, 0739)
nṭēḥ   bə-diyāy   əb-baṭn-a   yḥuṭṭūn   šwayyūn təmən
fall.IPFV.1PL with-chicken in-belly-3SG.M put.IPFV.3PL.M some   rice
w   kəšməš w   fəlfəl   ʔaswad w   l-ḥawār
and raisins and pepper black   and DEF-spices
'We take the chicken, in its belly we put some rice and raisins and black pepper and spices.'

2.11 Other obliques/verb

In the WOWA-corpus, of all 25 obliques labelled as "other" (which are mostly adverbs) 6 were found in pre-predicate position as in the following example (31):

(31) Khuzestani Arabic (Leitner 2021a: A, 0095)
waḷḷa   dāyman hēč   mā   nilbas
by_god always   like_this NEG dress.IPFV.1PL
'It's not always that we dress like this.'

2.12 Copular and become-constructions

In general, there is no present tense copula in Arabic. Several dialects have, as a consequence of contact with languages that do have obligatory present tense copulas, developed an obligatory copula for the imperfective (cf. e.g. Procházka 2019 on the dialects of Eastern Anatolia), but KhA has not despite its long term contact with Persian.

The WOWA-corpus features one imperfective copular-complement in pre-predicate position (out of 17 copular constructions[6]), (32), and two pre-predicate become-complements (out of 28), one of them is cited here in (33) (*ha-l-gadd-āt-ha tṣīr* 'it becomes about this size'). The latter structure again might be due to contact influence or due to information packaging and pragmatic reasons, as it seems to be an echo or a recall of the previous *yṣīr čibīr-e* 'it becomes big'.

Even though in the WOWA-corpus we find only one attestation of pre-predicate copulas, we know from previous studies that they are more often found in sentence-final position than other verbs in KhA (see Leitner 2022a).

(32) Khuzestani Arabic (Leitner 2021a: B, 0278)
hāḏanni ṭəlyān ṭəli yṣīr
DEM.PL.F lamb.PL lamb COP
'These are *ṭəlyān* ["lambs"], *ṭəli* ["lamb"] we call it.'

(33) Khuzestani Arabic (Leitner 2021a: D, 0508)
hāy əl-faḥla yṣīr čibīr-e w hāy ən-naxla la
DEM DEF-male_palm become.IPFV.3SG.M big-F and DEM DEF-palm no
zəġīr-e eh zəġīr-e taġrīban ha-l-gadd-āt-ha tṣīr
small-F yes small-F about DEM-DEF-size-PL.F-3SG.F become.IPFV.3SG.F
nəfšəg gadd čaff īd hāy waḥda-t ən-naṭye
split.IPFV.1PL size palm hand DEM thing-CS DEF-female_palm
'This male palm grows big, but this palm not, small, yes small, it becomes about this size, we split – as [big as] a hand – this thing of the female palm.'

3 Areal issues and information structure

As stated in the introduction, KhA is not entirely isolated from the rest of the Arabic-speaking world due to its border with Iraq, although it is spoken in a

[6]Of these 17 copular constructions, 15 copulas were imperfective forms and two were perfective forms; 14 copulas were forms of *ṣār* lit. 'to become' and only three were based on the lexeme *kān* 'to be'.

region in which the sociolinguistically dominant language is not Arabic but Persian. The use of KhA is mostly restricted to conversations within the family and among friends.

To the best of my knowledge, we do not have a detailed survey on word order in (Muslim/*gələt*) Iraqi Arabic, with which KhA is closely related. But since the overall picture of KhA word order shows that most word order features are most likely inherited, as the post-predicate position of objects and other complements is the default position, we can assume that KhA in this regard does not deviate much from its neighbor Iraqi Arabic. Against this background, this Section will focus on those cases of non-default word order trying to propose possible explanations. The following thus is a brief outline of some of the factors that may trigger pre-predicate position in KhA and discusses whether changes in word order and use of marked word order are likely to result from contact with Persian or are rather to be explained as information structural strategies. The latter builds on the theoretical approaches of Brustad (2000: 315–362) and her analysis of information packaging and its influences on word order in spoken Arabic and Ingham (1991) and his analysis of KhA sentence structure. Ingham proposes a basic division of sentence types into i) uninodal, in which new information comes first and which are usually verb-initial unless there is focus fronting (e.g. 36), and ii) binodal, in which the topic precedes new information/the comment. Example (34) illustrates the binodal sentence type (Ingham 1991: 722), where the object is fronted (fronted objects are usually given information and definite nouns[7]) is topical and indexed by a resumptive pronoun marking the original post-verbal position of the object. This example clearly shows how putting an object in pre-predicate position can be used as an information structural tool to indicate the topic of a sentence (cf. Brustad 2000: 348–349). The phrase *hāy əd-dār* 'this house' already appears earlier on in the sentence, but in each instance the speaker clearly refers to another room and the one following the conjunction *bass* 'but' is the one made the topic followed by the new information, vic. that it should not be opened. After the preceding rhythmic enumeration ("You may open this house, and you may open this room, and you may open this room"), the part after *bass* is also accompanied by a differing intonation contour (pitch goes up with *dār*), and a short pause before she goes on saying *lā thəddin-ha* '(this one) don't you open it'. Within the latter phrase, we find the feminine singular object pronoun *-ha* referring back to the object *hāy əd-dār* 'this house', thus the construction resembles that of a normal topicalization structure as found in all varieties of Arabic.

---

[7]Cf. Brustad (2000: 339) and the examples provided there on the fact that topic in Arabic also includes temporal verbs.

(34) Khuzestani Arabic (Leitner 2021a: E, 0635)
      gāl-ha   thəddīn   hāḏ əl-bīət   w   thəddīn
      tell.PFV.3SG.M-3SG.F open.IPFV.2SG.F DEM DEF-house and open.IPFV.2SG.F
      hāy əd-dār   w   thəddīn   hāy əd-dār   bass hāy əd-dār
      DEM DEF-room and open.IPFV.2SG.F DEM DEF-room but DEM DEF-room
      lā   thəddīn-ha
      NEG open.IPFV.2SG.F-3SG.F
      'He said: "You [may] open this house, and you [may] open this room, and
      you [may] open this room, but don't you open this room."'

Out of the 43 OV instances in the KhA data from the WOWA-corpus, 27 fea-
tured a resumptive pronoun on the verb as in the example above or in example
(23). As stated above, this structure is not foreign to Arabic and must not be at-
tributed to contact influence. Whether or not such topicalized sentences appear
more commonly in Arabic dialects that start shifting towards OV due to contact
with an OV language has yet to be determined.

The WOWA-corpus of KhA contains 16 OV phrases that do not feature a re-
sumptive pronoun on the verb such as (24) above cited here again (35) for the
sake of the discussion:

(35) Khuzestani Arabic (Leitner 2021a: A, 0100)
      ləbasne   əxwīəṣāt-ne   ləbasne   əḥžīəlāt-ne   yaʕni
      wear.PFV.1PL ring.DIM.PL-1PL wear.PFV.1PL bracelet.DIM.PL-1PL DM
      šīəla-t   balbūl ləbasne   yaʕni aṭ-ṭōg   u-māṣxa
      shawl.DIM-CS balbūl wear.PFV.1PL DM   DEF-necklace and-māsxa
      ləbasne
      wear.PFV.1PL
      '… we put on our rings, we put our bracelets, … we wore the *balbūl* shawl,
      [and] the necklace and the *māsxa* [kind of jewelry].'

In this example, in which the speaker (rhythmically) lists various items that
women used to wear in the past for weddings, she switches from the default VO
structure to the marked OV in the middle of the sentence (note that this turn is
introduced by the discourse marker *yaʕni*). The shift in word order is paralleled
by a shifted stress distribution in the second (OV) part of the sentence where
in both cases the speaker puts the main stress on the verb *ləbasne* 'we wore',
whereas in the first part the main stress lies on the objects. Information packag-
ing seems to be the most likely cause for such a change towards marked word
order. According to Brustad (2000: 343) "objects that are contrastive may occupy

pre-verbal position (OVS)" and later on adds that "objects without resumptive pronouns are highly contrastive" (Brustad 2000: 348). In this example, such a contrast may lie in the shift from the event-oriented first part (with a focus on the habitual event, vic. what they used to wear) towards a topic-oriented second part (with a focus on items that are part of a set of things they used to wear). In Ingham's (1991) terminology, the latter part of this sentence seems to be of the type "uninodal with focus fronting" (Ingham 1991: 721–722) and is thus also not new or undocumented for Arabic dialects. However, against the definition of this sentence type, the nuclear stress in this very example does not fall on the fronted item but on the following verb. It remains unclear, why the speaker once marks a pre-predicate object with the definite article (*əṭ-ṭōg*, albeit the only pre-predicate direct object with definite article in the WOWA corpus) and once without (*māṣxa*), as both seem to have a generic character representing a set of items. In general this example does not seem to support Ingham's (1991) assumption that possibly a "definite 'true' object" cannot appear in KhA in preverbal position without a resumptive pronoun (Ingham 1991: 722, Fn. 5).

Another instance of a pre-predicate object (*kəllšī* 'everything') without resumptive pronoun is the following (36):

(36) Khuzestani Arabic (Leitner 2021a: C, 0363)
      ḥaṣadna   lammēna   təbən lammēna   ğanēna
      harvest.PFV.1PL gather.PFV.1PL straw gather.PFV.1PL breed.PFV.1PL
      ḥaywān kəllšī   sawwēna   ya ʕazīz galb-i
      cattle   everything do.PFV.1PL VOC dear heart-1SG
      'We harvested, gathered straw, we gathered – we bred cattle, we made everything, my dear.'

This again is a uninodal sentence with focus fronting, in this case also fulfilling the requirement of heavy stress on the fronted object (*kəllšī* 'everything'). The indefinite pronoun *kəllšī* appears three times in the WOWA-corpus and always in pre-predicate position. It may also precede negated verbs as in the following example (37), in which we find additionally a fronted and topicalized object *əl-ʕarūs* (indexed by the resumptive pronoun *-hən* on the verb):

(37) Khuzestani Arabic (Leitner 2021a: A, 0072)
      əl-ʕarūs   kəllšī   mā nsawwī-l-hən
      DEF-bride everything NEG make.IPFV.1PL-DAT-3PL.F
      'The bride[s], we didn't do anything with them [like putting on henna, etc.].'

As stated above for example (8), where *kəllšāy* 'everything' precedes a negative possessive-construction, these structures may be calques on (Spoken) Persian structures such as *hičči na-kardam* 'we did nothing' and *har kari kardim* 'we did everything'. Their existence in Iraqi Arabic (see Section 2.2 above) might speak against this, as influence of Persian on Iraqi Arabic is mostly restricted to the lexical domain. Taking all this under consideration, it seems most likely to be an inherited structure in KhA, focus fronting of *kəllšāy* ~ *kəllšī* that indicates contrast[8], or, as in (36), some kind of closure of an enumeration (a category or context not mentioned by Brustad or Ingham). Of course, Persian word order might have reinforced the use of this structure and increased the frequency of fronting of *kəllšī*.

The following sentence (38) (already cited above as (27) and repeated here for the sake of convenience) is most likely also of the binodal type with the 1SG pronoun *āna* presenting the sentence topic and being taken up by a referential pronoun on the verb.

(38) Khuzestani Arabic (Leitner 2021a: C, 0412)
     waḷḷa āna iyā-ni   hād   əl-bīəhdāš[9]   māl
     by_god 1SG come.PFV.3SG.M-1SG DEM.SG.M DEF-healthcare_center GL.SG.M
     salf-i
     district-1SG
     'And I – He came to me, [from] this healthcare center of my district …'

A feature that can more likely by attributed to contact influence is the sentence-final position of auxiliaries (cf. Section 2.6 for examples), copulas and the verb 'to become' (cf. Section 2.12 for examples). This fits very well into the stages of shift towards XV structures as described by El Zarka & Ziagos (2020) for Southern Iranian Arabic. There, this shift seems to be more advanced than in KhA, but in both varieties it appears to have started with elements such as copulas and auxiliaries (see Leitner 2022a).

We can thus conclude that, overall, the data and its analysis clearly show that the inherited default word order (VX) is retained in Khuzestani Arabic. However, pragmatic factors related primarily to information structure and to a much lesser degree contact with Persian may cause that elements are moved to pre-predicate position, thus yielding XV word order. For future studies on KhA word order and information packaging, it would be interesting to further include a comparison of information structure in Persian and KhA.

---

[8]In example (37) the speaker was clearly contrasting wedding traditions of the past and the present.

[9]< Pers. *behdāšt* 'hygiene, healthcare' (Junker & Alawi 2002: 108).

Abbreviations

AUX	auxiliary	IPFV	imperfective
COLL	collective (noun)	KhA	Khuzestani Arabic
COP	copula (noun)	M	masculine
CS	construct state	MSA	Modern Standard Arabic
DAT	dative	NEG	negation
DEF	definite (article)	Pers.	Persian
DEM	demonstrative	PFV	perfective
DIM	diminutive	PL	plural
DM	discourse marker	PRO	pronoun
EXIST	existential particle	PROG	progressive marker
F	feminine	PST	past tense
GL	genitive linker	SG	singular
HORT	hortative particle	VOC	vocative particle
IMP	imperative		

References

Blanc, Haim. 1964. *Communal dialects in Baghdad.* Cambridge, MA: Harvard University Press.

Brustad, Kristen. 2000. *The syntax of spoken Arabic: A comparative study of Moroccan, Egyptian, Syrian, and Kuwaiti dialects.* Washington: Georgetown University Press.

El Zarka, Dina & Sandra Ziagos. 2020. The beginnings of word order change in the Arabic dialects of Southern Iran in contact with Persian: A preliminary study of data from four villages in Bushehr and Hormozgan. *Iranian Studies* 53(3–4). 465–488. DOI: 10.1080/00210862.2019.1690433.

Field, Henry. 1939. *Contributions to the anthropology of Iran.* Chicago: Museum of Natural History.

Gazsi, Dénes. 2011. Arabic-Persian language contact. In Stefan Weninger, Janet C. E. Watson, Michael Streck & Geoffrey Khan (eds.), *The Semitic languages: An international handbook*, 1015–1021. Berlin: De Gruyter.

Ingham, Bruce. 1973. Urban and rural Arabic in Khūzistān. *Bulletin of the School of Oriental and African Studies* 36. 523–553.

Ingham, Bruce. 1976. Regional and social factors in the dialect geography of Southern Iraq and Khuzistan. *Bulletin of the School of Oriental and African Studies* 39. 62–82.

Ingham, Bruce. 1991. Sentence structure in Khuzistani Arabic. In Alan S. Kaye (ed.), *Semitic studies: In honor of Wolf Leslau on the occasion of his eighty-fifth birthday, November 14th, 1991*, vol. 1, 714–728. Wiesbaden: Harrassowitz.

Ingham, Bruce. 1997. *Arabian diversions: Studies on the dialects of Arabia.* Reading: Ithaca Press.

Ingham, Bruce. 2008. Najdi Arabic. In Kees Versteegh (ed.), *Encyclopedia of Arabic language and linguistics*, vol. 3, 326–334. Leiden: Brill.

Junker, Heinrich F. J. & Bozorg Alawi. 2002. *Persisch-Deutsch: Wörterbuch, 9. unveränd. aufl.* Wiesbaden: Harrassowitz.

Leitner, Bettina. 2020. Khuzestan Arabic. In Christopher Lucas & Stefano Manfredi (eds.), *Arabic and contact-induced change*, 115–134. Berlin: Language Science Press.

Leitner, Bettina. 2021a. Arabic (Khuzestan). In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* Bamberg. http://multicast.aspra.uni-bamberg.de/resources/wowa/ (6 March, 2023).

Leitner, Bettina. 2021b. New perspectives on the urban–rural dichotomy and dialect contact in the Arabic Gələt dialects in Iraq and South-West Iran. *Languages* 6(4). 198. DOI: 10.3390/languages6040198. https://www.mdpi.com/2226-471X/6/4/198.

Leitner, Bettina. 2022a. Clause-final verbal copulae in Khuzestan Arabic and other Eastern Arabic varieties. In *Proceedings of the 13th International Conference (June 10-13, 2019) 'Association Internationale de Dialectologie Arabe (AIDA)'*, 165–175. Kutaisi: Akaki Tsereteli State University.

Leitner, Bettina. 2022b. *Grammar of Khuzestani Arabic: A spoken variety of South-West Iran* (Studies in Semitic Languages and Linguistics 108). Berlin: Brill.

Leitner, Bettina, Fady German & Stephan Procházka. 2021. *Lehrbuch des Irakisch-Arabischen: Praxisnaher Einstieg in den Dialekt von Bagdad.* German & Arabic (Semitica Viva 7). Wiesbaden: Harrassowitz.

Matras, Yaron & Maryam Shabibi. 2007. Grammatical borrowing in Khuzistani Arabic. In Yaron Matras, Georg Bossong & Bernard Comrie (eds.), *Grammatical borrowing in cross-linguistic perspective*, 137–149. Berlin: De Gruyter Mouton.

Nadjmabadi, Shahnaz Razieh. 2009. The Arab presence on the Iranian coast of the Persian Gulf. In Lawrence G. Potter (ed.), *The Persian Gulf in history*, 129–145. New York: Palgrave Macmillan US. DOI: 10.1057/9780230618459_7.

Nejatian, Mohammad Hossein. 2015. Ahvaz iv. Population, 1956-2011. In *Encyclopaedia Iranica online edition*. New York. http://www.iranicaonline.org/articles/ahvaz-04-population.

Oppenheim, Max Freiherr von. 1967. *Die Beduinen: 4,1: Die arabischen Stämme in Chūzistān (Iran). Pariastämme in Arabien.* Leipzig: Harrassowitz.

Procházka, Stephan. 2019. The Arabic dialects of Eastern Anatolia. In Geoffrey Haig & Geoffrey Khan (eds.), *The languages and linguistics of Western Asia: An areal perspective*, 159–189. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110421682-005.

Savory, R. M. 1986. Khūzistān. In Peri J. Bearman, Thierry Bianquis, Clifford E. Bosworth, Emericus J. van Donzel & Wolfhart P. Heinrichs (eds.), *Encyclopaedia of Islam, second edition*, vol. V, 80–81. Leiden: Brill.

Steingass, Francis Joseph. 2001. *A comprehensive Persian-English dictionary: Including the Arabic words and phrases to be met with in Persian literature being Johnson and Richardson's Persian, Arabic, and English dictionary.* London/New York: Routledge. https://dsal.uchicago.edu/dictionaries/steingass/.

Zarrinkūb, Abd Al-Husain. 1975. The Arab conquest of Iran and its aftermath. In Richard N. Frye (ed.), *The Cambridge history of Iran*, vol. 4, 1–56. Cambridge: Cambridge University Press.

Chapter 15

Neo-Aramaic in Iran and northeastern Iraq

Paul M. Noorlander
University of Cambridge

This chapter offers a brief overview of the word order typology of Neo-Aramaic dialects spoken by Jewish and Christian minorities of Iran and northeastern Iraq. A characteristic of the dialects in this region is the contact-induced shift from VO to OV word order under the influence of neighbouring Iranian and Turkic languages. In Iranian Azerbaijan, convergence with Azeri has resulted in an additional increase in Adjective-Noun order, and a different treatment of Addressees from Goals. In many respects, however, the constituent order remains consistent with that of so-called VO languages, such as prepositional marking and Noun-Genitive order.

1 Introduction

Aramaic[1] is a Semitic language that has been attested in writing since the first millennium BC and used to be spoken more widely in West Asia. The modern Aramaic vernaculars in Iran and northeastern Iraq mainly belong to the North Eastern Neo-Aramaic (NENA) subgroup.[2] Another relevant Neo-Aramaic subgroup, known as Mandaic, spoken by the Mandaeans of southwestern Iran and southern Iraq (e.g. Häberl 2011) lies beyond the scope of this chapter.

---

[1]The orthography has been adapted slightly to normalize transcription across dialects. Khan's (2016) <c>, <ɟ>, <ḳ> correspond to <k>, <g> and <q> here. Panoussi's (1990) <e> and Khan's (2004) <ɪ̆> both correspond to <ə> here. Garbell (1965b) and Khan's (2008b) <o> and <u> for /ø/ and /y/ in Jewish Urmi respectively correspond to <ö> and <ü> here. Superscript ⁺ indicates the following word or syllable is pronounced with additional velarization or pharyngealization.

[2]For the closely related Central Neo-Aramaic group in Anatolia, see Noorlander (2024 [this volume]).

NENA comprises a continuum of highly diverse and severely endangered dialects of Jewish (J.) and Christian (C.) communities that used to span an area from western Iran to southeastern Turkey. Most of the Jewish dialects are extinct or border extinction, and only a rapidly diminishing number of elderly speakers–generally known as *kurdim*–reside in Israel today. Apart from their regional identification, e.g. *sənaye* 'people from Səna, i.e. Sanandaj,' *'urməžnaye* 'people from Urmi, i.e. Urmia,' the Christian speakers self-identify as *suraye* 'Syrian Christian' and refer to their language as *surət* 'Syriac'. The Christians belong to various denominations, primarily the Chaldean Catholic Church and the Assyrian Church of the East, which may or may not coincide also with their linguistic and ethnic identification, respectively. Since more recent times, however, the self-identification among native speakers in both the homeland and diaspora as the Assyrian people, i.e. *'aturaye*, has extended beyond tribal, religious and geographic affiliations, and the same holds true for Chaldeans, albeit to a lesser extent. The havoc wreaked by the tumultuous 20th century and the recent atrocities in the name of Islamic State in the spread of the Syrian Civil War into Iraq has resulted in the massive displacement of Aramaic-speaking Christians and the destruction of entire villages. Consequently, the vast majority of speakers from Iran no longer reside in their original environment but as diaspora communities in Northern America (San Diego, California & Detroit, Michigan) and Australia (Sydney). Today, the largest Assyrian communities in Iran reside in Urmia and Tehran. The majority of Aramaic speakers in the Middle East, however, is found in Iraqi Kurdistan where the use of the literary *koine* based on the variety of the Urmia county (West Azerbaijan, Iran) has become increasingly widespread and accepted in education, media and sermons.

Figure 1 displays a selection of originally Neo-Aramaic speaking towns in Iran and northeastern Iraq. The Greater Zab River serves as an isogloss for both Jewish and Christian dialects, dividing the Jewish dialects into two major groups: *Lishana Deni*, e.g. Zakho and Duhok, in the west vs. the Trans-Zab Jewish subgroup in the east (Mutzafi 2008), such as Arbel (Erbil, Hewlêr), Urmi (Urmia, Orumiyeh) and Sanandaj (Sine). The dialects around the settlement Barzan represent a transition zone. The Trans-Zab cluster has been heavily influenced by contiguous Iranian languages (e.g. Kapeliuk 2004; Noorlander 2014; Khan 2020), in particular those dialects in the southeastern periphery, in Iranian Kurdistan and Kermanshah. The Trans-Zab Jewish dialects in the north also had outposts into southeastern Turkey, namely Başkale and Gawar (Yüksekova; Garbell 1965b). While the Christian dialects form a continuum from Turkey through Iraq to Iran, clusters can also be recognized in Iranian Azerbaijan and the Iraqi provinces of Erbil and Sulaymaniyah, with, however, only one easternmost outpost Sena
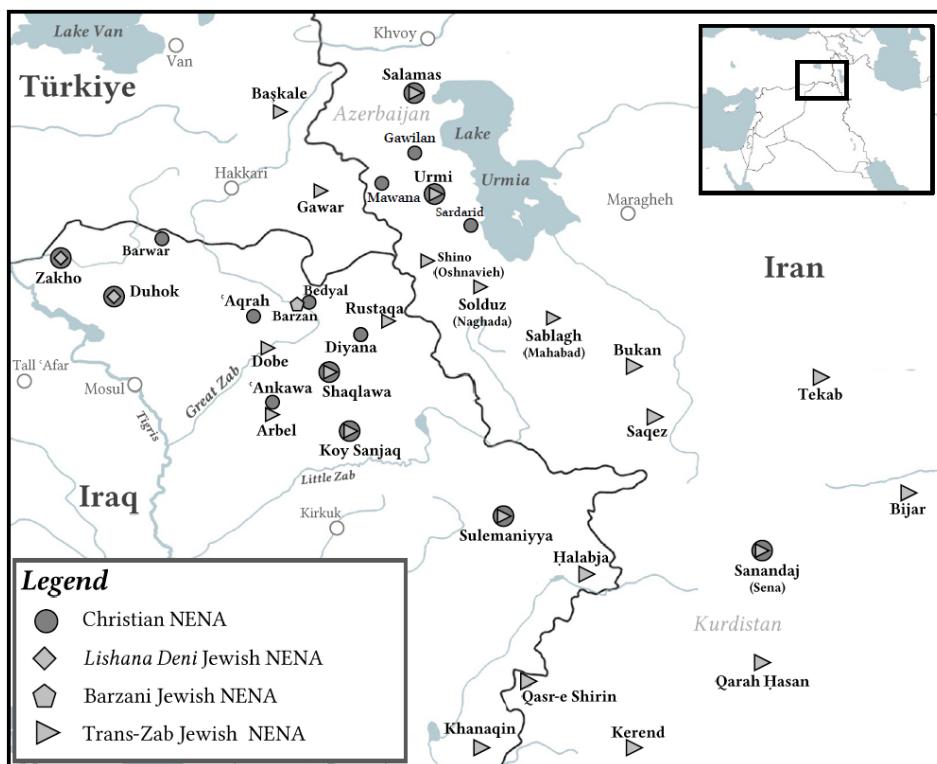
Figure 1: Location of the main Northeastern Neo-Aramaic dialects discussed in this chapter

(Sanandaj) in Iranian Kurdistan. The dialects near the Iraq-Iran-Turkey borders such as Diyana (Soran) constitute a transition zone.

In what follows, I shall focus on the NENA dialects in the eastern periphery whose statistically dominant ordering of subject, object and verb can be characterized as SOV, and where the post-predicate slot is reserved for Goals (for a definition, see §2.2.2), i.e. SVG. The object placement in these dialects is distinct from the typology of (Central) Semitic word order, as well as the majority of Neo-Aramaic dialects in modern-day Turkey and Iraq (see Noorlander 2024 [this volume]). This chapter will show, however, that the same basic, i.e. statistically dominant, word order does not hold to the same degree in every dialect for each argument type, in line with the general rationale of the WOWA project.

Table 1 shows a list of the datasets from the WOWA corpus with their sources[3] and partial metadata used for the analysis of non-subject arguments and their

---

[3]Numbered texts and numbered segments are separated by colons, e.g. 25:§2 means Text 25, Paragraph 2, and page numbers and segments by periods, e.g. 101.§2, Page 101, Paragraph 2, page numbers and lines by dots, e.g. 101.1, Page 101, Line 1.

respective position before or after the predicate in accordance with the framework and coding guidelines of the WOWA databank.[4] A handful of additional data were taken from Panoussi (1990: 120–128) for Christian Sanandaj, which is not part of the WOWA corpus. Concerning the J. Urmi doculect based on Khan (2008b) and C. Urmi doculect based on Khan (2016), it is likely that these dialects cannot be taken as representatives of NENA of Iranian Azerbaijan as a whole. Even a cursory glance at the material collected by Garbell (1965b) and Hopkins (1989) suggests that there seems to be more variation, and the same holds true for the Jewish and Christian dialects of Salmas documented by Duval (2009) and Tsereteli (1976). The final results from other texts could, therefore, be different and approximate more closely the typology of the NENA varieties elsewhere.

Hopkins (1999), Khan (2012a, 2019a,b), and Noorlander (2021: 100–206) provide general overviews of the Trans-Zab Jewish NENA dialects, especially in Iran. Gutman (2018)[5] provides a comparative overview of Noun-Genitive orders, Noorlander & Molin (2022) an overview of Verb-Object and Verb-Oblique, Khan (2020: 398–401) that of Auxiliary-Verb and Verb-Object. Most grammatical descriptions do not discuss word order in detail, except for Khan's voluminous grammars (Khan 2008b, 2016) and Coghill (2018), though, apart from Molin (2021) and Noorlander & Molin (2022), no statistics are provided. Nevertheless, for virtually all NENA dialects considered here, which pattern constitutes the basic word order is largely unquestioned, except for Christian Urmi and Sardarid (see §2.2.1).

The following sections provide a general overview of word order for which a synopsis is offered in Table 2, where plus (+) corresponds to placement after the head, and minus (–) corresponds to placement before the head, respectively. Word order configurations in NENA, however, are sensitive to pragmatic effects not coded in the WOWA corpus, e.g. any argument can undergo focalization to the immediately pre-verbal position or topicalization to clause-initial position (e.g. Noorlander & Molin 2022: 243–245; the difference between definite and indefinite arguments is coded, however, see §2.2.1.). For our purpose, word order patterns of the clause will be identified on the basis of frequency, as discussed in Haig et al. (2024 [this volume]), in line with (Dryer 2007: 73–78). Finally, we use the basic surface-syntax-based parameters coded for the WOWA corpus rather than pragmatic or formal criteria of movement.

---

[4] See https://multicast.aspra.uni-bamberg.de/resources/wowa/data/_docs/guidelines/wowa_-coding-guidelines.pdf.

[5] See especially p. 143 for J. Zakho, pp. 220–230 for J. Urmi, pp. 232–234 for J. Sanandaj, p. 291 for Kurdish. Compare also Cohen (2012) on Jewish Zakho.

Table 1: NENA datasets from the WOWA corpus discussed in this chapter

Doculect	Speakers	Total tokens	Analysed tokens	Source
J. Sanandaj	4	2837	1184	Noorlander 2022c based on Khan 2009
J. Urmi	1	923	502	Noorlander & Stilo 2022 based on Khan 2008b: 398–439
C. Urmi	2	865	724	Noorlander 2022b based on Khan 2016: Texts A2, A39
C. Shaqlawa	3	524	444	Noorlander 2022a based on Khan et al. 2022: Texts 4, 23 and 35

Table 2: Overview of dominant configurations

Doculect	VO	VAddr	VGoal	NGen	NAdj	AdjSt	CopPred
C. Barwar	+	+	+	+	+	+	+/−
C. Shaqlawa	−	+	+	+	+	+/−	+/−
C. Urmi	−	−	+	+	+	+/−	+/−
C. Sanandaj	−	+	+	+	+	(?)	−
J. Sanandaj	−	+	+	+	+	+/−	−
J. Urmi	−	−	+	+	+/−	−	−

2 Word order profile

2.1 Noun phrases

2.1.1 Determiner/noun

Across all doculects of NENA considered here, Demonstrative-Numeral-Noun order predominates, e.g.

(1) Demonstrative-Numeral-Noun
     C. Urmi (Khan 2016: A55:§7)
     ʾannə   tré   ʾojaxə
     DEM.PL two clan.PL
     'these two clans'

   Noun-Demonstrative-Adjective order also occurs, particularly in the expressions of 'the elder' or 'the eldest':

(2) Noun-Adjective-Demonstrative-Adjective
     C. Urmi (Khan 2016: A1:§29)
     kačala   ʾasli   ʾo   ⁺gur-a
     bald.person.M.SG original DEM.M.SG big-M.SG
     '**the** elder, original, baldhead'

2.1.2 Noun/attribute

Attributes, such as adjective phrases, follow the head noun they modify, e.g.

(3) Noun-Adjective
     J. Sanandaj (Khan 2009: B:§58)
     knəšta   rab-ta
     synagogue.SG.F big-SG.F
     'a big synagogue'

(4) Noun-Adjective
     C. Sanandaj (Panoussi 1990: 125.§6)
     ṣoma   rab-a
     fast.M.SG big-M.SG
     'the great fast'

(5) Noun-Modifier-Adjective
C. Urmi (Khan 2016: A42: §34)
brata   ʾuxča šapə̂r-ta
girl.SG.F such beautiful-SG.F
'**such a** beautiful girl'

Adjective-Noun order, e.g. (6), if tolerated, is pragmatically restricted–generally increasing the attribute's emotional significance–and its higher rate of occurrence is area-specific, namely specific to Iranian Azerbaijan, see §3.2.5 for discussion.

(6) Modifier-Adjective-Noun
C. Urmi (Khan 2016: A3:§81)
*ʾuxča **šapir-a**   qal-a*
such beautiful-M.SG voice-M.SG
'such a **beautiful** voice'

The primary adjectives denoting relative size, i.e. 'small' and 'big,' tend to remain closer to the head noun (cf. Khan 2016II: 44):

(7) J. Koy Sanjaq (Mutzafi 2004: 202.§26)
*xa belá   **ruww-á** jwan*
a house.M.SG big-M.SG beautiful
'a beautiful, **large** house'

(8) J. Sulemaniyya (Khan 2004: R:§144)
*bela   **ruww-á** hulaʾ-á*
house.M.SG big-M.SG Jewish-M.SG
'a **big** Jewish house'

Genitive constructions show Noun-Genitive order (see Gutman 2018, especially Chapter 4, for an overview and recent analysis of NENA genitive constructions). Prepositions similarly also serve as heads, e.g. J. Koy Sanjaq *qam-əd ʔod=belá* lit. front-of of=house 'in front of the house' (Mutzafi 2004: 175). Juxtaposition can also suffice, e.g. C. Diyana-Zariwaw *šəmma sawun-i* lit. name grandfather-my 'the name of my grandfather' (Napiorkowska 2015: 315).

(9) Noun-Genitive
C. Sardarid (Younansardaroud 2001: 13:§5)
šəmm-əd   d-o   naša
name-CSTR GEN-DEM.M.SG man.M.SG
'the name of that man'

(10)   Noun-Genitive
       J. Koy Sanjaq (Mutzafi 2004: 1B:§18)
       šuḷṭan-əd   ʾod=ḥaywan-é
       king-CSTR LINK=animal-PL
       'the king of the animals'

Genitive-Noun order is restricted, also known as the emotive genitive,[6] intensifying the speaker's emotional attitude, e.g.

(11)   Genitive-Noun
       C. Diyana (Napiorkowska 2015: 18.1:§35)
       ala   munix-əd   xəmyan-i
       God.M.SG give.rest.PTCP.M.SG-CSTR uncle.M.SG-my
       'my late uncle'

Other constituents can intervene between head and genitive, as shown in (12a). Pronominal possessors are suffixed directly to the head or expressed by a following independent genitive, e.g. (12b).

(12)   a.   Coordination, Noun-Genitive
            C. Diyana (Napiorkowska 2015: 18.7:§16)
            šop-əd   ʾaqle ʾu ʾəd=xzür-u
            print.PL-CSTR foot.PL and LINK=pig.PL-their
            'their (lit. the) footprints and also [those] of their piglets'
       b.   free pronoun, Noun-Genitive
            C. Diyana (Napiorkowska 2015: 18.6:§7)
            *dost-əd   **did-i***
            friend.M.SG-CSTR GEN-my
            'a friend of **mine**'

2.2 Verbal complements

2.2.1 Object/verb

While Verb-Object predominates in NENA dialects, the dialects in Iran and northeastern Iraq generally show Object-Verb order. Earlier treatments of the Christian dialects in Iran identified no primary word order on the basis of frequency (Younansardaroud 2001: 209; Khan 2020: 398–401). As shown in Table 3, when

---

[6]See Hopkins (2009), Cohen (2012: 100–102), Gutman (2018: 143, 182,315).

Table 3: Rate of post-predicate (PP) objects

Doculect	Object	
	n	**PP**
C. Urmi	258	16%
C. Shaqlawa	108	12%
J. Sanandaj	386	5%
C. Sanandaj	50	4%
J. Urmi	172	1%

lumping all types of direct objects together, the statistically dominant order over-
all in the NENA doculects considered here is OV. Table 3 gives the general nu-
merical data for direct object placement in the NENA doculects in Iran as well as
C. Shaqlawa (NE Iraq), excluding *wh*-elements.

Different types of objects, however, should be considered in their own right,
drawing on the distinctions made in the WOWA data (also possessums, see §3.2.1).
Table 4 gives the statistics for direct objects divided in accordance with the addi-
tional variables of definiteness and pronominal categories coded in the WOWA
corpus, which are illustrated in (13–14). "Pronoun," here, includes personal and

Table 4: Rate of post-predicate (PP) objects divided by definiteness and
argument type

Doculect	Indefinite nominal		Other		Definite nominal		Pronoun	
	n	**PP**	**n**	**PP**	**n**	**PP**	**n**	**PP**
C. Urmi	49	49%	19	16%	153	8%	37	8%
C. Shaqlawa	45	13%	7	29%	49	8%	7	29%
J. Urmi	65	0%	13b	8%	87	1%	13	0%
C. Sanandaj	32	3%	–	–	18	6%	(1	0%)
J. Sanandaj	244	3%	20[a]	5%	82[b]	7%	18	0%

[a]This also includes arguments bound as a possessor to the nominal element of light verb con-
structions.

[b]The four tokens with the idiomatic phrase 'May God give X rest' with VO order in J. Sanandaj
have been excluded here (see Noorlander & Molin 2022).

demonstratives, both bare and prepositional, such as (13b) and (14b), but excludes indefinite and reflexive pronouns, which are subsumed under "Other," such as (13c) and (14c). The number of pronominal tokens is, however, relatively low, especially in the case of C. Shaqlawa and C. Sanandaj. It is thus not possible to draw any conclusions about these two dialects without more tokens. Moreover, bound pronominal objects are more common than their independent counterparts in NENA (see Noorlander & Molin 2022).

(13)   a.   Nominal definite (flagged), Object-Verb
            J. Sanandaj (Khan 2009: A:§18)
            *ʾay   broná   **həl-d-ay**   **bratá**   gbe-Ø*
            DEM.SG boy.M.SG DOM-GEN-DEM.SG girl.SG.F IND.want-A.3SG.M
            'The boy loves **the girl**.'

       b.   Pronominal, Object-Verb
            J. Sanandaj (Khan 2009: C:§3)
            *ʾaná ʾ**ea** šmi-li   mən Bahrám*
            I   this heard.PFV-A.1SG from Bahram
            'I heard **this** from Bahram.'

       c.   Other, Object-Verb
            J. Sanandaj (Khan 2009: A:§48)
            kŭl-e   *kalw-ắ-wa-le*
            all-of.it.SG.M write-A.3PL-PST-O.3SG.M
            'They would write **everything** down.'

(14)   a.   Nominal indefinite, Verb-Object
            C. Urmi (Khan 2016: A39:§42)
            Ø-mayy-ət   ⁺raba goz-ə
            SBJV-bring-A.2SG.M many walnut-PL
            'You should bring many walnuts.'

       b.   Pronominal (flagged), Object-Verb
            C. Urmi (Khan 2016: A2:§25)
            *ʾatən   **qa-diyyi**   bət-⁺qaṭl-ət.ˈ*
            you.SG.M DOM-GEN.1SG FUT-kill-A.2SG.M
            'You shall kill **me**.'

       c.   Other, Object-Verb
            C. Urmi (Khan 2016: A2:§35)
            gan-o   *⁺rupp-a-la   ⁺ʾal-sepa*
            REFL-3SG.F threw.PFV-O.3SG.F-A.3SG.F on-sword.SG.M
            'She threw **herself** onto the sword.'

All else being equal, Table 4 demonstrates that OV order has grammaticalized completely in the Jewish doculects as well as Christian Sanandaj. J. Urmi, as represented in Khan (2008b), seems to have the most rigid kind of OV. It is possible, however, that Jewish NENA doculects of Iran collected by Garbell (1965b) and Hopkins (1989), although predominately OV, contain a higher rate of postpredicate Os than Khan (2008b). The higher rate of independent pronouns in Iranian Azerbaijan, namely J. and C. Urmi, may well be due to contact with Azeri. If OV order was completely grammaticalized in Christian Urmi, we would expect a rate similar to that in Jewish Urmi. Definiteness, however, is a major factor in object placement in C. Urmi. An overall decrease in the rate of post-predicate objects can be observed: the indefinite nominals and other pronouns are more likely to occur in pre-verbal position than, respectively, the definite nominals and personal and demonstrative pronouns. This also seems to hold true for C. Shaqlawa, but to a lesser extent still, i.e. only 13% of the indefinite objects are post-predicate.

2.2.2 Verb/goal

The endpoint of motion verbs and caused motion verbs, such as 'to come' and 'to bring' respectively, are subsumed under *Goal* (abbreviated G) here, while *Recipient* (R) refers to the human endpoint of a transfer like 'to give' and *Addressee* (Addr) to that of verbs of speech, e.g. 'to say,' 'to ask,' 'to talk'. *Beneficiaries* (Ben), i.e. indirect participants who are advantaged or disadvantaged by the action, have also been added here for completeness' sake. These argument classes are illustrated in (15–18) for C. Sanandaj. Table 5 displays the statistics resulting from the relevant datasets, which comprises all pronouns and full nominals. Here, the tokens from Younansardaroud (2001) for the dialect of Sardarid have also been added.

(15) a. Goal, motion verb
          C. Sanandaj (Panoussi 1990: 1:§4)
          say   ***arxe***
          IMP.go.SG.M mill.PL
          'Go **to the mill**!'
      b. Goal, caused motion verb
          C. Sanandaj (Panoussi 1990: 2:§14)
          tam-dāre-Ø-le   ***gaw ṣanoq-aw***
          PST.PFV-put-A.3SG.M-O.3SG.M in   chest-his
          'He placed him **in the wooden chest**.'

Table 5: Rate of post-predicate (PP) Goals, Recipients and Addressees (nominal and pronominal)

Doculect	G		R		Addr		Ben	
	n	**PP**	***n***	**PP**	***n***	**PP**	***n***	**PP**
J. Urmi	59	86%	19	11%	55	33%	19	32%
C. Urmi	129	92%	11	73%	37	24%	7	43%
C. Sardarid	—	—	11[a]	100%	5[b]	0%	—	—
C. Shaqlawa	44	91%	28	96%	31	97%	18	100%
C. Sanandaj	44	84%	4	75%	5	80%	11	55%
J. Sanandaj	207	91%	38	87%	32	72%	16	81%

[a]Younansardaroud 2001: 11:§5, 13:§1, 13:§2, 15:§3, 17:§1.

[b]Younansardaroud 2001: 9:§3, 15:§4, 16:§2, 2x 16:§3, 2x 16:§5, 2x 16:§6, 17:§8, 17:§10.

(16) Addressee
C. Sanandaj (Panoussi 1990: 4:§13)
mere **tlas-a** *gor-əd* *baxta*
said.PFV.PST.3SG.M to-3SG.F husband.SG.M-CSTR woman.SG.F
'The woman's husband said **to her**…'

(17) Recipient
C. Sanandaj (Panoussi 1990: 3:§16)
tm-ēw-ə́n-wa-lu **tlas-ox**
PST.PFV-give-A.1SG.M-PST-O.3PL to-2SG.M
'I had given them **to you**.'

(18) Beneficiary
C. Sanandaj (Panoussi 1990: 2:§2)
ayət **ta kalba** *hādax gī-wəd-lox*
you.SG.M for dog.SG.M such ANT-did-A.2SG.M
'You have done such a thing **for a dog**.'

These data are consistent with the findings in Noorlander & Molin (2022). The post-verbal position is preferred for Goals across all dialects, and the same holds true for Recipients and Addressees in Iranian Kurdistan, here represented by the Jewish and Christian dialects of Sanandaj. The handful of tokens in C. Sanandaj are relatively low, but suggest a typology similar to that of its Jewish counterpart,

except in the case of beneficiaries, which in general do not seem to betray a clear tendency. It is far more common for Addressees than for Recipients to be placed before the verb in the Christian NENA dialects of Urmi and Sardarid (see §3.2.3 for the areal significance of this Addressee/Recipient split), even though Recipients and Addressees are generally marked by the same preposition *qa-*, e.g.

(19) a. Verb-Recipient
          C. Sardarid (Younansardaroud 2001: 17:§10)
          *ʾaxnan xa ton čapač   jarāy Ø-yav-ax   **qa dar⁺bar***
          we   a ton sawdust must sвjv-give-1pl to court
          'We must give a ton of sawdust **to the court**.'
      b. Addressee-Verb
          C. Sardarid (Younansardaroud 2001: 15:§3)
          *⁺Šāh ⁺ʾAbbās **qa vazir** mār=ələ*
          Shah Abbas   to vizier GRD.say=COP.3SG.M
          'Shah Abbas says **to the vizier**…'

In comparison to objects, the placement of the aforementioned endpoint roles turns out to be more flexible overall. In the rare occasion that a ditransitive clause contains two full nominal objects, each argument class typically occurs at either side of the verb: the Theme, like the O, before the verb, but the Recipient, like Goals, after it, and thus OVR as illustrated in (19a). The same order, i.e. OVR, is also common in most Kurdish varieties (Haig 2022). This constituent order is cross-linguistically rare, as most languages reflect a preference to place both arguments at either side (Haspelmath 2015).

In the Jewish Urmi doculect (Khan 2008a), the pre-verbal position of O and R as well as Addressees is apparently the norm. The relative position of the Theme (O) and the Recipient (R) in a ditransitive clause is not entirely fixed, e.g.

(20) a. Recipient-Theme-Verb
          J. Urmi (Khan 2008a: §122)
          ba-⁺yal-i   ⁺ruzi   fərya   höl-Ø
          to-children-my provision abundant IMP.give-SG
      b. Theme-Recipient-Verb
          J. Urmi (Khan 2008a: §113)
          ⁺ruzi   fərya   ba-⁺yal-i   höl-Ø
          provision abundant to-children-my IMP.give-SG
          'Give abundant provision to my children.'

The most common order for the J. Urmi dialect described in Khan (2008a), however, is ROV, especially for pronominal Recipients. A contrastive or topical O may precede the R, and immediately pre-verbal placement may add narrow focus to the Recipient (see also Noorlander & Molin 2022: 244–246).

This notwithstanding, the dominant order in the majority of Trans-Zab Jewish dialects is OVR. Statistics based on others doculects of Jewish NENA in Iran approximate more closely the typology of that of Jewish varieties in Iranian Kurdistan. Texts in Garbell (1965b) and Hopkins (1989) contain far more cases of post-verbal Recipients and Addresseees than our J. Urmi doculect here (Khan 2008a),[7] as illustrated in (21) below, which suggests VR and VAddr are the more frequent position among Trans-Zab Jewish NENA dialects as a whole.

(21) a. Theme-Verb-Recipient
          J. Urmi (Garbell 1965b: 149.18)
          *əsrá dehwé   hwəl-le   **ba d-ö**   **mar**
          ten   gold.PL gave.PFV-A.3SG.M to GEN-DEM.SG owner.CSTR
          xmará
          donkey.SG.M
          'He gave ten pieces of gold **to the donkey owner**.'
      b. Verb-Addressee
          J. Urmi (Garbell 1965b: 149.20)
          mər-a   **ba d-ö**   **görá**
          said.PFV-A.3SFG to GEN-DEM.SG man.SG.M
          'He said **to that man**…'

2.2.3 Become/complement

In contradistinction to direct objects but similarly to Goals, the final state of change-of-state verbs, such as 'to become,' 'to turn into,' typically follows the predicate (e.g. Khan 2008a: 323), as shown in (22). Under this class one may also subsume the complements of 'to name' and 'to fill,' although, here, the object complement does not represent the final outcome of the primary object, but rather specifies the content of the verb. Neverthless, the pre-verbal position seems to be more frequent due to language contact (see §3.1.1). The complement can also be treated as a Recipient in J. Urmi and J. Sanandaj and flagged as such (see Noorlander & Molin 2022: 251–252).

---

[7]Khan's (2008a) texts are based on one male speaker, as the number of speakers available was much smaller than at the time of Garbell (1965b), who was able to consult more speakers.

(22)   a. Become-Complement
            J. Solduz (Garbell 1965b: 209)
            pra   *xdər-e*   ***dehwé***
            earth.SG.M became.PFV-S.3SG.M gold.PL
            'The earth turned **into pieces of gold**.'

        b. Object-Verb-Complement
            J. Solduz (Garbell 1965b: 231)
            tunnú xurjine   *⁺məly-i-la*   ***dehwé***
            both   saddle.bag.PL filled.PFV-O.3PL-A.3SG.F gold.PL
            'She filled both saddle bags **with pieces of gold**.'

With the verbs 'to say' and 'to make,' the double object construction shifts the semantics to that of 'to name X Y' and 'to make X into Y'. Thus, with the ambivalent verb (C.) *(h)wd* or (J.) *(h)wl* 'to make,' the post-verbal placement of the object correlates with its two-argument valence and the semantics of the resulting condition, i.e. 'to turn into,' rather than the effect of a single argument verb, i.e. 'to make' (Noorlander & Molin 2022: 252–253); contrast *kăbáb* (23a) with (23b) below.

(23)   a. Object-Verb
            J. Sanandaj (Khan 2009: B:§35)
            kăbáb   *kol-i-wa*
            kebab.SG.M make-A.3PL-PST
            'They made **kebab**.'

        b. Verb-Complement
            J. Sanandaj (Khan 2009: B:§35)
            kol-í-wa-le   ***kăbáb***
            make-A.3PL-PST-O.3SG.M kebab.SG.M
            'They made it into **kebab**.'

2.2.4 Other obliques

Here, obliques are confined to constituents related to *Place*, such as the locative complement of position verbs like 'to sit,' e.g. (24), and the *Source* of motion, e.g. (25). They are more likely post-predicate than objects, as given in Table 6 for the same datasets, which comprises both lexical and pronominal arguments. In Table 6, we observe that the rate of post-predicate locatives is higher in the Christian varieties overall and in Christian Urmi especially, whereas the Jewish varieties show a stronger verb-final preference. Thus, even in C. Sanandaj, the

rate of post-predicate Oblique is high, while the dialect otherwise patterns almost exactly like its Jewish counterpart. The relatively high rates of post-predicate Obliques suggests a general tendency for local case relations (Source, Place, Goal) to occur after the predicate, at least more frequently than objects. This is matched by similar findings from Balochi, see Nourzaei & Haig (2024 [this volume]).

(24) Place/Locative
      C. Sanandaj (Panoussi 1990: 4:§11)
      *ay   xōr-e   **gāw mezgəd** ītīw-a   =le*
      DEM.SG friend.SG.M-his in   mosque seated.PTCP-SG.M =COP.3SG.M
      'That friend of his is sitting **in the mosque**.'

(25) Source/Ablative
      C. Sanandaj (Panoussi 1990: 3:§11)
      *kod   yōma gaz   m-šāqə́l-Ø-wa-le   **mən šūqa***
      each day   honey.SG.M FUT-take-A.3SG.M-PST-O.3SG.M from market
      'Every day he bought Turkish honey **from the market**.'

Table 6: Rate of post-predicate (PP) place and source constituents (both nominal and pronominal)

Doculect	Place		Source	
	n	PP	*n*	PP
C. Urmi	64	81%	39	44%
C. Sanandaj	11	64%	11	100%
C. Shaqlawa	12	58%	7	57%
J. Urmi	40	40%	19	21%
J. Sanandaj	64	39%	28	14%

2.3 Other predicate types

2.3.1 Copulas

Post-predicate copula placement correlates with verb-final syntax (e.g. Dryer 2007: 91; see §2.4.2, for the use of the copula in verbal clauses), and constitutes a typological trait of the languages in the area (e.g. Matras 2009: 270; Haig 2017: 404–405). The syntax of the copula in main clauses in the Trans-Zab Jewish dialects differs from that in the Christian dialects in the same region (Khan 2012b).

Post-predicate placement is almost categorical in these Jewish dialects, with the only exception being certain modal contexts, consistent with an overall higher rate of OV in these Jewish dialects. The post-predicate position is favoured but less fixed in the Christian dialects.

NENA dialects generally distinguish between two copula bases:

(a) pronominal copulas, e.g. C. Urmi *ʾina* 'they are,' J. Urmi *ʾilu* 'they are';

(b) verbal copulas, i.e. *(h)wy* or *(h)vy* 'to be,' e.g. C. Urmi *ʾavi* 'they may be,' J. Urmi *haweni* 'they are'.

Other particles can be added to either base to express negation, past tense, subordination, and deixis, depending on the dialect. Importantly, the latter is excluded from this discussion, since the deictic (or presentative) copula has a fixed pre-predicate position throughout (e.g. Molin 2021: 227–247).

When the copula is placed after a constituent, there is a strong tendency for the copula to undergo cliticization and attachment to the immediately preceding element, for which reason I shall distinguish between bound, i.e. enclitic, and unbound copulas.

Post-predicate, thus often bound, copulas are used in present and/or past tense affirmative clauses, e.g.

(26) a. Predicate-Copula, present (bound)
J. Koy Sanjaq (Mutzafi 2004: 1A:§1)
*ʾoni=š   beʿeṛǝx   tremma   nafar-e   =**lu***
they=ADD approximately two.hundred person-PL =COP.3PL
'They **are** about two hundred people.'

b. Predicate-Copula, past
J. Koy Sanjaq (Mutzafi 2004: 1A:§1)
*kullú   ʾoni xet   mšǝlmān-é **we-lū***
all.of.them those others Muslim-PL COP.PST-3PL
'All those others **were** Muslims.'

(27) a. Predicate-Copula, present (bound)
C. Koy Sanjaq (Askar 2021: 215.§10)
*šm-ew   šúmʿŭn =**ile***   ba ăṣǝl
name.SG.M-his Simon   =COP.3SG.M in origin
'His name **is** actually Simon.'

b. Predicate-Copula, past
C. Koy Sanjaq (Askar 2021: 220.§25)
*ana dkan-əd osta   akram **yən-wa***
I   shop-of artisan Akram cop.1sg.m-pst
'I **was** in the artisan shop Akram.'

In deontic contexts, especially in idiomatic wishes, Copula-Predicate order is used in all dialects, e.g.

(28) Copula-Predicate
J. Shino (Garbell 1965b: 231.7)
⁺*šultaná* **Ø-hawe-Ø**   *basim-a*
king.sg.m sbjv-be-3sg.m healthy-sg.m
'May the king **be** blessed!'

(29) Copula-Predicate
C. Urmi (Khan 2016: A2:§4)
malka   **Ø-ʾav-ət**   *basim-a*
king.sg.m sbjv-be-2sg.m healthy-sg.m
'May the king **be** well!'

All copula forms, whether bound or unbound, follow the predicate in all Jewish dialects of Iran and most Jewish dialects of northeastern Iraq, except in these deontic contexts.

The copula placement is more free in the Christian varieties, and the Jewish dialects in the Erbil region. Moreover, the Trans-Zab Jewish dialects in Iraq differ in negative copula placement. It can either follow or precede the predicate in J. Arbel (Khan 1999: 320), while the negative copula follows the nominal predicate in J. Koy Sanjaq (Mutzafi 2004: 107) and J. Sulemaniyya (Khan 2004: 254), as it does in the Jewish dialects of Iran. Copula syntax is summarized in Table 7 by contrasting Christian and Jewish Urmi.

The unbound copula freely occurs before the predicate in the dialects of Iranian Azerbaijan, and similarly also other dialects in Iraq such as C. Diyana (Napiorkowska 2015), as illustrated in (30a), instead of being cliticized to the predicate, e.g. (30b). The cliticization of the copula to the *subject* constituent, such as *garda* 'net' in ⁺*ham gárdə=la* ⁺*allo* 'also a net is on her' (Khan 2016: II: 289), is rare. The complement occurs after the copula in 16/96 (17%) cases in the C. Urmi doculect and 11/86 (13%) cases in the C. Shaqlawa doculect, which, as expected, occurs more frequently than in the J. Sanandaj doculect, which only has 5/215 (2%) cases. Copula-Predicate order with a lexical subject, as shown in (30a), can

Table 7: Copula placement in Jewish and Christian Urmi

	C. Urmi	J. Urmi
'My son is hungry.'	*bruni kpä́nə=**lə***	*brönä́ kpiná=**ile***
'He is hungry.'	*kpä́nə=**lə***	*kpiná=**ile***
'I am a king.'	*ʾana xa malk=**ən***	*ʾaná xa ⁺šültane=**len***
'Who is your friend?'	*xorux máni=**lə***	*⁺barüxŏx mǎni=**le***
'My son is not hungry.'	*bruni **lelə** kpina*	*brönä́ kpiná **lewe***
'He is/IS hungry.'	*ʾ**ilə** kpina*	—
'I am/AM a king.'	*ʾana ʾ**ivən** xa malka*	—

be analysed as a cleft sentence, i.e. 'I am (the one who is) the vizir of your father,' which is used in contexts of identification and specification and the expression of properties that are permanent or contra-presuppositional (Khan 2016: II: 158–162).

(30) a. Copula-Predicate
C. Urmi (Khan 2016: A2:§25)
*ʾana ʾ**īn-va** vazzir-ət bab-ət diyy-ux*
I cop.1sg-pst vizier-cstr father-cstr gen-2sg.m
'I **was** the vizier of your father.'

b. Predicate-Copula
C. Urmi (Khan 2016: A2:§25)
*ʾana ʾatxa naš=**ən-va***
I such man.sg.m=cop.1sg-pst
'I **was** such a **man**.'

As compared in (31) and (32), while the affirmative copula tends to be post-predicate, the negative copula is generally pre-predicate in Christian dialects, contrast *=la* 'she is' and *lewa* 'she is not' in (31) and *=ən* 'I am' and *lēn* 'I am not' in (32).

(31) a. Copula-Predicate
C. Shaqlawa (Khan et al. 2022: Text 23:§28)
*har máre-wən bráte =**la***
every.time say.inf-1sg.m girl.sg.f =cop.3sg
'I keep saying **it is** a girl.'

b. Predicate-Copula
    C. Shaqlawa (Khan et al. 2022: Text 23:§28)
    *har   máre-wat   **lewa***   *brata*
    every.time say.INF-2SG.F NEG.COP.3SG.F girl.SG.F
    'You keep saying **it is not** a girl.'

(32) a. Copula-Predicate
          C. Urmi (Khan 2016: A43:§15)
          *ʾána=da **lēn***   *tliqa yala*
          I=ADD   NEG.COP.1SG lost child
          'I **am not** a lost child.'

      b. Predicate-Copula
          C. Urmi (Khan 2016: A43:§15)
          *ána=da brūn-malk =**ən***
          I=ADD   son-king   =COP.1SG
          'I **am** the son of a king.'

Finally, the Christian dialects may have a special copula that fuses the relativizer *d* with the copula, e.g. *d* 'that/which/who' + *ʾile* 'is' > *ṱ-ile* 'that/which/who is,' e.g.

(33) Copula-Predicate
      C. Shaqlawa (Khan et al. 2022: Text 4:§42)
      *ṣăṱáne   =le   **ṱ-ile***   *xor-a*
      devil.SG.M =COP.3SG REL-COP.3SG.M friend.SG.M-her
      'It is the devil **who is** her friend.'

2.3.2 Auxiliaries

The combination of Verb-Auxiliary and Object-Verb ordering is considered to be a form of harmonic word order (e.g. Dryer 1992: 100), since both are considered head-final in the standard assumption that the auxiliary constitutes the head of auxiliary in content verb pairings, since it bears verbal inflectional properties pertinent to the clause. The copula can serve as an auxiliary in NENA dialects, expressing several TAM properties as well as negation together with the nonfinite verb. The syntax of the copula as an auxiliary in a verbal predicate can sometimes differ from its syntax in a non-verbal predicate.

First of all, as expected, the Object-Verb-Auxiliary pattern of the copula is what we find in the Jewish Trans-Zab dialects, as illustrated in (34a). Future and modal

auxiliaries, such as ⁺*msy* 'be able' in (34b), show Auxiliary-Object-Verb order in–
at least historically–same subject modal complements.

(34) a. Object-Verb-Auxiliary
           J. Urmi (Garbell 1965b: 194)
           balki   xa danká mən-nu   əl-d-o   araqčə́n
           maybe a   CLF   from-3PL DOM-GEN-DEM.SG cap.SG.M
           xəzy-á   Ø-hawe-la
           seen.PTCP-O.SG.M SBJV-be-A.3SG.F
           'Maybe one of them **would have seen** that cap.'

       b. Auxiliary-Object-Verb
           J. Urmi (Garbell 1965b: 194)
           ⁺məssé   ⁺maé   m-mešá   ⁺palə́t-Ø
           can.3SG.M water.PL in-forest.SG.F take.out.SBJV-A.3SG.M
           'He **would be able to find** water in the forest.'

In northeastern Iraq, however, the negative copula, which functions as a neg-
ative auxiliary with certain verbal forms, always precedes the content verb in
J. Koy Sanjaq, such as *lewan* in (35), and contrasts with its post-predicate place-
ment as a copula (Mutzafi 2004: 108) in §2.3.1. Other Trans-Zab Jewish dialects
do not make use of this, but place a negator (*la*) before the verb phrase, such as
Neg-V-Aux in (36).

(35) Object-Negative Auxiliary-Verb
       J. Koy Sanjaq (Mutzafi 2004: 2A:§4)
       *ʾixalá   **le-wan**   xəl-tá*
       food.SG.M NEG-COP.A.1SG.F eaten.PTCP-A.SG.F
       '**I have not** eaten any food.'

(36) Object-Negator-Verb-Auxiliary
       J. Sulemaniyya (hypothetical example based on Khan 2004)
       *xalá   **la-xəlte**   =yan*
       food.SG.M NEG-eaten.PTCP-A.SG.F =COP.A.1SG.F
       '**I have not** eaten any food.'

The Christian NENA varieties, in turn, favour Auxiliary-Verb order through-
out, as illustrated in (37), without affecting the Object-Verb order. C. Urmi could
be characterised as 'VO' in light of the pre-verbal auxiliary placement in contrast

to the Verb-Auxiliary order in others dialects where OV predominates through-out (Khan 2020: 399). There is no a priori reason, however, to consider the position of the auxiliary more 'basic' than that of the object in the manifestation of Aux-OV order, cf. also Dutch and German, which also exhibit Aux-OV order in main clauses and are generally considered V-final in formal approaches to word order. See also Stilo (2015: 345) on Aux-OV in same subject 'want' complements in Colloquial Armenian and Azeri; Aux-OV is also the norm for Kurdish (see Mohammadirad 2024 [this volume]).

(37)   a. Auxiliary-Object-Verb
        C. Urmi (Khan 2016: B17L§4)
        *ʾaxnan **k-av-ax-va**   makkə   praxa*
        we   IND-be-1PL-PST maize.PL hull.INF
        'while we **were hulling** the maize'
   b. Auxiliary-Object-Verb
        C. Urmi (Khan 2016: A16:§3)
        lá-⁺ʾams-ən   *ʾid-i*   *yavv-ən-na*
        NEG-can-S.1SG.M hand.SG.F-my give-A.1SG.M-O.3SG.F
        'I **cannot give** my hand.'

2.3.3 Complement of non-finite verbs

Finally, the complement of non-finite verb follows the syntax of the complement of finite verbs. Thus, phasal verbs like 'to begin' can combine with a nominal or non-finite form of the verb, such as an infinitive or gerund, respectively. If the nominal/non-finite form of the verb had followed the common pattern of Noun-Attribute, the order would have been Verb-Complement, as would be expected for the majority of NENA dialects. However, in the Trans-Zab Jewish dialects, if the complement corresponds with an object, it also betrays the same syntax, and is thus placed before the non-finite verb, as shown in (38a). The placement of the prefixal preposition *b-* before the complement *ʔərbe* 'sheep' of the infinitive *zwana* 'to sell' suggests a type of compounding or noun incorporation strategy, i.e. 'He began sheep-selling'. Interestingly, the same order occurs in C. Urmi, as shown in (38b), although here the preposition *b-* always attaches to the verbal form. How common this is among the other NENA dialects in the region requires further investigation.

(38)   a.   Object-Verb(Non-finite)
            J. Arbel (Khan 1999: S:§72)
            bde-le   *bə-ʔərbe*   *zwana*
            begin.PFV-A.3MSG at-sheep.PL selling.INF
            'He began selling sheep.'

       b.   Object-Verb(Non-finite)
            C. Urmi
            ⁺*šuri-lun*   *ʔərbə*   *bə-zvana*
            begin.PFV-A.3PL sheep.PL at-selling.INF
            'They began selling sheep.'

3 Areal issues

In the NENA-speaking area, predicate-final order becomes more prevalent in the east, where, today, varieties of Central Kurdish are dominant.[8] The convergence towards OV in the Christian NENA varieties of Iranian Azerbaijan, however, is incomplete, as it still maintains a greater degree of flexibility for object placement and betrays features of predicate-initial typology (§3.2) in its ordering of indefinite objects and possessums (§3.2.1), and that of light verb complements (§3.2.2, see also §2.3.2 on auxiliaries). Thus, although Christian Urmi is by sheer frequency characterizable as predicate-final, it is, on closer examination, typologically more mixed than the Tran-Zab Jewish group.

   The areal convergence in word order can be merely incidental, such as in the case of Noun-Genitive, Demonstrative-Noun-Adjective, and the prevalence of Verb-Goal and Become-Complement, which are common to Semitic and Iranian. The higher rates of OV in the main dialects discussed in this chapter, however, are doubtless due to influence from neighbouring OV languages, and yet the outcome differs per region and community. For example, the word order profile of the Jewish and Christian dialects of Sanandaj as representative of the dialects of Iranian Kurdistan matches in many ways that of the local Iranian varieties, i.e. Central Kurdish and Gorani (see Mohammadirad 2024 [this volume]). The fact that they show a similar general tendency towards Object-Verb and Predicate-Copula order is most likely due to contact with Iranian. However, many configurations are still liable to language-internal developments, such as the syntax of objects, which is sensitive to definiteness in dialects like C. Urmi. Moreover, the same syntactic structure is attested where contact with neighbouring OV languages presumably played no direct role, such as Aux-OV order

---

[8]This corresponds to Stilo's (2005) Buffer Zone 1, bordering a low OV and high OV zone.

and the pre-verbal placement of the object before non-finite verbal forms (see 2.3.3). The syntax of copulas and auxiliaries (see 2.3.2 and 2.3.2) also exhibits language-internal peculiarities. Convergence in word order can thus be partly understood as contact-induced reinforcement of pre-existing parallel patterns (e.g. Silva-Corvalán 1994, 2008[9]; Heine 2008: 42–43, 48–49), which presupposes that syntactic OV presumably goes back to an original situation of more fluid order driven by pragmatic configurations.

Areal features are present in all Trans-Zab Jewish varieties, but the effect of contact on Christian varieties seems to be more varied. Doubtless, speakers' attitudes play a crucial role in susceptibility to contact-induced change in NENA dialects (Noorlander 2014) and thus word order shifts. One may think that Christian NENA varieties show considerably higher rates of VO than their Jewish peers. This seems to hold true for the Christian dialects of Iranian Azerbaijan, but it does not seem to apply to the southeasterly located dialects such as Christian Shaqlawa, Koy Sanjaq, Sulemaniyya and Sanandaj, where OV seems to be as rigid as it can be in Trans-Zab Jewish dialects.

Regional effects, in turn, may also cut across communal differences. Both Jewish and Christian NENA dialects of Iranian Azerbaijan, for instance, show a higher rate of Addressee-Verb order for verbs of speech, which parallels the northernmost dialects of Kurdish (Haig 2022), and Adjective-Noun order for evaluative adjectives, which parallels Azeri. On the other hand, other features are completely insensitive to word order shifts: flagging, for instance, remains prepositional in all NENA dialects, and none of the Neo-Aramaic varieties discussed here have developed a system of postpositions, despite having undergone a shift from VO to OV.

Further research is required on the embedding of the shift from VO to OV in other internally and externally caused developments, as well as on the social-historical circumstances. Moreover, it is possible that the shift documented for NENA, especially in the east, is ultimately rooted in the spoken varieties of Mesopotamia during the Achaemenid period.[10]

3.1 Iraqi and Iranian Kurdistan

3.1.1 Complement/become

Similarly to Kurdish, a general post-predicate proclivity becomes apparent in object complements (see §2.2.3) of verbs of naming and turning something into

---

[9]I am indebted to G. Haig for directing my attention to this reference.

[10]See Haig et al. (In press) for a discussion of this from a wider areal and typological perspective.

something else in the Jewish varieties of northeastern Iraq and Iranian Kurdistan. The Iranian preposition *ba-* (see Mohammadirad 2024 [this volume]) has been transferred into these Jewish varieties along with the pattern, e.g.

(39) a. Verb-Complement
       J. Sanandaj (Khan 2009: D:§1)

       xir-Ø   ***ba-xá***   *broná*
       became.PFV-S.3SG.M   to-INDEF   boy.SG.M

       'He became a boy.'

   b. Verb-Complement
       J. Sanandaj (Khan 2009: B:§41)

       kol-i-wa-le   ***ba-lešá***
       make-A.3PL-PST-O.3MS   to-dough.MS

       'They made it into dough.'

This tendency for post-predicate placement is, however, not supported by the statistics, as the pre-verbal position, for example in (40) below, occurs far more often than in the NENA dialects outside of the relevant area. The rate of post-predicate final states is 22% (4/18) in C. Shaqlawa, 27% (3/11) in J. Urmi, 30% (12/41) in J. Sanandaj, and 62% (8/13) in C. Urmi, which can be contrasted with 90% (18/20) in C. Barwar.

(40) a. Complement-Verb
       C. Shaqlawa (Khan et al. 2022: Text 35:§33)

       ʾen-aw   ***kor***   *pəš-lu*
       eye.PL-his   blind   became.PFV-S.3SG.M

       'His eyes became **blind**.'

   b. Complement-Verb
       C. Shaqlawa (Khan et al. 2022: Text 28:§19)

       ʾāt   *bet-i*   ***nura***   *qam-ʾawd-ət-e*
       you.SG   house.SG.M-my   fire.SG.F   PST.PFV-make-A.2SG.M-O.3SG.M

       'You have turned my home **into a hell**.'

3.1.2 Light-verb complements

Light verb constructions are also consistent with their usage in neighbouring languages, the non-referential nominal element preceding the light verb. The object NP, as expected, occurs before the complex verb, e.g. (41). In Christian dialects of northeastern Iraq, recent Arabic loans are also incorporated into this strategy (Hakeem 2021: 474–475), e.g. (42).

(41) Noun-Light Verb
C. Sanandaj (Panoussi 1990: 126:§16)
*xay-u   ləbas-i   **ḥazər k-od-i-lu***
one-of.them clothes-my ready IND-make-A.3PL-O.3PL
'One of them **prepares** my clothes.'

(42) Noun-Light Verb
C. Shaqlawa (Khan et al. 2022: Text 12:§28)
dăbi ʾana ʾistəraḥat Ø-ʾawd-ən
must I   rest   SBJV-do-1SG.M
'I must **rest**.'

3.1.3 Attribute/noun

The NENA dialects in this area have, in general, not changed the placement of attributes, contrary to the dialects in Iranian Azerbaijan (§3.2.5). Attribute-Noun order is infrequent in the majority of NENA dialects, drawing attention to the listener with additional emotional weight from the speaker's perspective. This position thus rarely occurs with more objective statements, except when two or more attributes are contrasted.[11] In several cases, however, the lexical item or morpheme is transferred along with the pattern. In (43a) below, for instance, the Persian loan-adjective *behtarīn* 'best' precedes the noun, and thus follows the expected Iranian order for superlatives. Interestingly, in (43b) below, the originally Kurdish adjective *zirej* precedes the head, presumably for pragmatic reasons, even though Noun-Adjective order is also the convention in Kurdish.

(43) a. Adjective-Noun
C. Sanandaj (Panoussi 1990: 122:§9)
behtarín ixale
best   food.SG.M
'the best kinds of food'

b. Adjective-Noun
C. Diyana (Napiorkowska 2015: 301)
rɒba zirej   naš-e
much clever.INVAR person-PL
'very clever people'

---

[11]See Gutman (2018: 143, 224, 232–233, 246–247, 250, 255).

The adjective *xoš* is another case in point: it is a Persian *Wanderwort* in the NENA speaking area – also found in Turkish and Iraqi Arabic – and seems to be equally compatible with Adjective-Noun order in virtually all NENA dialects, in particular in combination with the noun *naša* 'man':

(44)   Adjective-Noun
        C. Shaqlawa (Khan et al. 2022: Text 17:§12)
        xoš   naša
        good man.SG.M
        'a good man'

(45)   Adjective-Noun
        J. Arbel (Khan 1999: 462.§326)
        xoš   naše=le
        good man=COP.3SG.M
        'He is a good man.'

Furthermore, ordinal numbers in the Jewish NENA dialects of northeastern Iraq and Iranian Kurdistan all add the Iranian suffix *-(a)min* to the native Aramaic numeral, e.g. *tmanyamán* 'eighth' from *tmanya* 'eight' + *min*. Ordinals follow their heads as attributes as they do in the majority of NENA dialects, and this order incidentally convergences with Kurdish. In J. Sanandaj, however, the ordinal can also precede the noun as it does in Gorani (MacKenzie 1966: 24), e.g.

(46)   Ordinal-Noun
        J. Sanandaj (Khan 2009: 213)
        arba-mán gorá
        four-ORD man.SG.F
        'fourth man'

3.2 Iranian Azerbaijan

3.2.1 Object/verb and possessum/existential

There is a direct correlation between object and possesum placement in NENA dialects. In locative-exisential possessor constructions, exemplified in (47–48), the possessum can be considered the object-like argument, and hence, not surprisingly, it occupies the same position as object NPs in verbal clauses in the majority of NENA dialects. Figure 2 shows the post-predicate possessum placement and the corresponding figures for definite objects. The VO dialects of NENA,
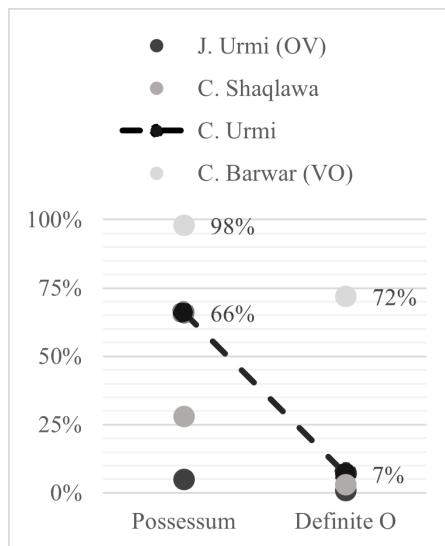
Figure 2: Rate of post-predicate (PP) possessums and definite objects

represented here by Barwar (northwestern Iraq) in the top, and the OV dialects of NENA, represented by J. Urmi in the bottom, are largely consistent throughout, although definiteness is a factor of preverbal object placement in C. Barwar. The possessum (25 out of 38), however, occurs post-verbally far more often than definite objects in C. Urmi, and is comparable to *in*definite objects (24 out of 49, see §2.2.1). It is conceivable that discourse organisation also plays a role in possessum-possessor constructions. Since possessums are generally indefinite, we would expect to observe the same distribution. The post-predicate possessums occur more frequently still, although not in a statistically significant way. This higher rate of post-predicate possessums might be due to formulaic language, e.g. opening formulas introducing discourse-new arguments, as illustrated in (47), which would favour post-predicate placement.

(47)   Existential-Possessum
       C. Urmi (Khan 2016: A39:§1)
       ʾət-va   ***xa-malka****.*   *ʾaha*   *malka*   *ʾə́t-va-le*   ***⁺ṭla***
       EXIST-PST one-king.SG.M DEM.SG.M king.SG.M EXIST-PST-3SG.M three
       bnunə
       son.PL
       'There was **a king**. This king had **three sons**.'

(48)   Possessum-Existential
       C. Sanandaj (Panoussi 1990: 3:§1–2)
       xa gora   *k-awe... ay*   *gora*   *məšulman-a **tre išənyase***
       one man.SG.M IND-be   DEM.SG man.SG.M muslim-SG.M two woman.PL
       k-āwe-le
       IND-be-POSS.3SG.M
       'There was (lit. is) **a man**... This Muslim man had (lit. has) **two wives**.'

3.2.2 Light-verb complements

Moreover, similarly to indefinite objects and to possessums, the nominal element
of light verb constructions from Kurdish, Persian and Azeri can occur after the
light verb in C. Urmi, e.g. *kullə ʾodilun ⁺**hazər*** 'they should prepare everything'
(Khan 2016: A3:§70). By contrast, such non-referential nominal elements precede
the light verb in J. Urmi as do all types of objects and possessums (Garbell 1965a:
173), e.g. *ixalé ⁺**hazər** wudun* 'prepare food!' (J. Urmi, Garbell 1965b: 156).

3.2.3 Addressee/verb & verb/goal

Addressees, in turn, seem to have shifted in the same predicate-final direction as
objects in Iranian Azerbaijan (§2.2.2), which is more prevalent in the northeast.
Here, the syntactic organisation is role-specific and area-specific. The pre-verbal
Addresseee but post-verbal Goal and Recipient split in C. Urmi and Sardarid Azer-
baijan matches the findings for some varieties of Central Kurdish and northern-
most dialects of Northern Kurdish (Haig 2022: 358–359; Mohammadirad 2024
[this volume]). Influence from especially Azeri and Persian but also Armenian,
however, cannot be ruled out either. A higher rate of Verb-Goal order, for exam-
ple, also occurs in local Azeri dialects due to Colloquial Persian influence (Kıral
2001: 75–77).

3.2.4 Complement/become

The Complement, i.e. a resulting condition, is generally placed before the verb
in the NENA dialects of Iranian Azerbaijan studied here, which may even be
fronted before the object, as can be observed in (49). This pre-predicate placement
matches the word order in Turkic and Persian, rather than what is expected in
Kurdish. The same holds true for Recipient-Theme-Verb order in the Jewish Urmi
material in Khan (2008b), as exemplified by (20) in §2.2.2.

(49) Complement-Object-Verb
   C. Sardarid (Younansardaroud 2001: 15:§1)
   trə ʾaxunvatə ⁺tla d-ani vəd-li
   two brother.PL three GEN-DEM.PL made.PFV-1SG
   'Into two brothers I turned three of them.'

3.2.5 Attribute/noun

Adjective-Noun and Genitive-Noun, alongside Object-Verb, are the dominant orders in Turkic languages that NENA dialects in Iranian Azerbaijan have been in contact with,[12] which could be considered parallel to the shift from VO to OV (Gutman 2018: 222, 233, 250–251). The complement of nominal forms of the verb, such as agent nominalisations or participles ending in *-ana*, precedes the verb in compound-like attributes across Trans-Zab Jewish NENA dialects, similarly to their OV typology, contrast (50a) with (50b) and (50c). In C. Urmi, the attribute generally follows the head, as given in (51a), but sporadically the opposite order does occur, as illustrated in (51b). Also, when the adjective constitutes the head of a compound, J. Urmi patterns in a head-final fashion but C. Urmi head-initial, cf. J. Urmi *dəqná-xwará* lit. 'beard-white' and C. Urmi lit. *xvār-dəqna* 'white-beard' for 'grey-bearded'.

(50) a. Noun-Genitive
         J. Urmi (Garbell 1965a)
         gör-ət **xalünt-xun**
         husband.SG.M-CSTR sister.SG.F-your.2PL
         'the husband of **your sister**'
      b. Genitive-Noun
         J. Urmi (Garbell 1965b: 212)
         xalünt-xun *gör-an-a*
         sister.SG.F-your.2PL marry-AGN-SG.M
         '**your sister's** husband'
      c. Genitive-Noun (compound-like)
         J. Urmi (Garbell 1965b: 86)
         masy-e *döq-an-a*
         fish-PL catch-AGN-SG.M
         'fisherman, lit. **fish** catcher'

---

[12]See also Garbell (1965b: 171–172), Khan (2016: 39), Gutman (2018: 220–224, 332–334).

(51)   a. Noun-Genitive
        C. Urmi (Khan 2016: III:44)
        doq-an-əd   **nuyn-e**
        catch-AGN-CSTR fish-PL
        'fisherman, lit. catcher of **fish**'

     b. Genitive-Noun
        C. Urmi (Khan 2016: A48:§32)
        ʾo   **prəzla**   **⁺taptəpp-án-a** *damurči*
        DEM.SG iron.SG.M bash-AGN-SG.M blacksmith.SG.M
        'the **iron hammerer** blacksmith'

Similarly, ordinal numbers formed out of the fusion of the Iranian suffix *-amin* and Turkic suffix *-inji* (-IncI) also precede the noun in J. Urmi as they do in the source language of this morphological transfer, e.g. *xa tre-minji baxtá* 'a second woman' (Garbell 1965b: 206).

Finally, more frequent Adjective-Noun order due to Azeri occurs among NENA speakers in Iranian Azerbaijan. Contrast the following two near-identical descriptions in J. Sanandaj (W Iran) and J. Urmi (NW Iran):

(52)   Noun-Adjective
      J. Sanandaj (Khan 2009: B:§158)
      kništa   **rab-ta**=w *kništa*   **zor-ta**
      synagogue.SG.F big-SG.F=and synagogue.SG.F small-SG.F
      'a **big** synagogue and a **small** synagogue'

(53)   Adjective-Noun
      J. Urmi (Khan 2008b: §156)
      xa **rab-ta** *knəšta*   *xa* **zör-ta** *knəšta*
      a   big-SG.F synagogue.SG.F a   small-SG.F synagogue.SG.F
      'a **big** synagogue and a **small** synagogue'

Table 8 contrasts corpus-based frequencies of Adjective-Noun order for the adjectives 'small, young' and 'red' in Jewish and Christian Urmi. This is consistent with a dominant Noun-Adjective order in C. Urmi (Khan 2016: II:39). There is a greater degree of flexibility in J. Urmi,[13] however, and a dominant order cannot be identified for more evaluative adjectives like *zora* 'small, young' in this dialect.

---

[13]See Garbell (1965b: 82–84), Khan (2008a: 216–219); Gutman (2018).

Table 8: Rate of AdjN in NENA dialects of Iranian Azerbaijan

	J. Urmi (Garbell 1965b[a]; Khan 2008b)		C. Urmi (Khan 2016)	
	n	ADJ-N	***n***	ADJ-N
small, young	33	49%	81[b]	6%
red	20	0%	18	0

---

[a]including Jewish Shino.
[b]excluding the high frequency idiom *yala sura* 'baby.'

When two adjectives modify the noun, they can also be placed at either side, cf. (54c). When the adjective modifies a noun that is part of a nominal annexation construction, however, it tends to follow the noun phrase, as illustrated in (54d).

(54) a. Adjective-Noun
         J. Urmi (Garbell 1965b: 83)
         zür-ta   *xalünt-u*
         small-SG.F sister.SG.F-their
         'their **youngest** sister'
      b. Noun-Adjective
         J. Urmi (Garbell 1965b: 172)
         brata   ***zür-ta***
         girl.SG.F small-SG.F
         'the **youngest** daughter'
      c. Adjective-Noun-Adjective
         J. Urmi (Garbell 1965b: 192)
         *xa **zür-ta***   *tkana*   ***šüšaband***
         a   small-SG.F shop.SG.F glass.covered
         'a **small glass-covered** shop'
      d. Noun-Genitive-Adjective
         J. Urmi (Garbell 1965b: 86)
         brat-ət   *⁺šültana*   ***zür-ta***
         girl.SG.F king.SG.M small-SG.F
         'the king's **youngest** daughter'

Adjective/Noun configurations are said not to correlate with other configurations (e.g. Dryer 1992: 95–96), but the above finding for Jewish Urmi warrants

further research into this understudied area. For example, is the syntax of adjectives more likely to be affected by contact than other nominal attributes or not?

Furthermore, sporadically, the attribute can consist of a gerundial verb phrase preceding the head noun, reminiscent of the Relative-Noun order in Turkic (Garbell 1965a: 173), e.g. (55–56).

(55) Numeral-Adjective-Gerund-Noun
C. Urmi (Khan 2016: A56:§1)
xa šap̂ər-ta   max ⁺šrá   bəllaya   brata
a   beautiful-SG.F like lantern shine.GRND girl.SG.F
'a beautiful girl shining like a lantern'

(56) Gerund-Noun
J. Shino/Solduz (Garbell 1965b: 84)
ba-šatoe   xriw-e ⁺mae
for-drink.GRND bad-PL water.PL
'water bad for drinking'

3.2.6 Standard/adjective

Finally, in the Jewish dialects of Iranian Azerbaijan, all instances of the standard of comparison seem to occur consistently before the adjective rather than after it, compare (57) with (58). A stronger preference for the position before the predicate converges with the typology of local OV languages.

(57) Standard-Adjective
J. Solduz (Garbell 1965b: 211.19)
mənn-áw ⁺raba ⁺raba bis-sqil-é   ita   g-d-ay   ⁺ahrá
from-her   very very more-beautiful-PL EXIST in-GEN-DEM.SG.F city.SG.F
'There are far, far more beautiful women **than she** in this city.'

(58) Adjective-Standard
J. Koy Sanjaq (Mutzafi 2004: 1B:§24)
*ʾo bä́š-faqir Ø-hawé-Ø   **mənn-éw***
he more-poor SBJV-be-3SG.M from-3SG.M
'Even if he is poorer **than him**.'

Paul M. Noorlander

Abbreviations

1	1st person	IMP	Imperative
2	2nd person	J.	Jewish (linguistic variety)
3	3rd person	LINK	linker
A	agent	M	masculine
ADD	additive	n	total number of tokens
Addr	addresseee	NENA	Northeastern Neo-Aramaic
ADJ	adjective	NP	Noun phrase
AGN	agent nominalization	O	object
ANT	anterior	PL	plural
AUX	auxiliary	PP	post-predicate
BEN	beneficiary	PRED	predicate
C.	Christian (linguistic variety)	PST	past
COP	copula	PTCP	participle
CSTR	construct	R	recipient
DEM	demonstrative	REFL	Reflexive
DOM	Diffrential Object Marking	S	subject (intransitive)
EXIST	existential	SG	singular
F	feminine	St	standard of comparison
G	Goal	V	verb
GEN	genitive	WOWA	= Haig et al. (2022)
GRD	gerund		

Acknowledgements

References

Askar, Kawther N. 2021. *Der neuaramäische Dialekt von ʾAṛmoṭa*. Heidelberg: Heidelberg University. (Doctoral dissertation).

Coghill, Eleanor. 2018. Information structure in the Neo-Aramaic dialect of Telkepe. In Evangelia Adamou, Katharina Haude & Martine Vanhove (eds.), *Information structure in Lesser-described languages*, 293–327. Amsterdam: Benjamins.

Cohen, Eran. 2012. *The syntax of Neo-Aramaic: The Jewish dialect of Zakho*. Piscataway, NJ: Gorgias Press. DOI: 10.31826/9781463234737.

Dryer, Matthew S. 1992. The Greenbergian word order correlations. *Language* 68. 81–138.

Dryer, Matthew S. 2007. Word order. In Timothy Shopen (ed.), *Language typology and syntactic description*, vol. 1, 61–131. Cambridge: Cambridge University Press.

Duval, M. Rubens. 2009. *Les dialectes néo-araméens de Salamas*. Piscataway, NJ: Gorgias Press. DOI: 10.31826/9781463213961.

Garbell, Irene. 1965a. The impact of Kurdish and Turkish on the Jewish Neo-Aramaic dialect of Persian Azerbaijan and the adjoining regions. *Journal of the American Oriental Society* 85(2). 159–177. DOI: 10.2307/597988.

Garbell, Irene. 1965b. *The Jewish Neo-Aramaic dialect of Persian Azerbaijan*. De Gruyter Mouton. DOI: 10.1515/9783110877991.

Gutman, Ariel. 2018. *Attributive constructions in North-Eastern Neo-Aramaic*. Berlin: Language Science Press. DOI: 10.5281/zenodo.1182527.

Häberl, Charles G. 2011. Neo-Mandaic. In Stefan Weninger, Geoffrey Khan, Michael P. Streck & Janet C. E. Watson (eds.), *The Semitic languages*, 725–737. Berlin: De Gruyter. DOI: 10.1515/9783110251586.725.

Haig, Geoffrey. 2017. Western Asia: East Anatolia as a transition zone. In Raymond Hickey (ed.), *The Cambridge handbook of areal linguistics*, 396–423. Cambridge: Cambridge University Press.

Haig, Geoffrey. 2022. Post-predicate constituents in Kurdish. In Yaron Matras, Geoffrey Haig & Ergin Öpengin (eds.), *Structural and typological variation in the dialects of Kurdish*, 335–377. Cham: Springer International Publishing. DOI: 10.1007/978-3-030-78837-7_8.

Haig, Geoffrey & Geoffrey Khan. 2019. *The languages and linguistics of Western Asia: An areal perspective*. Berlin: De Gruyter Mouton.

Haig, Geoffrey, Paul M. Noorlander & Nils N. Schiborr. In press. Which word order features are stable in a contact setting? Corpus-based evidence from the Western Asian transition zone. In Jeroen Darquennes, Joe Salmons & Wim Vandenbussche (eds.), *Language contact: An international handbook*, vol. 2. Berlin: De Gruyter Mouton.

Haig, Geoffrey, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils Schiborr. 2024. Post-predicate elements in the Western Asian Transition Zone: Data, theory, and methods. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 3–54. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266331.

Haig, Geoffrey, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr. 2022. *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation*. Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

Hakeem, Salam N. H. 2021. Arabic loanwords in the Neo-Aramaic dialect of Ankawa. In Geoffrey Khan & Paul M. Noorlander (eds.), *Studies in Neo-Aramaic grammar and lexicon* (Cambridge Semitic Languages and Cultures 5), 469–486. Cambridge: Cambridge University Faculty of Asian, Middle Eastern Studies & Open Book Publishers.

Haspelmath, Martin. 2015. Ditransitive constructions. *Annual Review of Linguistics* 1(1). 19–41. DOI: 10.1146/annurev-linguist-030514-125204.

Heine, Bernd. 2008. Contact-induced word order change without word order change. In Peter Siemund & Noemi Kintana (eds.), *Language contact and contact languages*, 33–60. Amsterdam: Benjamins. DOI: 10.1075/hsm.7.04hei.

Hopkins, Simon A. 1989. A tale in the Jewish Neo-Aramaic dialect of Naγada. *Jerusalem Studies in Arabic and Islam* 12. 243–281.

Hopkins, Simon A. 1999. The Neo-Aramaic dialects of Iran. *Irano-Judaica* 4. 311–327.

Hopkins, Simon A. 2009. "That monster of a man" and the emotive genitive. In Gideon Goldenberg & Ariel Shisha-Halevy (eds.), *Egyptian, Semitic and general grammar: Studies in memory of H. J. Polotsky*, 363–389. Jerusalem: The Israel Academy of Sciences & Humanities.

Kapeliuk, Olga. 2004. Iranian and Turkic structural interference in Arabic and Aramaic dialects. *Jerusalem Studies in Arabic and Islam* 29. 176–194.

Khan, Geoffrey. 1999. *A grammar of Neo-Aramaic: The dialect of the Jews of Arbel*. Leiden: Brill.

Khan, Geoffrey. 2004. *The Jewish Neo-Aramaic dialect of Sulemaniyya and Ḥalabja*. Leiden: Brill.

Khan, Geoffrey. 2008a. *The Jewish Neo-Aramaic dialect of Urmi*. Piscataway, NJ: Gorgias Press.

Khan, Geoffrey. 2008b. *The Neo-Aramaic dialect of Barwar*. Leiden: Brill. DOI: 10.1163/ej.9789004167650.i-2198.

Khan, Geoffrey. 2009. *The Jewish Neo-Aramaic dialect of Sanandaj.* Piscataway, NJ: Gorgias Press.

Khan, Geoffrey. 2012a. North Eastern Neo-Aramaic. In Stefan Weninger, Geoffrey Khan, Michael P. Streck & Janet C. E. Watson (eds.), *The Semitic languages: An international handbook*, 708–724. Berlin: De Gruyter Mouton.

Khan, Geoffrey. 2012b. The copula in North-Eastern Neo-Aramaic. In Federico Corriente, Gregorio del Olmo Lete, Ángeles Vicente & Juan-Pablo Vita (eds.), *Dialectology of the Semitic Languages: Proceedings of The IV Meeting On Comparative Semitics Zaragoza 06/9–11/2010*, 25–31. Sabadell/Barcelona: Editorial Ausa.

Khan, Geoffrey. 2016. *The Neo-Aramaic dialect of the Assyrian Christians of Urmi I-IV.* Leiden: Brill.

Khan, Geoffrey. 2019a. The Neo-Aramaic dialects of eastern Anatolia and northwestern Iran. In Geoffrey Haig & Geoffrey Khan (eds.), *The languages and linguistics of Western Asia: An areal perspective*, 190–236. Berlin: De Gruyter Mouton.

Khan, Geoffrey. 2019b. The Neo-Aramaic dialects of western Iran. In Geoffrey Haig & Geoffrey Khan (eds.), *The languages and linguistics of Western Asia: An areal perspective*, 305–353. Berlin: De Gruyter Mouton.

Khan, Geoffrey. 2020. Contact and change in Neo-Aramaic dialects. In Bridget Drinka (ed.), *Historical linguistics 2017: Selected papers from the 23rd international conference on historical linguistics, San Antonio, Texas, 31 July – 4 August 2017*, 387–407. Amsterdam: Benjamins.

Khan, Geoffrey, Masoud Mohammadirad, Dorota Molin & Paul M. Noorlander. 2022. *Neo-Aramaic and Kurdish folklore from Northern Iraq.* Cambridge: Open Book Publishers. DOI: 10.11647/OBP.0307.

Kıral, Filiz. 2001. *Das gescprochene Aserbaidschanisch von Iran: Eine Studie zu den syntaktischen Einfüssen des Persischen.* Wiesbaden: Harrassowitz.

MacKenzie, David N. 1966. *The dialect of Awroman (Hawraman-i Luhon): Grammatical sketch, texts, and vocabulary.* English. Kobenhavn: Kommissionaer/Munksgaard.

Matras, Yaron. 2009. *Language Contact.* Cambridge: Cambridge University Press.

Mohammadirad, Masoud. 2024. Zagros region: The Kurdish-Gorani continuum. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 245–279. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266347.

Molin, Dorota. 2021. *The Jewish Neo-Aramaic dialect of Dohok: A comparative grammar.* Cambridge: University of Cambridge. (Doctoral dissertation).

Mutzafi, Hezy. 2004. *The Jewish Neo-Aramaic dialect of Koy Sanjaq.* Wiesbaden: Harrassowitz.

Mutzafi, Hezy. 2008. Trans-Zab Jewish Neo-Aramaic. *BSOAS* 1. 409–431.

Napiorkowska, Lidia. 2015. *The Neo-Aramaic dialect of Diyana-Zariwaw.* Leiden: Brill.

Noorlander, Paul M. 2014. Diversity in convergence: Kurdish and Aramaic variation entangled. *Kurdish Studies* 2(2). 201–224.

Noorlander, Paul M. 2021. *Ergativity and other alignment types in Neo-Aramaic: Investigating morphosyntactic microvariation.* Leiden: Brill.

Noorlander, Paul M. 2022a. NE Neo-Aramaic (Christian) Shaqlawa. In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA – Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

Noorlander, Paul M. 2022b. NE Neo-Aramaic (Christian) Urmi. In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA – Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

Noorlander, Paul M. 2022c. NE Neo-Aramaic (Jewish Sanandaj). In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA – Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

Noorlander, Paul M. 2024. Arabic and Aramaic in eastern Anatolia. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 471–514. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266361.

Noorlander, Paul M. & Dorota Molin. 2022. Word order typology in North Eastern-Neo-Aramaic: Towards a corpus-based approach. In Hiwa Asadpour & Thomas Jügel (eds.), *Word order variation Semitic, Turkic and Indo-European languages in contact*, 235–258. Berlin: De Gruyter Mouton.

Noorlander, Paul M. & Donald Stilo. 2022. NE Neo-Aramaic (Jewish Urmi). In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA – Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

Nourzaei, Maryam & Geoffrey Haig. 2024. Balochi: A cross-dialect investigation of post-verbal elements. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 121–154. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266337.

Panoussi, Estiphan. 1990. On the Senaya dialect. In Wilhelm Heinrichs (ed.), *Studies in Neo-Aramaic*, 107–129. Atlanta, GA: Scholars Press.

Silva-Corvalán, Carmen. 1994. *Language contact and change: Spanish in Los Angeles*. Oxford: Clarendon.

Silva-Corvalán, Carmen. 2008. The limits of convergence in language contact. *Journal of Language Contact* 2(1). 213–224.

Stilo, Donald. 2005. Iranian as a buffer zone between the universal typology of Turkic and Semitic. In Éva 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. 2015. An introduction to the Atlas of the Araxes-Iran linguistic area. In Bernard Comrie & Lucía Golluscio (eds.), *Language contact and documentation/Contact lingüistico y documentación*, 343–356. Berlin: De Gruyter Mouton.

Tsereteli, Konstantin. 1976. Образцы современной ассирийской речи (саламасские тексты). In Institut naradov Azii & Institut vostokovedeniya (eds.), *Семитские языки: сборник статей III*, 208–219. Moskva: Nauka.

Younansardaroud, Helen. 2001. *Der neuostaramäische Dialekt von Särdä:rid*. Wiesbaden: Harrassowitz.

Chapter 16

Arabic and Aramaic in eastern Anatolia

Ⓘ Paul M. Noorlander
University of Cambridge

This chapter offers a brief overview of the word order typology of *Qəltu*-Arabic and Neo-Aramaic dialects spoken by minorities in southeastern Anatolia. Constituent ordering is generally consistent with the typology of VO languages, and representative of the majority of Central Semitic languages. Convergence with local languages, however, has resulted in the development of often post-predicate copulas and a higher rate of OV order, and in some doculects even a complete shift to OV.

1 Introduction

For centuries, Jews, Christians and Muslims have co-existed in the historical region of eastern Anatolia. Jewish (J.) and Christian (C.) communities used to speak their own Aramaic and/or Arabic variety, predating the arrival of Turkish and Kurdish. Aramaic was one of the principal languages in Syria, Anatolia and Mesopotamia before the Islamic period, and following the Arab conquests, most of the Jews and Christians gradually shifted to Arabic, leading to a diversity of regional and communal dialects. Nowadays, Anatolia is characterized by increasing nationalization and homogeneity within a predominantly Turkish-speaking Muslim society. The ethnic cleansing during the First World War alongside continuous persecution and systematic marginalization of minorities led to a massive displacement of these minorities. Virtually all Jews left the region under duress to Israel after 1948.

Linguistically, eastern Anatolia constitutes part of a continuum of Arabic and Aramaic dialects that once extended from Syria-Palestine to modern-day Iraq and Iran. Figure 1 displays a map of the original locations and distribution of several Arabic and Aramaic dialects in Anatolia.
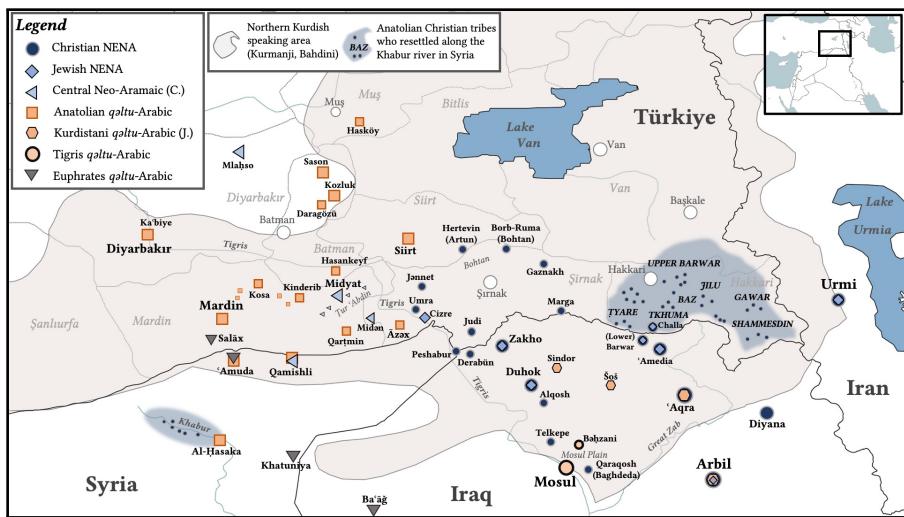
Figure 1: *Qəltu*-Arabic and Neo-Aramaic dialects in southeastern Anatolia

The present outline takes a corpus-based approach to word order following the design of the WOWA corpus (Haig et al. 2022 and Haig et al. 2024 [this volume]) and relies on datasets from within and without WOWA. Those approaches (e.g. Dahlgren 1998; El Zarka & Ziagos 2020) that exclude pragmatically-driven fronting or topicalization to clause-initial position from word order statistics may reach different conclusions than the current chapter. Table 1 lists the WOWA datasets in *Qəltu*-Arabic, as well as Central and Northeastern Neo-Aramaic, used in this chapter with their respective sources.[1] These datasets were designed for the corpus-based analysis of non-subject arguments (see Dahlgren 1998 for a corpus-based study of subjects in Arabic, and Molin 2021 for that in NENA) and their respective position before or after the predicate in accordance with the framework and coding guidelines of the WOWA databank.[2] The sets for Barwar and J. Dohok, though dialects originally spoken in northwestern Iraq, are, for all practical purposes, considered also representative of the majority of NENA dialects in eastern Anatolia (see Noorlander 2024 [this volume] for a discussion of NENA in Iran and northeastern Iraq). A handful of object tokens were also counted for the NENA dialect of Bohtan (Fox 2009: 116–137) and Hertevin (Jas-

---

[1]Numbered texts and numbered segments are separated by colons, e.g. 25:§2 means Text 25, Paragraph 2, and page numbers and segments by periods, e.g. 101.§2, Page 101, Paragraph 2.

[2]See https://multicast.aspra.uni-bamberg.de/resources/wowa/data/_docs/guidelines/wowa_coding-guidelines.pdf

trow 1988), which are not part of the WOWA corpus. Moreover, Arabic data outside of the WOWA corpus were taken from Noorlander (2023) for Kinderib and Daragözü *Qəltu*-Arabic and Cilician Levantine Arabic.

Table 1: Datasets from the WOWA corpus discussed in this chapter

Doculect	Speakers	Total tokens	Analysed tokens	Source
Ṭuroyo, Midyat, CNA	2	1014	778	Noorlander (2022d) based on the digitalization by with Lyavdansky et al. (2020) of Ritter's (1967) texts 1–2, 24, 27
Mlaḥso, CNA	2	824	703	Noorlander (2022c) based on Jastrow (1994: 74–129)
J. Dohok, NENA	4	916	517	Molin (2022)
Barwar, NENA	2	963	963	Stilo (2022) based on Khan (2008a)
Kaʿbiye, *Qəltu*-Arabic	3	788	643	Noorlander (2022a) based on Jastrow (2022: Texts II, IX, XII, XIV)
J. Baghdad, *Qəltu*-Arabic	2	1339	490	Noorlander (2022b)

1.1 Arabic

The Arabic dialects of modern-day Turkey belong to four major groups (Jastrow 2006):

1. *Anatolian Arabic*, i.e. sedentary Mesopotamian, divided into at least four subgroups

   a) Mardin (Muslim, Christian, Jewish);

   b) Siirt (Muslim, Christian);

   c) Diyarbakir (Christian, Jewish);

   d) Kozluk-Sason-Muş (Muslim).

2. *Khawetna Arabic* (Khatuniya), i.e. Bedouin Mesopotamian, stretching from northern Syria into Turkish Mardin (Talay 1999)

3. *Cilician-Antiochian Arabic* (Arnold 1998; Procházka 2002), a type of Levantine Arabic spoken by various ethnoreligious groups, of which Alevis constitute the majority, on the coastal region along the Mediterranean (Turkish Mersin, Adana and Hatay);

4. *Shawi Arabic*, originally Bedouin, in modern-day Urfa (Procházka 2003);

The Anatolian dialects belong to the so-called *qəltu*-subgroup of *Mesopotamian Arabic* that preserves the voiceless uvular stop /q/ and first singular suffix *-tu*, as in *qəltu* 'I said,' against the later arrivals in Mesopotamia of ultimately Bedouin origin which innovated a corresponding velar stop /g/ and lost the final *-u*, thus *gələt* 'I said' (Blanc 1964; Jastrow 1978, 2006; Talay 2012). Today, only a few Muslim speakers of *Qəltu*-Arabic remain in Siirt, Mardin and the mountains between Kozluk-Muş. Outside of Turkey, *Qəltu*-Arabic is also represented by:

- dialects spoken in Syria (Behnstedt 1992);

- Jewish dialects of Iraqi Kurdistan (Jastrow 1990);

- several varieties along the Tigris, including the various ethnoreligious communities of the Mosul Plain (Jastrow 1979) and the Jewish (Bar-Moshe 2019) and Christian (Abu-Haidar 1991) communities of Baghdad;

- and finally Muslim and Jewish communities along the Euphrates, spanning from Hīt in Iraq to Khatuniya in Syria and Salāx in Turkey (Talay 1999).

Anatolian *Qəltu*-Arabic dialects form a continuum with Levantine Arabic (Talay 2014), and, as minorities outside of the core Arabic-speaking regions, they

share similarities with other peripheral varieties of Arabic such as Cypriotic Arabic and Central Asian Arabic (Akkuş 2017); though, the former is of Levantine origin and the latter presumably of Iraqi origin.

The various other Bedouin dialects spoken in Iraq subsumed under Mesopotamian Arabic have more in common with the varieties of Arabia, which includes the Muslim *gələt*-dialects of Baghdad in contradistinction to the Jewish and Christian *qəltu*-dialects of the same city (Blanc 1964) as well as the Arabic of Khuzestan (Leitner 2024 [this volume]).

Overviews of Mesopotamian Arabic are offered by Jastrow (1978) and (Talay 2012), and that of Arabic varieties in Turkey by Jastrow (2006), Arnold (2015) and Procházka (2019). Word order has been an understudied area, although Jastrow (1978: 131–141) and Birnstiel (2022: 204–218) provide comparative studies of copula syntax, for example, and Dahlgren (1998) offers large-scale, corpus-based studies of Arabic word order, especially concerning the position of subjects. A comparative, corpus-based study of object placement in several Anatolian Arabic dialects can be found in Noorlander (2023).

1.2 Aramaic

Aramaic is represented in eastern Anatolia by

- the diverse group of *Northeastern Neo-Aramaic* (NENA) spoken by Jews and Christians of Iranian Kurdistan, Iranian Azerbaijan, Iraqi Kurdistan, and southeastern Anatolia;

- *Central Neo-Aramaic* (CNA) consisting of the Neo-Aramaic dialects known as Ṭuroyo spoken by the Christians of Ṭur ʿAbdin in modern-day Mardin and Şırnak, and the extinct dialect of Mlaḥso (Turkish: Lice) in Diyarbakir (Jastrow 1994).

This chapter focuses on the dialects in southeastern Anatolia, i.e. the dialects of Ṭuroyo and Mlaḥso, as well as the NENA dialects in the western periphery. (The eastern periphery is treated in Noorlander 2024 [this volume]).

The Neo-Aramaic dialects of rural Ṭur ʿAbdin and that of the city of Midyat exhibit slight variation, while Ṭuroyo in general is distinct from Mlaḥso. Waltisberg (2016), relying primarily on Ritter (1967), provides a corpus-based overview of Ṭuroyo syntax.

NENA used to be spoken in Christian villages in Şırnak and Siirt, south of the Bohtan river, notably Hertevin (Aramaic name: Artun; Jastrow 1988), Borb-Ruma (i.e. Bohtan, Fox 2009) and Mount Judi (Sinha 2000). The Jewish NENA dialects

of Cizre (Nakano 1973) in Şırnak and of Challa (Fassberg 2010) in Hakkari belong to the so-called *Lishana Deni* cluster whose core region is in northwestern Iraq. The Christian NENA dialects of the tribes in Hakkari used to form a densely populated area, e.g. Tkhuma and Upper/Lower Ṭyare, extending into Turkish Van (Tsereteli 1963), Iranian Azerbaijan, and Iraqi Kurdistan, such as Lower Barwar (Khan 2008a). Several communities originating in Anatolia re-settled in Iraq or along the Khabur river northwest of Al-Hasaka in Syria (Talay 2008, 2009).

Talay (2008) and Khan (2019) provide overviews of the NENA varieties of eastern Anatolia. Individual grammars, for instance, of Barwar (Khan 2008a: 823–950), of Tel Kepe (Mosul Plain; Coghill 2018), of Jewish Zakho (Cohen 2012) and of Jewish Dohok (Molin 2024) offer detailed studies of information structure and word order. Noorlander & Molin (2022) offer corpus-based word order comparisons in Jewish NENA.

2 Word order profile

2.1 Noun phrases

Noun phrases display Numeral-Noun-Adjective order, e.g.

(1) *Qəltu*-Arabic Kinderib (Jastrow 2003: 2.5:§19)
 fə-θəθ   aṛbaʕ xams šəberi   zġār
 in-three.F four.F five.F straw_basket.PL little.PL
 'in three, four, five little straw baskets'

(2) NENA Txuma (Talay 2009: 166.§16)
 ʔarpa xamšá plaš-e   xelan-e
 four   five   battle.M-PL severe-PL
 'four, five severe battles'

(3) CNA Mlaḥso (Jastrow 1994: 130.§137, §139)
 tre aḥé...   ə=aḥó   rab-ó
 two brother.M.PL DEF.SG=brother.MSG big-MSG
 'the two brothers... the elder brother'

Noun-Numeral occurs with the numeral 'one' in *Qəltu*-Arabic and Ṭuroyo, as exemplified in (4–5). Adjective-Noun order also sporadically occurs (see §3.1.7.).

(4) *Qəltu*-Arabic Kaʿbiye (Jastrow 2022: 99)
    kaṛm   *wēḥəd*
    vineyard.MSG one.MSG
    'a vineyard'

(5) CNA Ṭuroyo, Midən (Jastrow 1985: 275.§9)
    barθo   *ḥðo*
    girl.FSG one.FSG
    'a girl'

Demonstratives precede the noun everywhere except in CNA where the demonstrative is a suffix added to the determined noun and/or to the adjective that immediately follows it (Waltisberg 2016: 46–47), as shown in (7c), which originated in the patterns Noun-Demonstrative and Noun-Adjective-Demonstrative. Adjectives agree in definiteness with their head nominal in *Qəltu*-Arabic and Ṭuroyo, cf. (6) and (7). Attributive demonstratives in the majority of NENA correspond to prefixal definite articles in CNA, e.g. *u=* < **hū* in (7a) and *a=* < **han* in (7b). The near deixis proclitic attributive demonstratives can also be augmented with a deictic suffix *-ha* in the NENA dialect of Hertevin (Jastrow 1988: 33–34), e.g. *ʾád=ʾoda-ha* 'this room over here,' which parallels the situation in Kurdish (see Mohammadirad 2024 [this volume]).

(6) *Qəltu*-Arabic Kaʿbiye (Jastrow 2022: XXI:§7)
    ād   *əl-ḥāfez*   *əl-məskin-∅*
    DEM.SG DEF-blind.MSG DEF-poor-MSG
    '**this** poor blind beggar'

(7)   a. CNA Ṭuroyo, Kfarze (Ritter 1967: 67:§92)
       ú=səsy-ayði   *ú=kom-o*
       DEF.MSG=horse.MSG-my DEF.MSG=black-MSG
       'my black horse'

    b. CNA Ṭuroyo, Anḥəl (Ritter 1967: 58:§119)
       *ám=medon-**ani***   *áḥ=ḥren-e*
       DEF.PL=thing.MPL-DEM.PL DEF.PL=other-PL
       '**those** other things'

    c. CNA Ṭuroyo, Midən (Jastrow 1985: 266.§9)
       á=[tre   *kŭrfe*   *kom]-**anək***
       DEF.PL=two.M snake.MPL black-DEM.PL
       '**those** two black snakes'

(8) NENA Hertevin (Jastrow 1988)
    ʔád=naša   *ṭaw-a*
    DEM.SG=man.MSG good-MSG
    '**this** good man'

(9) NENA Bohtan (Fox 2009: 122.§87)
    at   *abra*   *xen-a*
    DEM.SG man.MSG other-MSG
    '**this** other boy'

While the morphology of attributive constructions varies considerably, Noun-Genitive order predominates in both Arabic and Aramaic for both genitive nouns and pronouns, e.g.

(10) *Qəltu*-Arabic Āzəx (Jastrow 1981: VI6:§40)
    bayt   *əl-ḥakəm*
    house.MSG DEF-judge.MSG
    'the house of the judge'

(11) *Qəltu*-Arabic Kaʿbiye (Jastrow 2022: IV:§16)
    bayt   *abu-y*
    house.MSG father.MSG-my
    'the house of my father'

(12) CNA Ṭuroyo, Midyat (Ritter 1967: 24:§55)
    ú=bayto   *d-ú=taǧər*   *u*
    DEF.MSG=house.MSG GEN-DEF.SG=merchant.MSG and
    í=zangan-ayðe   *kul-a*
    DEF.FSG=wealth.FSG-his all-FSG
    'the house of the merchant and all his wealth'

(13) CNA Mlaḥso (Jastrow 1994: 126.§126)
    beytó   *d-ə=malkó*
    house.MSG GEN-DEF.SG=king.MSG
    'the house of the king'

(14) NENA Hertevin (Jastrow 1988: 156.§583)
    l-ʔarʔ-əd   *bab-ew*   *Yaqo*
    to-land-CSTR father.MSG-his Jakob
    'to the land of his father Jakob'

2.2 Verbal complements

2.2.1 Verb/object

Verb-Object order predominates in all relevant languages,[3] with the exception of Mlaḥso and the NENA dialect of Bohtan (Borb-Ruma), where Object-Verb predominates. Drawing on distinctions made in the WOWA corpus, however, argument type and definiteness are major factors in several VO doculects, as borne out by the statistics in Figure 2 categorized according to referentiality and identifiability, i.e. definite as opposed to indefinite NPs, and argument type, i.e. pronouns as opposed to nouns. "Pronoun," here, comprises independent personal and demonstrative pronouns, as illustrated in (15c) for Arabic and (16a) for Aramaic, while all remaining independent pronouns, such as reflexive and indefinite pronouns, are subsumed under "Other," exemplified in (15d) and (16d). In general, Figure 2 shows there is an increasing likelihood for post-verbal position across all doculects if the object is indefinite. Even in the case of Mlaḥso where OV order is largely grammaticalized, indefinite objects are slightly more likely post-verbal than their definite counterpart (see §3.1.1 on variation among speakers and the role of contact in Mlaḥso). This, however, does not mean that definiteness plays a significant role in every dialect. An additional, preliminary study of 96 direct object NPs and their placement in NENA Hertevin (Jastrow 1988: §§86–100, §§166–172, §§307–323, §§419–466) reveals no significant difference between definite and indefinite object NPs: 25 out of 35 (0.71) indefinites are post-verbal and 47 out of 61 (0.77) definites are post-verbal.

(15)   a.   Definite, OV
            Qəltu-Arabic Kaʿbiye (Jastrow 2022: II:§19)
            *mō na-ʕref   **madṛa**   əšṭōr ya-ḥṣəd-ū-**nu***
            NEG A.1PL-know millet.MSG how A.3-harvest-A.PL-O.3MSG
            'We do not know how to harvest **the millet**.'

       b.   Indefinite, VO
            Qəltu-Arabic Kaʿbiye (Jastrow 2022: II:§5)
            *tə-t-ṛūḥ-o   tá-ḥṣəd-o   **madṛa***
            FUT-S.2-go-S.PL A.2-harvest-A.PL millet.MSG
            'You shall go to harvest **millet**.'

---

[3]See Dahlgren (1998) on Arabic, Waltisberg (2016: 289–290) on Ṭuroyo, and Noorlander & Molin (2022) on NENA.
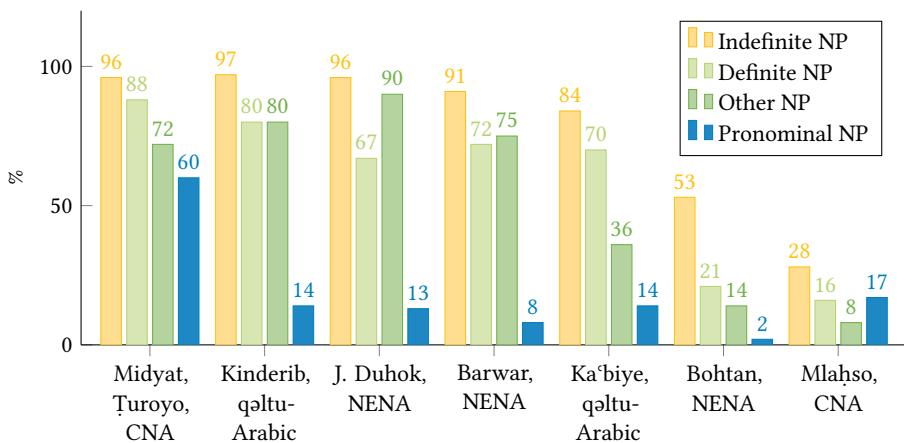
Figure 2: Rate of post-predicate (PP) nominal and pronominal objects

c. Pronominal, OV
   Qəltu-Arabic Kaʿbiye (Jastrow 2022: XIV:§2)
   ***ənti** ana t-āxəd-ki*
   2FSG 1SG FUT-A.1SG.take-O.2FSG
   'I will take **you**.'

d. Other, OV
   Qəltu-Arabic Kaʿbiye (Jastrow 2022: II:§7)
   ***kəll-en** qatl-ōn*
   all-of.them killed-A.3PL
   'They killed **them all**.'

(16) a. Definite, OV
   CNA Ṭuroyo, Midyat (Ritter 1967: 27:§25)
   ***í=hadiy-ayo** gə-mšayáʕ-no-le=yo*
   DEF.FSG=gift.FSG-DEM.FSG FUT-send-A.1MSG-R.3MSG=T.3MSG
   'I will send him **the present**.'

   b. Indefinite, VO
   CNA Ṭuroyo, Midyat (Ritter 1967: 27:§25)
   *ṭlə́b-le-le **hadiye** men-i*
   asked.PFV-A.3MSG-R.3MSG gift.FSG from-1SG
   'He asked **a present** for himself from me.'

    c. Pronominal, OV
       CNA Ṭuroyo, Midyat (Ritter 1967: 83:§50)

       haθe   *ono ló=ḥəzy-o-li*
       DEM.F.SG 1SG NEG=saw.PFV-O.3F.SG-A.1SG

       'I did not see **this one**.'

    d. Other, VO
       CNA Ṭuroyo, Midyat (Ritter 1967: 27:§54)

       mayta-lle   **kul-le**
       brought.PFV-A.3PL all-of.them

       'He brought **them all**.'

The tokens of object pronouns are low, since they are, by default, expressed with object suffixes on the verb; their independent expression thus attracts attention in the discourse. This notwithstanding, the data indicate their word order preference should not be characterised in the same manner as nouns, suggesting OV order instead in the NENA doculects and Kaʿbiye *Qəltu*-Arabic doculect.

A fronted definite object generally triggers object indexing, i.e. clitic doubling or cross-referencing verbal object suffixes, as illustrated in (15a) for Arabic and (16a) for Aramaic. Topicalization with pronominal resumption is well-known in Arabic linguistics and also occurs in Anatolian Arabic, e.g. see Wittrich (2001: 164–165) for examples of topicalization in the *Qəltu*-Arabic dialect of Āzəx. Moreover, pre-verbal placement is conditioned by specificity in at least the Kozluk-Sason-Muş dialects (Akkuş 2017). The object indexing originally used to be dedicated to topicalization and subsequently developed into a differential object coding strategy. Other studies of Arabic word order (Dahlgren 1998; El Zarka & Ziagos 2020), however, have treated such definite objects – often in clause-initial position – with additional object indexing on the verb as clause-external arguments, and not instances of OV. How frequently this pre-verbal placement occurs, be it due to topicalization or specificity, has not been studied in detail. Preposed objects of any kind – be they topicalized to clause-initial position or simply fronted before the verb – were subsumed under OV in the current approach, and the importance of treating them as such is borne out by the statistics in Figure 2. The additional factor of object indexing requires further investigation, but see Noorlander & Molin (2022) on Jewish NENA and Noorlander (2023) on Anatolian Arabic.

2.2.2 Verb/goal

Lexicalized arguments – in contrast to bound verbal person markers – denoting endpoint roles (see Haig et al. 2024 [this volume]) like the Goal of verbs of (caused) motion, such as 'go' and 'bring,' are generally post-verbal throughout, as given in the first column of Table 2. The second column of Table 2 combines the tokens of *Recipients* of ditransitive verbs like 'give' and *Beneficiaries*, and the last column shows the tokens of *Addressees* of verbs of speech like 'say'. These respective roles are illustrated for Kaʿbiye Arabic and Mlaḥso Aramaic in (17–18). Thus, endpoints are generally treated the same across doculects, except for Recipient/Beneficiaries and especially Addressees in Mlaḥsó, which are pre-verbal more often than in the other doculects (this is an areal phenomenon; see §3.1.4).

Table 2: Rate of post-predicate (PP) goal-like arguments

Doculect	G		R/Ben		Addr	
	n	PP	*n*	PP	*n*	PP
Midyat, Ṭuroyo, CNA	140	97%	24	96%	25	100%
J. Dohok, NENA	112	99%	34	95%	31	100%
Kinderib, *Qǝltu*-Arabic	76	100%	21	95%	(5	60%)
Kaʿbiye, *Qǝltu*-Arabic	87	92%	29	79%	12	75%
Mlaḥso, CNA	152	97%	42	69%	34	50%

(17) a. Verb-Goal
      Qǝltu-Arabic Kaʿbiye (Jastrow 2022: XIV:§4)
      *tǝ-ḥreb   tǝ-ği   **bayt***
      S.2fsg-flee S.2fsg-come house.msg
      'She flees **home**.'

   b. Verb-Recipient
      Qǝltu-Arabic Kaʿbiye (Jastrow 2022: XIII:§1)
      *ana t-a-ḥṭi   bǝnt-i   **šān ǝben   axū-y***
      I   fut-A.1sg-give daughter.fsg-my to   son.msg.cstr brother-my
      'I shall give my daughter **to the son of my brother**.'

   c. Verb-Addressee

      Qəltu-Arabic Kaʿbiye (Jastrow 2022: XIV:§24)

      əl-xōǧa   *y-qūl-∅*   ***šā-ll-ūlād***

      DEF-Khoja A.3M-say-A.SG to-DEF-boy.PL

      'The Khoja says **to the lads**.'

(18) a. Verb-Goal

      CNA Mlaḥso (Jastrow 1994: 108.§25)

      ase-lan   ***ʕayni beytó***

      went.PFV-S.1PL same house.MSG

      'We went **to the same house**.'

   b. Verb-Recipient

      CNA Mlaḥso (Jastrow 1994: 163.§117)

      ə=brat-ezav   *hiv-le*   ***lə=nošk-ano***

      DEF.SG=daughter.FSG-his gave.PFV-A.3MSG DEF.SG=person-DEM.MSG

      'He gave **that person** his daughter.'

   c. Addressee-Verb

      CNA Mlaḥso (Jastrow 1994: 90.§115)

      malkó ***el-áv*** *emir-le*

      king.MSG to-3MSG said.PFV-A.3MSG

      'The king said **to him**…'

Relative object ordering in three-argument clauses depends on construction type. In the *prepositional dative construction*, the prepositional Recipient comes second. In a *double object construction*, both arguments are treated like O, and the Recipient comes first. This alternation is shown in (19) for NENA.[4]

(19) a. Theme(O)-Recipient(Prepositional)

      NENA Hertevin (Jastrow 1988: §437)

      daḥ *b-yăw-aḥ-la*   *ḥakimut-an*   ***l-ohá***

      how FUT-give-A.1PL-O.3FSG government.FSG-our to-DEM.MSG

      'How are we to grant lordship over us **to that one**?'

   b. Recipient(O)-Theme(O)

      NENA Hertevin (Jastrow 1988: §437)

      yăw-á-lehən   ***ʔan* *čičoke*** *beʔe*

      give-A.3FSG-O.3PL DEM.PL chick.PL egg.PL

      'She would give **these chicks** eggs.'

---

[4]For more examples, see (Coghill 2014) and Noorlander (2018: 144–153) for NENA, Waltisberg (2016) for Ṭuroyo, Camilleri et al. (2014) and Birnstiel (2022) for Arabic.

The default order of the prepositional dative construction in both Ṭuroyo (Waltisberg 2016: 298) and NENA (e.g. Noorlander & Molin 2022: 251) as well as Anatolian *Qəltu*-Arabic is *Verb-Theme-Recipient*:[5]

(20) CNA Ṭuroyo, Midyat (Ritter 1967: 27:§30)
húle-le=ste *kallat* *ġalabe* **l-ú-ḥaloq-ano**
gave.PFV-A.3MSG=ADD money.PL much to-DEF.MSG-barber-DEM.MSG
'He gave **the barber** much money.'

(21) a. *Qəltu*-Arabic Kinderib (Jastrow 2003: 6.1:§24)
tə-ʕṭi-∅ *rġīf* **lə-ṣāḥdət** **ət-tannōṛ**
3F-give-SG flatbread.MSG to-owner.FSG.CSTR DEF-oven
'She gives **the owner of the oven** a flatbread.'

b. *Qəltu*-Arabic Hasankeyf (Fink 2020: 6.4.2:§5)
īlzam yə-ʕṭi-∅ *paṛāṭ* **l-əmm** **w-abū**
must A.3M-give-A.SG money.PL to-mother.FSG and-father.MSG
'He must give **his mother and father** money.'

c. *Qəltu*-Arabic Hasköy (Talay 2002: I.2.2.:§9)
qən-na *n-āxəz* *pēlāv* **šā ixṭ-aṭ-na**
wanted.PFV-A.1PL A.1PL-buy shoe.SG for sister-FSG.CSTR-our
'We wanted to buy shoes **for our sister**.'

2.2.3 Become/complement

Final sates of change-of-state verbs, such as 'become,' 'make,' 'turn into (tr./intr.),' tend to follow the predicate:

(22) a. *Qəltu*-Arabic Kaʿbiye (Jastrow 2022: XIII:§2)
ṣār-∅ *kačal*
became.PFV-S.3MSG bald
'He turned bald.'

b. *Qəltu*-Arabic Kaʿbiye (Jastrow 2022: IX:§11)
n-say-en *kəde* **pūšiy-āt**
A.1PL-make-O.3PL such headscarf-PL
'We make them **into headscarves**.'

---

[5]Birnstiel (2022: 218–230) points to the default order of Verb-Recipient-Theme, but this does not seem to hold for all of *Qəltu*-Arabic.

(23)   a.   CNA Mlaḥso (Jastrow 1994: 112.§50)

kul-én   ve-len   nayar-ezan

all-of.them became.ᴘꜰᴠ-S.3ᴘʟ enemy.ᴘʟ-our

'They all became our enemies.'

   b.   CNA Mlaḥso (Jastrow 1994: 116.§75)

*băṭrăk   Elyás sim-le   el-áv   **šammás***

patriarch.ᴍsɢ Elyas made.ᴘꜰᴠ-A.3ᴍsɢ ᴅᴏᴍ-3ᴍsɢ deacon.ᴍsɢ

'Patriarch Elyas made him **deacon**.'

In dialects with a higher rate of OV, such as the NENA dialect of Bohtan, the relative placement of the argument of the verb 'make,' distinguishes between the preverbal direct object and the postverbal endpoint denoting the final state, as reflected in the following example (see Noorlander 2024 [this volume] for parallels in Iran):

(24)   a.   Object-Verb

NENA Bohtan (Fox 2009: 55)

xa kaboba*   *iwad-le

one kebab.ᴍsɢ make.ɪᴍᴘᴠ-O.3ᴍsɢ

'Make **a kebab**!'

   b.   Verb-Complement

NENA Bohtan (Fox 2009: 56)

*wəd-lo-le   **kaboba***

made.ᴘꜰᴠ-A.3ꜰsɢ-O.3ᴍsɢ kebab.ᴍsɢ

'She made it **into a kebab**.'

2.2.4 Other obliques

Post-verbal position is the most common for obliques. The first column of Table 3 gives the statistics of place and source constituents, illustrated in (25a) for Arabic and (26a) for Aramaic. Pre-verbal placement, by contrast, is common in Kaʿbiye Arabic, e.g. (25b), and predominates in Mlaḥso, e.g. (26b). Moreover, while all obliques generally behave similarly across doculects, instruments, given in the second column of Table 3, are more likely pre-verbal than other obliques in Ṭuroyo Aramaic.

(25)   a.   Verb-Oblique

Qəltu-Arabic Kinderib (Jastrow 2003: 7.6:§3)

*t-haššəs-a   **fə-l-ṃayy***

A.3ꜰsɢ-make.swell-O.3ꜰ in-ᴅᴇꜰ-water

'She swells them **in the water**.'

Table 3: Rate of post-predicate (PP) oblique arguments (pronominal and nominal)

Doculect	Place/Source		Instrument	
	n	**PP**	***n***	**PP**
Kinderib, *Qəltu*-Arabic	43	98%	13	92%
Barwar, NENA	177	74%	21	71%
Midyat, Ṭuroyo, CNA	121	88%	22	55%
Kaʿbiye, *Qəltu*-Arabic	82	61%	19	47%
Mlaḥso, CNA	56	34%	13	15%

b. Oblique-Verb
   Qəltu-Arabic Kaʿbiye (Jastrow 2022: II:§16)
   fə-l-ṃayye *faṭṭs-uw-a*
   in-DEF-water drowned.PFV-A.3PL-O.3FSG
   'They drowned her **in the water**.'

(26) a. Verb-Oblique
   CNA Ṭuroyo (Ritter 1967: 1:§1)
   yətáw-no-wo   ***b-Băbak***
   be.settled-S.1MSG-PST in-Bebek
   'I used to live **in Bebek**.'

   b. Oblique-Verb
   CNA Mlaḥso (Jastrow 1994: 80.§50)
   b-ə́-hawše   ***d-ə́-deyro***   *yativ-ina*
   in-DEF.SG-court.MSG GEN-DEF.SG-church.MSG sat.ANT-S.1PL
   'We sat down **in the churchyard**.'

2.3 Other predicates

2.3.1 Copulas

Semitic languages generally lack a verbal copula in present realis clauses. The juxtaposition of a (pro)noun and a nominal predicate are sufficient, and this occurs in all relevant major subgroups of Arabic and Aramaic in eastern Anatolia. The use of pronominal copulas, however, is a hallmark of *Qəltu*-Arabic as well as Central and Northeastern Neo-Aramaic (and a well-documented areal

phenomenon, see §3.2). Their relative position not only varies across dialects, but also depends on clause type. Overall, pronominal copulas exist for all clause types, except past and irrealis clauses, where a copular 'be' verb is preferred, e.g. Arabic *kwn*, Aramaic *hwy*. The paradigmatic organisation of pronouns and copulas is closely intertwined in Central Neo-Aramaic and *Qəltu*-Arabic, as compared for the peripheral dialects in Table 4. It is not unlikely that the pronominal inflection in Mlaḥso was based on an Arabic model. An overview with contrastive illustrations from a sample of the relevant languages is given in Table 5.[6]

Table 4: Comparison of personal pronouns and copula in Sason Arabic and Mlaḥso

	Sason Arabic (Akkuş 2017: 14)		**Mlaḥso, CNA** (Jastrow 1988)	
3MSG	*iyu*	*=ye*	*hiye*	*=yo*
3FSG	*iya*	*=ye*	*hiya*	*=yo*
3PL	*iyen*	*=nen*	*hiyen*	*=ne*

Generally, the negative and relative copula precede the predicate, as does the so-called deictic copula–or "demonstrative" or "presentative" copula–, which asserts an actual state of affairs in the immediately observable present, often with speaker or listener deixis. The latter is thus frequently used after verbs of perception, generally incompatible with questions, and ultimately derived from a deictic particle combined with a pronominal element. These three copula types are illustrated in (27) and (28).

(27) a. Negative copula
          Qəltu-Arabic Qartmin (Jastrow 1978: 137)
          ma-nne   *fə-l-bayt*
          NEG-COP.3MSG in-DEF-house.MSG
          '**They are not** at home.'

      b. Relative copula
          Qəltu-Arabic Qartmin (Jastrow 1978: 138)
          la-nne   *fə-s-suri*
          REL-COP.3PL in-DEF-Syria
          '(Yazidis) **who are** in Syria'

---

[6]On copula syntax in *Qəltu*-Arabic, see Jastrow (1978: 131–141) and Birnstiel (2022: 204–218), on that in NENA, see e.g. Khan (2002: 322–330, 2008a 823–842), Cohen (2012: 30–65) and Molin (2024: 140–174), on that in Ṭuroyo, see Waltisberg (2016: 112–122, 208–210, 238–240).

c. Deictic copula
   Qəltu-Arabic Qartmin (Jastrow 1978: 140)
   kəná   *rəfqət-u*   *ǧaw*
   DEIC.COP.3PL companion.MPL-his came.PFV.3PL
   '**Look! There they are**, his companions are coming.'

Table 5: Comparison of copula placement in *Qəltu*-Arabic and Central Neo-Aramaic

Qəltu-Arabic				
Siirt[a]	**Kinderib**[b]	**Kaˤbiye**[c]	**Hasankeyf**[d]	
mayye fəlbayt	*mayye fəlbayt*	*fəlbayt maye*	*mo fəlbayt=e*	'She is not at home.'
əlbənt iyy fəlbayt	*əlbənt fəlbayt=ye*	*əlbənt fəlbayt=ye*	*əlbənt fəlbayt=e*	'The girl is at home.'
əlbənt ayysap iyye?	*əlbent aynī=ye?*	*əlbənt əndaḥ=ḥe?*	*əlbənt angəs=e?*	'Where is the girl?'
NENA		**CNA**		
J. Challa[e]	**Bohtan**[f]	**Ṭuroyo**	**Mlaḥso**[g]	
lewa go besa	*lewa bata*	*latyo bú=bayto*	*b-beytó letyo*	'She is not at home.'
brata (ʔila) go besa(=la)	*brata bata=la*	*í=barθo bú=bayto=yo*	*ə=bartó b-beyto=yo*	'The girl is at home.'
ʔaya ma=yle?	*awa m=ile?*	*hawo mən=yo?*	*awo mən=yo?*	'What is that?'

These are hypothetical examples to illustrate the patterns.

---

[a]Jastrow (1978: 137)
[b]Jastrow (1978: 137), Lahdo (2009: 73–76, 172–175)
[c]Jastrow (2022: 47)
[d]Fink (2020: 76–77, 152–153)
[e]Fassberg (2010: 100–101)
[f]Fox (2009)
[g]Jastrow (1994: §35, §62, §142)

(28)   a.   Negative copula
       CNA Midən, Ṭuroyo (Ritter 1967: 115:§206)
       ám=may-ani   *dəθxu* **lan-ne**   *basim-e*
       DEF.PL=water.PL-DEM.PL GEN.2PL NEG-COP.3PL nice-PL
       'That water of yours **is not** tasty.'

       b.   Relative copula
       CNA Midən, Ṭuroyo (Ritter 1967: 115:§24)
       d-kət-ne   *yatiw-e*
       REL-EXIST-COP.3PL seated-PL
       '(those) **who are** seated'

       c.   Deictic copula
       CNA Midən, Ṭuroyo (Ritter 1967: 116:§48)
       ka-lən=ne   *tamo*
       DEIC-3PL=COP.3PL there
       '**Look, they are** there!'

The default position of the copula in affirmative present realis clauses is after
the nominal predicate in the majority of the Anatolian *Qəltu*-Arabic varieties as
well as in that of Neo-Aramaic. The copula, however, is not necessarily clause-
final, as shown in (29b), nor is it obligatory, as shown in (29c). The predicate
is expected to be downgraded to the background when following the copula in
a focalisation strategy like (29b) where the subject pronoun is in narrow focus
for the purpose of identification or specification. The copula can be completely
lacking in clauses like (29c), a structure that is akin to the original Semitic non-
verbal clause, reflecting *Topic-Comment* order. The frequency of this structure is
unknown, and this has not been coded in the WOWA corpus.

(29)   a.   Default
       CNA Midən, Ṭuroyo (Ritter 1967: 79:§12)
       *ono Ḥóre=**no**!*
       I   Hore=COP.1SG
       'I **am** Hore!'

       b.   Focalisation
       CNA Midən, Ṭuroyo (Ritter 1967: 82:§40)
       *óno=**no***   *Šēx*   *Dhām*
       I=COP.1SG Sheikh Dham
       '**It is** I, Sheikh Dham.'

c. Absent
CNA Midən, Ṭuroyo (Ritter 1967: 116:§1)
ono Slemắn Ḥanna Maskobi
I   Sleman Hanna Maskobi
'I am Sleman Hanna Maskobi.'

Pre-posed copulas, in turn, are a typical trait of the Arabic dialects of Siirt (Lahdo 2009), comparable to the situation in Levantine Arabic, except for interrogative clauses (Jastrow 1978: 132, and §3.2). Across NENA dialects in the region, pre-predicate placement of the affirmative copula is possible when the predicate serves the purpose of identification or when it expresses a transitory state, in which case the predicate will be most often adverbial, e.g. (30a) (see, for example, Molin 2022: 219–262 regarding J. Dohok, NENA). The latter parallels the structure of the deictic copula, as shown in (30b).

(30)   a. Present affirmative copula
NENA J. Challa (Fassberg 2010: 5.6:§17)
*ʔắya =**le***   *l-tama*
he   =COP.3MSG on-there
'He **is** there.'

b. Present deictic copula
NENA J. Challa (Fassberg 2010: 102)
ʕaqida   *dexun,*   ***wəl-le***   *l-axxa*
military.leader 2PL.GEN DEIC-COP.3MSG on-here
'Your leader, **he is** right here.'

2.3.2 Auxiliaries

Tense-Aspect-Mood is generally expressed by preverbal uninflected particles, which may have originated in auxiliaries. Complements of modal and phasal verbs immediately follow, as given in (31). Deictic copulas are also pre-verbal, conveying imminent or ongoing events in the immediately observable present (see §3.2 for examples).

(31)   Auxiliary-Verb-Object

a. *Qəltu*-Arabic Kaʕbiye (Jastrow 2022: II:§2)
***mo** **y-ṭiq**-∅*   *yə-šreb-∅*   *ṃayy əl-fəšqi*
NEG S.3-can-MSG A.3-drink-MSG water DEF-dung
'**He cannot** drink the sewage water.'

b. CNA Midyat, Ṭuroyo (Ritter 1967: 24:§12)

hat **lú-k-qudr-ət** Ø-*mbaṭl-ət* *úw-*
you.sg neg-ind-can-S.2msg sbjv-stop-A.2msg def.msg-
amro *d-aloho*
command.msg gen-god.msg
'You **cannot** thwart God's command.'

2.3.3 Complements of non-finite verbs

The bare object of infinitives regularly precedes the verb in the NENA dialects of Hertevin and Bohtan, illustrated in (32-33) below; cf. by contrast, NENA Barwar (Khan 2008b: B5:§175) *lewa mšurya xala **gəlla*** 'It had not started eating **grass**'; see also §2.3.3 in Noorlander (2024 [this volume]). Interestingly, however, in Bohtan the general rate of post-verbal object NPs is 0.33 (out of 93), while this is 0.75 (out of 96) in Hertevin. In Herevin, therefore, only objects of infinitives have shifted to OV.

(32) NENA Hertevin (Jastrow 1988: §318-§319)

a. Object-Verb(Infinitive)
b-dar-aḥ *ʾida* **gəlla** *člaʾa*
fut-throw-A.1pl hand.fsg grass.msg mow.inf
'We started mowing **grass**.'

b. Verb(Finte)-Object
člé^ʾ-laḥ-le **gəll-an**
mowed.pfv-A.1pl-O.3msg grass.msg-our
'We mowed our grass.'

(33) NENA Bohtan (Jastrow 1988: §118)

a. Object-Verb(Infinitive)
ani darəš-i **qanyon-e** *xlowa*
they begin-A.3pl sheep-pl milk.inf
'They started milking **sheep**.'

b. Verb(Finite)-Object
*oyün=se xolü-*Ø **qanyon-e**
he=add milk-A.3msg sheep-pl
'He too was milking sheep.'

3 Areal issues

Turkish and Northern Kurdish have been the dominant languages in eastern Anatolia since recent times, and Aramaic-Armenian bilingualism presumably also occurred (Jastrow 1994: 3). Aramaic and Arabic have been in contact with each other from the beginning and both at various stages with Persian – and Greek –in Antiquity. Moreover, Aramaic and *Qəltu*-Arabic varieties of Anatolia and the Mediterranean Region were an integral part of dialect continua stretching from the Levant through Anatolia to Mesopotamia (Jastrow 2007). Intra-Semitic contact and Semitic-Iranian contact therefore has a long and complex history.[7]

Finding themselves between mainstream Arabic, a prototypical example of a head-initial language, and Turkish, that of a head-final language, the relevant Arabic and Aramaic varieties exhibit numerous hallmarks that can be characterized against the backdrop of these two spheres of influences. At the same time, these contact-induced typological traits cannot be disentangled from language-internal developments. Iranian languages like Kurdish, in turn, have a more mixed word order typology, being in several respects rather similar to Semitic, except for the more rigid preverbal object position, the fixed clause-final copula placement, and the higher number of postpostions, which exhibit distinctly head-final syntax. While the influence of such neighbouring V-final languages – be they Iranian or Turkic – is undeniable, intensive exposure to contact with such languages may not always radically change word order typology. For example, the Arabic and Aramaic varieties – originally – spoken in Iranian Khuzestan are still characterized as VO (Häberl 2011: 735; El Zarka & Ziagos 2020), suggesting that, if this characterization is correct – which is still a matter of debate –, word order can be remarkably stable. This notwithstanding, local OV languages doubtless affected the peripheral Anatolian Arabic dialects belonging to the Diyarbakir or Kozluk-Sason-Muş cluster and the Neo-Aramaic dialect of Mlaḥso in Diyarbakir and that of Bohtan in Siirt. Furthermore, a higher rate of OV can be indicative of greater word order flexibility for pragmatically driven configurations, which could yet need not result from contact, and this is presumably the situation in the majority of NENA dialects. In fact, it is plausible that continuous interaction with the wider Arabic-speaking world reinforced more rigid VO syntax, thereby serving as an anticatalyst against a shift to OV or more flexibile word order, although much more comparative data is needed to establish

---

[7]For areal perspectives on Anatolia, see inter alia Haig (2001, 2014, 2017); Matras (2009: 270); Haig & Khan (2019); Khan (2019); Donabedian & Sitaridou (2021); on Aramaic and Kurdish specifically, see Noorlander (2014), and on Arabic in Anatolia, e.g. Talay (2007), Procházka (2020) and Akkuş (2020).

how frequently, for instance, object topicalization occurs across Arabic dialects. Thus, the rather inflexible VO syntax in Ṭuroyo may well be the result of Arabic influence, and not necessarily an archaic feature. The greatest extent of Arabic influence on NENA, in turn, is observed on the Mosul Plain in northern Iraq (Khan 2002; Coghill 2020).

3.1 Kurdish and Turkish influence

3.1.1 Object/verb

The data suggest that a combination of both the area-specific contact situation and the language-specific syntax of definite objects reinforced OV dominance. Figure 3 shows the relative frequencies in percentages of post-predicate objects in several Arabic and Aramaic dialects with the variables of definiteness and indefiniteness. The more rigid VO varieties are found on the left of the figure, with higher rates of VO overall in Jewish Baghdadi *Qǝltu*-Arabic (Noorlander 2022b), and the more rigid OV varieties on the right, with the lowest rates of VO overall in Mlaḥso Neo-Aramaic (Diyarbakir). The difference between definite
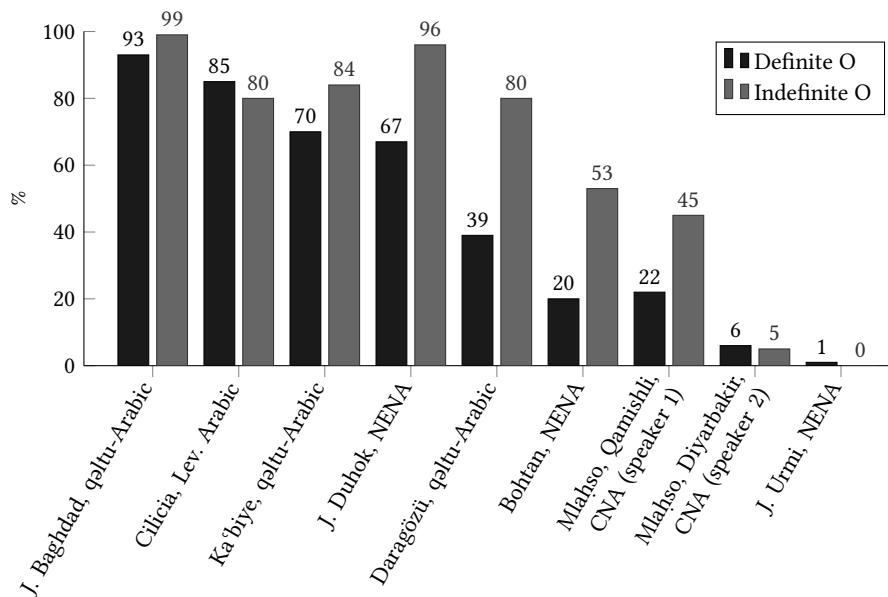
Figure 3: The rate of post-predicate (PP) definite and indefinite objects in Arabic and Aramaic

and indefinite objects, as indicated by the first and second bars, respectively, is smallest in those dialects closer to the ends of the spectrum.

Starting from right (OV) to left (VO), significant variation is observed among Mlaḥso speakers. Speaker 2 recorded in Diyarbakir, has conventionalized OV order consistently, showing extensive imposition of Turkic and/or Iranian, much like the Jewish NENA dialect of Urmi (Noorlander 2024 [this volume]). This is different from the other idiolect of Mlaḥso, Speaker 1, recorded in Qamishli, where the slightly higher rate of VO order is likely due to interference with Arabic and/or Ṭuroyo. In the NENA dialect of Bohtan (Borb-Ruma), the high rate reflects a conventionalization of OV word order, but definiteness remains a factor, and indefinite objects are lagging behind in the VO-to-OV shift. Definiteness is even a stronger factor in the *Qǝltu*-Arabic dialect of Daragözü, which is part of Kozluk-Sason-Muş group of Anatolian Arabic, where the OV/VO split depending on definiteness is largely grammaticalized: OV order for definite objects against VO order for indefinite objects. While the same tendency is also reflected in especially the Jewish NENA dialect of Dohok (Molin 2022) and to some extent also in the *Qǝltu*-Arabic variety of Kaʿbiye, these two varieties waver more strongly towards VO. Finally, the statistics are also given for Cilician Levantine Arabic[8], showing, like Jewish Baghdadi, no significant difference, presumably due to contact with mainstream Arabic.

Historically, the complete syntacticization of OV as in Central Asian Arabic (e.g. Seeger 2002; Jastrow 2004) and in Jewish NENA of Iran (e.g. Khan 2020; Noorlander & Molin 2022; Noorlander 2024 [this volume]) is only sporadically observed in Anatolia, and definiteness remains a major factor in regulating object placement. While VO is certainly the more archaic order, alternative pragmatically driven configurations were presumably part of Central Semitic syntax as a whole, but perhaps not with equal frequency across the entire subgroup. The shift to OV in Central Asian Arabic is generally explained in terms of the conventionalization and thus increase in frequency of a former topicalization strategy, e.g. *That man – I saw him yesterday*, due to the exposure to intensive contact with Uzbek and Tajik (e.g. Versteegh 1984: 452; Ratcliffe 2005; Souag 2017: 56); the same holds for OV in NENA under Kurdish influence (Haig 2015: 410–412). These explanations are consistent with the data as well as with the view of contact-induced word order convergence as a result of the extension of a pre-existing parallel construction (e.g. Silva-Corvalán 1994, 2008[9]; Heine 2008).

---

[8]Contact with Turkish could still play a role here, see Noorlander (2023).

[9]I am indebted to G. Haig for directing my attention to this reference.

3.1.2 Possessum/existential

Possessum placement in locative-existential predicative possession correlates with object placement, as illustrated in (34–35). Lower rates of post-predicate possessums correlate with lower rates of post-predicate objects. The dialects are contrasted in Figure 4 where the line indicates the frequency of post-predicate possessums across dialects decreasing significantly from core into periphery, and the bars indicate the frequency of post-predicate definite objects.
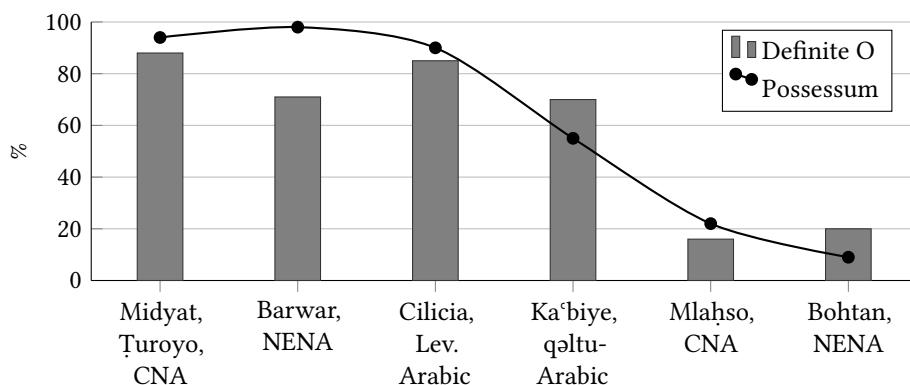
Figure 4: The rate of post-predicate definite and indefinite objects in Arabic and Aramaic

(34) a. Existential-Possessum
          Cilician Arabic Çukurova (Procházka 2002: 2.4:§1)
          *kān   il-a   **ibin***
          PST.COP.3MSG to-3FSG son.MSG
          'She had **a son**.'

      b. Possessum-Existential
          Qəltu-Arabic Kaʿbiye (Jastrow 2022: XIV:§12)
          *ənti   mādām **bayt**   **abu**   lə-ki*
          you.FSG because house.MSG father.MSG to-2FSG
          'because you have **a father house**'

(35) a. Existential-Possessum
          CNA Ṭuroyo (Ritter 1967: 27:§1)
          *kət-wo   šulṭono   kət-wo-le   **bəstono***
          EXIST-PST sultan.MSG EXIST-PST-A.3MSG garden.MSG
          'There was once a sultan who had **a garden**.'

b. Possessum-Existential
CNA Mlaḥso (Jastrow 1994: §71)
karme *hito el-əna*
vineyards.PL EXIST to-1PL
'We have **vineyards**.'

Sporadically, existential possessors – otherwise predicative – can even be expressed adnominally as they are in Kurdish (cf. Fox 2009: 116 fn. 103) and Turkish, cf. (36) and (37) with (38) and (39).

(36) Northern Kurdish
bira-yek=î ***min*** *he-ye*
brother-INDEF=EZ.MSG my   EXIST-COP.3SG
'**I have** a brother.'

(37) Turkish
*oğlu-**m** yok*
son-my NEG.EXIST
'**I have** no son.'

(38) CNA Mlaḥso (Jastrow 1994: 106.§12)
*aḥ-**i*** *hito*
brother.MSG-my EXIST
'**I have** a brother.'

(39) NENA Bohtan (Fox 2009: 4.1:§1)
oyün iwa *baxt-**əw**=u* *abr-**əw***
he   PST.COP.3SG wife.FSG-his=and son.MSG-his
'He **had** a wife and a son.'

3.1.3 Light-verb complements

Similarly, the non-referential complement of light verb constructions follows the verb where VO predominates, as given in (40–42).[10]

(40) T. *banyo etmek*
Qəltu-Arabic Mardin (Jastrow 1981: I.1:§60)
t-a-ği *a-sawiy-u* *ḅānyo*
FUT-S.1SG-come A.1SG-do-O.3MSG bathroom
'I shall come to wash him.'

---

[10]For more examples, see e.g. Talay (2007: 184); Akkuş (2020: 150); Procházka (2020: 97).

(41) T. *telefon etmek*, K. *telefon kirin*
      CNA Midyat, Ṭuroyo (Ritter 1967: 7:§13)
      səm-li   talafṓn   l-ú=   ḥakimo   d-áḥ=   ḥəyewən
      did.ᴘꜰᴠ-A.1sɢ telephone to-ᴅᴇꜰ.ᴍsɢ= doctor.ᴍsɢ ɢᴇɴ-ᴅᴇꜰ.ᴘʟ= animal.ᴘʟ
      'I called the vet.'

(42) T. *idare etmek*
      NENA Hertevin (Jastrow 1988: 156.§505)
      ∅-ʔod-aḥ-be   ʔidara
      sʙᴠᴊ-do-A.3ᴍsɢ-by.it management
      'so that we come through it'

The complement regularly precedes the light verb in dialects with a higher OV rate:

(43) T. *keyf etmek*, K. *kêf kirin*
      Qəltu-Arabic Ka'biye (Jastrow 2022: XIII:§4)
      kēf saw-ōn
      joy did.ᴘꜰᴠ-A.3ᴘʟ
      'They celebrated.'

(44) K. *kar kirin*
      CNA Mlaḥso (Jastrow 1994: 106.§17)
      l-á=ṭay-e   kar   sim-no
      for-ᴅᴇꜰ=muslim-ᴘʟ labour do-1ᴍsɢ
      'I shall work for the Muslims.'

(45) Russ. *sobraniye*
      NENA Bohtan (Fox 2009: 4.2:§2)
      sabroni   yawd-i
      meeting do-A.3ᴘʟ
      'They held a meeting.'

3.1.4 Addressee/verb & verb/goal

The Addressee placement in Mlaḥso also converges with Kurdish word order typology. In Mlaḥso the post-predicate rate of addressees is only 56% (18/32), whereas this rate is a 100% in Ṭuroyo (Midyat; 26/26). Mlaḥso also shows a higher rate of Beneficiary-Verb order (11/23) as opposed to Verb-Recipient order (16/17). Addressee-Verb and Beneficiary-Verb order as opposed to Verb-Goal order corresponds to the Northern Kurdish word order pattern in the same region (Haig

2022). Imposition from Iranian – rather than Turkish – is thus the most likely explanation for this syntactic split in Mlaḥso. Oblique-Verb order in Mlaḥso (§2.2.4) could be due to contact with either Iranian or Turkish.

3.1.5 Wh-in-situ

Wh-elements, or interrogatives, regularly remain in-situ in Kurdish, as given in (46a) for direct objects and (46b) for goals, although the latter can also undergo fronting as in (46c) (Haig 2022: 339). The same order would be obtained for Turkish, although, here, most arguments, including goals, regularly precede the predicate contrary to Kurdish (see Haig et al. 2024 [this volume]).

(46) Northern Kurdish

    a. Object *wh*-in-situ
       tu   *di-zan-î*   *min*   **çi** *kir*
       A.DIR.2SG IND-know-A.2SG A.OBL.1SG what did.PST
       'You know **what** I did.'

    b. Goal *wh*-in-situ
       tu   *di-č-î*   **kîve**
       A.DIR.2SG IND-go-S.2SG where
       '**Where** are you going?'

    c. Goal *wh*-fronted
       tu   *kîve*   *di-č-î*
       A.DIR.2SG **where** IND-go-S.2SG
       '**Where** are you going?'

Generally, wh-fronting occurs in the relevant Arabic and Aramaic dialects, as expected for VO typology. Occasionally, the wh-element stays in situ, as illustrated in (47). In the NENA dialect of Bohtan – and the CNA dialect of Mlaḥso – the object interrogative, stays in situ in line with their OV typology, as shown in (48). When wh-fronting occurs, the clause-initial slot remains open for a topical element, which usually is the subject. The resulting order of Subject-Interrogative-Predicate converges with that found in Northern Kurdish and Turkish, as illustrated in (48–50) below, and in §3.2.

(47) CNA Midən (Ritter 1967: 115:§7)
    adlalyo g-əzz-ano   **l-ayko**
    tonight FUT-go-S.1PL to-where
    '**Where** shall we go tonight?'

(48) NENA Bohtan (Fox 2009: 4.3:§38)
*yad-ət   ona **moy** wid-ena*
know-A.2MSG I   what do.PERF-A.1MSG
'You know **what** I have done.'

(49) CNA Midən (Ritter 1967: 73:§273)
*hat   **mə** ko-saym-ət*
you.SG what IND-do-A.2MSG
'**What** are you doing?'

(50) *Qəltu*-Arabic Hasköy (Talay 2002: I.2.4)
*ina **šəna** āsi*
I   what A.1SG.do
'**What** should I do?'

3.1.6 Postpositions?

The relevant Aramaic and Arabic dialects in contact with OV languages have maintained prepositional marking. In one case, however, the NENA dialect of Borb-Ruma (Bohtan) developed a postposition *=ləl* out of the preposition *lal-* (Fox 2009: 101–102), e.g. (51a). There is, however, no direct correspondence to a postposition in any of the neighbouring languages, which, in fact, would not generally use a posposition with inanimate goals, cf. (52b). When it attaches to the predicate, this is presumably an instance of convergence with the Northern Kurdish directional particle *=e* (Fox 2009: 101–102), cf. (51a) with (52a).

(51) a. Postposition
      NENA Bohtan (Fox 2009: 126:§139)
      *üzü-∅-wa   matwota=**ləl***
      go.ANT-S.3MSG-PST **village**.PL=DRCT
      'He had gone **to the villages**.'

   b. Directional particle
      NENA Bohtan (Fox 2009: 118:§35)
      *duwa   yar-o=**lal**   gawr-aw*
      woman.FSG say-A.3FSG=DRCT man.MSG-her
      'His mother says **to her husband**...'

(52)   a.   Directional particle and oblique
             Bahdini, Northern Kurdish
             ez   *di-çû-m=e*   ***mal-ê***
             1SG.DIR IND-go-S.1SG=DRCT house-OBL.F
             'I'm going **home**.'

        b.   Directional particle and oblique
             Bahdini, Northern Kurdish
             min   *got=e*   ***Mesûd-î***
             1SG.OBL say.PST=DRCT Masoud-OBL.M
             'I said **to Masoud**.'

3.1.7 Other OV correlates

Several other phenomena are related to the head-final typology of especially Turkish, which are again imposed more strongly in the northern and western periphery. A case in point is copula placement: in Mlaḥso Neo-Aramaic and the *Qəltu*-Arabic dialects of Diyarbakir and Kozluk-Sason-Muş areas, the post-predicate position has been conventionalised for also the negated copula. In the dialect of Hasankeyf, in turn, the exact parallel to Kurmanji occurs: the negator itself is pre-predicate but the copula post-predicate (see 5 in §2.3.1). This tendency is reflected in the statistics of the doculects: the rate of post-predicate copula complements is only 3% (2/65) in Mlaḥso (CNA) and 5% (4/79) in Kaʿbiye (Qəltu-Arabic) against 20% (14/71) in Ṭuroyo (CNA) and 35% (54/156) in C. Barwar (NENA).

While it is difficult to determine the language which ultimately provided the model for the development of post-predicate copulas in these Semitic languages, it is reasonable to assume an interplay of language-internal changes as well as contact-induced reinforcement and areal diffusion. The placement of a pronoun serving as the subject of a non-verbal clause, for instance, was not entirely fixed and its post-predicate position was part of the repertoire of Central Semitic. The starting point could have been interrogative clauses, as in the *Qəltu*-Arabic dialect of Siirt: the copula otherwise precedes the predicate but it is placed after the interrogative, which incidentally converges with Northern Kurdish syntax, as compared in (53–56). *Subject-Interrogative-Copula* is the common order in the majority of other varieties of Anatolian Arabic besides Neo-Aramaic. One can contrast this with the dialects of Arabic that did not develop an analytical copula, such as (55).

(53) Northern Kurdish (p.c. with Ergin Öpengin)
ev   çi-qas=e
DEM what-value=COP.3SG
'How much is this?'

(54) CNA Ṭuroyo
hano   məq-qa=yo
DEM.MSG what-value=COP.3MSG
'How much is this?'

(55) *Qəltu*-Arabic Āzəx (Jastrow 1978: 135)
hāza   b-áš-qad=u
DEM.MSG at-what-value=COP.3MSG
'How much is this?'

(56) *Gələt*-Arabic Khawetna (Talay 1999: 54)
b-əš-qadd   hāða
at-what-value DEM.MSG
'How much is this?'

The further extension of this post-predicate position to other contexts and its increasing obligatorisation was presumably not only due to contact with clause-final copula languages such as Iranian *par excellence*, but also embedded in a cluster of changes in the verbal system, and this applies especially to Aramaic. See Noorlander & Stilo (2015) for further parallel developments in the verbal system of Eastern Neo-Aramaic, which is completely derived from original verbal adjectives and enclitic pronouns that used to be mobile clitics, for example in Syriac (e.g. Noorlander 2018: 15), and that of other languages in the area, such as Northern Kurdish, where present tense endings, e.g. *-im* as in *diçim* 'I go,' are identical to the original verb 'to be,' e.g. *=im* as in *li vir=im* 'I am here'.

Moreover, Adjective-Noun order would be highly marked in all relevant Semitic languages but typical of Turkic. Incidentally, the Turkish loan adjective *dēri* 'another, last, next,' i.e. Turkish *diğeri* 'the other,' has been transferred with its corresponding Adjective-Noun order, e.g.

(57) *Qəltu*-Arabic Āzəx (Wittrich 2001: 121)
dēri yawm
next day.MSG
'the next day'

(58) CNA Ṭuroyo
deri yawmo
next day.MSG
'the next day'

In Kaʿbiye *Qəltu*-Arabic, Adjective-Noun order sporadically occurs under the influence of Turkish (Jastrow 2022: 7–8), cf. (59) and a tentative Turkish rendering in (60). Further research is necessary to investigate its frequency and distribution.

(59) *Qəltu*-Arabic Kaʿbiye (Jastrow 2022: VI:§45)
ktir kwayyəs faqad ṣūf   y-ṣīr-∅
very nice   INDEF wool.MSG S.3M-become-S.SG
'It will become some very nice wool.'

(60) Turkish
çok güzel bir   yün ol-acak-∅
very nice   INDEF wool become-FUT-S.3SG
'It will become some very nice wool.'

Similarly, there are numerous cases of pre-predicate final states of change-of-state verbs (§2.2.3) as shown in example (59) in peripheral dialects like Kaʿbiye Arabic due to Turkish influence. Thus, the rate of post-predicate final states is 60% (15/25) in Kaʿbiye (*qəltu*-Arabic) and 67% (24/36) in Mlaḥso (CNA) against 90% (18/20) in C. Barwar (NENA) and 100% (20/20) in Ṭuroyo (CNA).

Finally, in both Aramaic and Arabic, the standard of comparison, introduced with the source preposition *mən-* 'from,' seems to *precede* the adjective in Diyarbakir, as it does in local Turkish and Northern Kurdish varieties, which is consistent with the higher rate of OV order. The same also holds for Sason Arabic (Akkuş 2020: 144–145).

(61) *Qəltu*-Arabic Kaʿbiye (Jastrow 2022: IX:§19)
mən sáyn-na=ste   *ʔáxrab=we*
from language-our=ADD worse=COP.3MSG
'It is even worse **than our language**.'

(62) CNA Mlaḥso (Jastrow 1994: 112.§48)
hay-ó   *ṭaw*   **m-á=ṭay-e**=zi   *tə*
become.PERF-S.3SFG Muslim.MSG from-DEF.PL.Muslim-PL=ADD more
ḥarb-ó
bad-MSG
'She became Muslim, worse **than the Muslims** themselves.'

The opposite Adjective-Standard order predominates elsewhere, cf. (63–64) (and see Waltisberg 2016: 50–51, 117–118 for more examples), even in the NENA dialect of Bohtan (Borb-Ruma) where OV order is the most common, as shown in (65).

(63) *Qəltu*-Arabic Mardin (Jastrow 1981: I2:§32)

ʕamm-i	*agbaṛ*	***mən abū-y***	*kān*
uncle.msg-my	bigger	from father.msg-my	cop.pst.3msg

'My uncle was older **than my father**.'

(64) CNA Midən, Ṭuroyo (Ritter 1967: 83:§39)

ono rab-∅	***mín-ux=no***
I big-cmpr	from-2msg=cop.1sg

'I am older **than you**.'

(65) NENA C. Bohtan (Fox 2009: 95)

ay	*brota*	*ṭo*	*qaryan-ita=la*	***mənn-ət***	***d-aw***
dem.fsg	girl.sg	more	short-fsg=cop.3fsg	from-cstr	gen-dem.msg

abra
boy.msg

'This girl is shorter **than that boy**.'

3.2 Levant-Anatolia continuum

Several typological features indicate diffusion into eastern Anatolia from the Levant and Mesopotamia, resulting in many parallels between Aramaic and Arabic (see e.g. Weninger 2012 and Procházka 2020), and to some extent also Iranian, to name a few: First of all, differential object indexing (also known as clitic doubling) possibly combined with the preposition *l-* is a feature shared by Aramaic (e.g. Coghill 2014; Noorlander 2021: 290–294, 307–308, 350–370) and Arabic (e.g. Souag 2017); this, however, correlates with pre-posed objects especially in Anatolian *qəltu*-Arabic and Ṭuroyo. The correlation between differential object indexing and word order requires further investigation, but see also §3.1.1 on definiteness, of which indexing may be an epiphenomenon.

Verb-Goal and Become-Complement order is shared by Semitic and Kurdish more widely (Haig 2015, 2022), more specifically Verb-Addressee placement converges with Bahdini, i.e. southeastern varieties of Northern Kurdish (Haig 2022: 354–359).

In nominal syntax, Arabic adjectives such as *ʔawwəl* 'first,' *θēni* 'next, another,' *ġēr* 'other,' and *flān* 'so-and-so' are borrowed with their respective

Adjective-Noun order in Neo-Aramaic (see also Waltisberg 2016: 40–41), e.g.
NENA Hertevin (Jastrow 1988) *plan dokta* 'such-and-such a place', NENA Bohtan (Fox 2009) *fǝllan mota* 'such-and-such a village', and oftentimes also in Kurdish, e.g. *filan kes* 'such-and-such a person'. Noun-Numeral order for the numeral 'one', as illustrated in (4–5), as well as the development of a prefixal definite article have presumably been reinforced in Ṭuroyo through contact with Arabic. In Mlaḥso, when only the genitive noun is marked for definiteness, this is presumably based on an Arabic model, cf. (11) and (13) in §2.1.

Semitic and Iranian converge in Noun-Attribute order. Here, the attachment of proclitic determiners to the following adjective in Aramaic (see Waltisberg 2016 for more examples), as shown in (68) and (69), converges not only with the *ezafe* in Northern Kurdish, e.g. (66), but also with the definite article in Arabic dialects, e.g. (67).

(66) Northern Kurdish
      biray=ê   min=ê   mezin
      brother=ᴇᴢ.ᴍsɢ my=ᴇᴢ.ᴍsɢ big
      'my elder brother'

(67) *Qǝltu*-Arabic Kaʿbiye (Jastrow 2022: III:§3)
      axū-y   lǝ-gbīr-∅=ste
      brother.of.ᴍsɢ-my ᴅᴇꜰ-big-ᴍsɢ=ᴀᴅᴅ
      'my elder brother'

(68) CNA Midyat (Ritter 1967: 11:§36)
      ʔaḥun-i   ú=rab-o
      brother.ᴍsɢ-my ᴅᴇꜰ.ᴍsɢ-big-ᴍsɢ
      'my eldest brother'

(69) NENA Upper Barwar (Talay 2009: 516.§2)
      xon-i   ʔó=goṛ-a
      brother.ᴍsɢ-my ᴅᴇᴍ.ᴍsɢ-big-ᴍsɢ
      'my eldest brother'

The effects of convergence with both Arabic and Iranian can be further illustrated by the divergent usage patterns of post-predicate person markers. This pronominal series, often cliticized to the immediately preceding predicate, occurs across Arabic and Aramaic dialects to indicate both the present affirmative copula *and* the pronominal theme-object in ditransitive constructions (e.g. Retsö

Table 6: Comparison of copula placement in *Qəltu*-Arabic and Central Neo-Aramaic

Syrian Arabic Cilician	*Qəltu*-Arabic Mosul	Mardin	
ʕaṭa-ni hinni	*ʕaṭā́-nī=yəm*	*ʕaṭā́-nī=nne*	'he gave them to me'
hinni fəl-bayt	*hīyəm fəl-bēt*	*fəl-báyt=ənne*	'they are at home'
	NENA Qaraqosh	**CNA Ṭuroyo**	
	kewí-li=na	*kobí-li=ne*	'they give them to me'
	ṭawe=na	*ṭáwwe=ne*	'they are good'

1987, Birnstiel 2022). The first is partly modelled on the clause-final copula in the neighbouring Iranian languages. The second, however, suggests close interaction with the Arabic-speaking Levant and Arabia. While the post-predicate copula is a feature common to all languages of the West Asian transition zone (e.g. Haig 2017: 404–405), there are notable differences, such as the pre-predicate placement of negated copulas and relative copulas (see §2.3.1) in the majority of both Aramaic and Arabic dialects. In Syriac, however, the enclitic copula and the bound pronominal objects of the third person plural were also identical in form, cf. *šappirin=ennun* 'they are beautiful' and *qṭal=ennun* 'he killed them', derived from *hennun* 'they'. This copula-object syncretism, compared in Table 6, applies especially to the *Qəltu*-Arabic and Neo-Aramaic dialects of Mardin (Grigore (2007)) and Siirt provinces in Turkey and that of the Mosul Plain (Jastrow 1979; Khan 2002) in Iraq, but is also common in the Levant and Ḥejaz (Retsö 1987). The Mardini dialects in Syria have enclitic pronouns only for the third person in both the copula and the object marking function (Isaksson & Lahdo 2002). They are confined to the object marking function in Mosul *Qəltu*-Arabic (Jastrow 1979). Other *Qəltu*-Arabic dialects do not have such enclitic copulas, notably in Iraq and Syria, although they may be restricted to the negative copula, e.g. *Gələt*-Arabic Khawetna *ma-hi* 'she is not' (Talay 1999: 54–55).

Pre-predicate deictic copulas are a shared feature of Arabic, Aramaic and southeastern Northern Kurdish (Bahdini), compared in Table 7. Deictic copulas characterized by an initial deictic morpheme *k-* occur across *Qəltu*-Arabic dialects, Ṭuroyo and the NENA dialects of the Mosul Plain, denoting a situation in the immediately observable present or the imminent future. The Bahdini future particle *=ê*, derived from *dê* and *wê* – presumably eroded forms of the 3sg. present

Table 7: Comparison of deictic copula and Auxiliary-Verb order

	Bəḥzani *Qəltu*-Arabic (Jastrow 1978: 139)	Qaraqosh NENA (Khan 2002)	Midyat CNA	Bahdini Kurdish
Copula-Predicate 'He is at home.'	*kū fəl-bayt*	*kile b-beθa*	*kəlé bú=bayto*	*ew=ê li malê*
Auxiliary-Verb	*kū ∅-yakəl*	*kile k-axəl*[a]	-1 *kəlé k-oxəl*[b]	*ew=ê bi-xwe*[c]
'He is eating/he is about to eat.' (Arabic/Aramaic).' 'He will eat.' (Kurdish)				

[a] The verbal forms in Neo-Aramaic are in the realis/indicative, lit. 'Here he is, he eats', as opposed to the irrealis/subjunctive in Arabic and Kurdish.

[b] The verbal forms in Neo-Aramaic are in the realis/indicative, lit. 'Here he is, he eats', as opposed to the irrealis/subjunctive in Arabic and Kurdish.

[c] The future particle *=ê* in Bahdini Kurdish does not inflect for number/gender in contradistinction to the *ezafe* copula *=ê* that does inflect for number/gender.

form of 'want' – may also have been influenced by preverbal TAM markers in the same region. At the same time, the morpheme attaches to the subject pronoun and effectively inflects it for TAM similarly to the pre-predicate *ezafe*-based copula and similarly to the pre-predicate deictic copulas in Arabic and Aramaic. For a discussion of the situation in Bahdini Kurdish, see Chyet (1995: 247–249) and Haig (2017: 405–407). Since Neo-Aramaic dialects in general have a plethora of deictic copulas, it is conceivable that these deictic copulas spread from Aramaic into *Qəltu*-Arabic – unless they are a parallel development – and possibly from Semitic even also into Bahdini varieties of Northern Kurdish.

Abbreviations

1	first person	DEIC	deictic
2	second person	DEM	demonstrative
3	third person	DIR	direct case
A	agent	DOM	Differential Object Marking
ADD	additive		
ANT	anterior tense	DRCT	directional
CMPR	comparative	EXIST	existential
COP	copula	EZ	ezafe
CSTR	construct state	F	feminine
DEF	definite article	FUT	future

GEN	genitive	SG	singular
IND	indicative	T	theme
M	masculine	CNA	Central Neo-Aramaic
n	total number of tokens	NENA	Northeastern
NEG	negator		Neo-Aramaic
O	object	V	verb
OBL	oblique case	LOC	locative
PERF	perfect	Lev.	Levantine
PFV	perfective	ABL	ablative
PL	plural	BEN	beneficiary
PP	post-predicate	ADDR	addressee
PST	past	INSTR	Instrumental
R	recipient	WOWA	= Haig et al. (2022)
REL	relative	ZAL	Zeitschrift für Arabische
S	subject (intransitive)		Linguistik
SBJV	subjunctive		

Acknowledgements

References

Abu-Haidar, Farida. 1991. *Christian Arabic of Baghdad*. Wiesbaden: Harrassowitz.

Akkuş, Faruk. 2017. Peripheral Arabic dialects. In Elabbas Benmamoun & Reem Bassiouney (eds.), *The Routledge handbook of Arabic linguistics*. London/New York: Routledge. DOI: 10.4324/9781315147062-26.

Akkuş, Faruk. 2020. Anatolian Arabic. In Christopher Lucas & Stephano Manfredi (eds.), *Arabic and contact-induced change*, 135–158. Berlin: Language Science Press.

Arnold, Werner. 1998. *Die arabischen Dialekte Antiochiens*. Wiesbaden: Harrassowitz.

Arnold, Werner. 2015. Die arabischen Dialekte der Christen in der Türkei. *Mediterranean Language Review* 22. 105–117.

Bar-Moshe, Assaf. 2019. *The Arabic dialect of the Jews of Baghdad*. Wiesbaden: Harrassowitz.

Behnstedt, Peter. 1992. Qəltu-dialekte in Ost-Syrien. *Zeitschrift fur Arabische Linguistik* 24. 35–59.

Birnstiel, Daniel. 2022. Copulas and target phrase positioning in the Arabic dialects of Kurdistan. In Hiwa Asadpour & Thomas Jügel (eds.), *Word order variation*, 197–234. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110790368-010.

Blanc, Haim. 1964. *Communal dialects in Baghdad.* Cambridge, MA: Harvard University Press.

Camilleri, Maris, Shaimaa Elsadek & Louisa Sadler. 2014. A cross dialectal view of the Arabic dative alternation. *Acta Linguistica Hungarica* 61(1). 3–44. DOI: 10.1556/ALing.61.2014.1.1.

Chyet, Michael. 1995. Neo-Aramaic and Kurdish: An interdisciplinary consideration of their influence on each other. *Israel Oriental Studies* 15. 219–252.

Coghill, Eleanor. 2014. Differential object marking in Neo-Aramaic. *Linguistics* 52(2). 335–364. DOI: 10.1515/ling-2013-0065.

Coghill, Eleanor. 2018. Information structure in the Neo-Aramaic dialect of Telkepe. In Evangelia Adamou, Katharina Haude & Martine Vanhove (eds.), *Information structure in Lesser-described languages*, 293–327. Amsterdam: Benjamins.

Coghill, Eleanor. 2020. Neo-Aramaic. In Christopher Lucas & Stephano Manfredi (eds.), *Arabic and contact-induced change*, 371–402. Berlin: Language Science Press.

Cohen, Eran. 2012. *The syntax of Neo-Aramaic: The Jewish dialect of Zakho.* Piscataway, NJ: Gorgias Press. DOI: 10.31826/9781463234737.

Dahlgren, Sven-Olaf. 1998. *Word order in Arabic.* Göteborg: Acta Universitatis Gothoburgensis.

Donabedian, Anaïd & Ioanna Sitaridou. 2021. Anatolia. In Evangelia Adamou & Yaron Matras (eds.), *The Routledge handbook of language contact*, 404–433. Milton Park/Abingdon/Oxon: Routledge.

El Zarka, Dina & Sandra Ziagos. 2020. The beginnings of word order change in the Arabic dialects of Southern Iran in contact with Persian: A preliminary study of data from four villages in Bushehr and Hormozgan. *Iranian Studies* 53(3–4). 465–488. DOI: 10.1080/00210862.2019.1690433.

Fassberg, Steven E. 2010. *The Jewish Neo-Aramaic dialect of Čalla*, vol. 54. Leiden: Brill. DOI: 10.31826/9781463211615-005.

Fink, Andreas. 2020. *Der arabische Dialekt von Hasankeyf am Tigris (Osttürkei)* (Semitica Viva). Wiesbaden: Harrassowitz. DOI: 10.2307/j.ctv11sn58j.

Fox, Samuel Ethan. 2009. *The Neo-Aramaic dialect of Bohtan.* Piscataway, NJ: Gorgias Press. DOI: 10.31826/9781463217327.

Grigore, George. 2007. *L'arabe parlé à Mardin: Monographie d'un parler arabe"périphérique"*. Bucharest: Editura Universităţii din Bucureşti.

Häberl, Charles G. 2011. Neo-Mandaic. In Stefan Weninger, Geoffrey Khan, Michael P. Streck & Janet C. E. Watson (eds.), *The Semitic languages*, 725–737. Berlin: De Gruyter. DOI: 10.1515/9783110251586.725.

Haig, Geoffrey. 2001. Linguistic diffusion in present-day East Anatolia: From top to bottom. In A. Aikhenvald & R. M. W. Dixon (eds.), *Areal diffusion and genetic inheritance: Problems in comparative linguistics*, 195–224. Oxford: Oxford University Press.

Haig, Geoffrey. 2014. East Anatolia as a linguistic area? Conceptual and empirical issues. In Lale Behzadi, Patrick Franke, Geoffrey Haig, Christoph Herzog, Birgitt Hoffmann, Lorenz Korn & Susanne Talabardon (eds.), *Bamberger orientstudien*, vol. 1, 13–35. Bamberg: University of Bamberg Press. file : / / /C : /Users/ba4as01/Downloads/Bost1HerzogopusseA2.pdf.

Haig, Geoffrey. 2015. Verb-goal (VG) word order in Kurdish and Neo-Aramaic: Typological and areal considerations. In Geoffrey Khan & Lidia Napiorkowska (eds.), *Neo-Aramaic and its linguistic context*, 407–425. Piscataway, NJ: Gorgias Press.

Haig, Geoffrey. 2017. Western Asia: East Anatolia as a transition zone. In Raymond Hickey (ed.), *The Cambridge handbook of areal linguistics*, 396–423. Cambridge: Cambridge University Press.

Haig, Geoffrey. 2022. Post-predicate constituents in Kurdish. In Yaron Matras, Geoffrey Haig & Ergin Öpengin (eds.), *Structural and typological variation in the dialects of Kurdish*, 335–377. Cham: Springer International Publishing. DOI: 10.1007/978-3-030-78837-7_8.

Haig, Geoffrey & Geoffrey Khan. 2019. *The languages and linguistics of Western Asia: An areal perspective*. Berlin: De Gruyter Mouton.

Haig, Geoffrey, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils Schiborr. 2024. Post-predicate elements in the Western Asian Transition Zone: Data, theory, and methods. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 3–54. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266331.

Haig, Geoffrey, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr. 2022. *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation*. Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

Heine, Bernd. 2008. Contact-induced word order change without word order change. In Peter Siemund & Noemi Kintana (eds.), *Language contact and contact languages*, 33–60. Amsterdam: Benjamins. DOI: 10.1075/hsm.7.04hei.

Isaksson, Bo & Ablahad Lahdo. 2002. Three border towns between Turkey and Syria: ʿĀmūda, Dərbēssiyye and Rās əl-ʿAyn. In Werner Arnold & Harmut Bobzin (eds.), *„Sprich doch mit deinen Knechten aramäisch, wir verstehen es!" 60 Beiträge zur Semitistik Festschrift für Otto Jastrow zum 60. Geburtstag*, 311–335. Wiesbaden: Harrassowitz.

Jastrow, Otto. 1978. *Die mesopotamisch-arabischen qəltu-Dialekte. Vol. 1: Phonologie und Morphologie*. Wiesbaden: Steiner.

Jastrow, Otto. 1979. Zur arabischen Mundart von Mossul. *ZAL* 2. 76–99.

Jastrow, Otto. 1981. *Die mesopotamisch-arabischen qəltu-Dialekte. Vol. 2: Volkskundliche Texte in elf Dialekten*. Wiesbaden: Steiner.

Jastrow, Otto. 1985. *Laut- und formenlehre des neuaramäischen Dialekts von Midin im Tur ʿAbdin*. Wiesbaden: Harrassowitz.

Jastrow, Otto. 1988. *Der neuaramäische Dialekt von Hertevin (Province Siirt)*. Wiesbaden: Harrassowitz.

Jastrow, Otto. 1990. *Der arabische Dialekt der Juden von ʿAqra und Arbīl*. Wiesbaden: Harrassowitz.

Jastrow, Otto. 1994. *Der neuaramäische Dialekt von Mlaḥsô*. Wiesbaden: Harrassowitz.

Jastrow, Otto. 2003. *Arabische Texte aus Kinderib*. Wiesbaden: Harrassowitz.

Jastrow, Otto. 2004. Uzbekistan Arabic: A language created by Semitic-Iranian-Turkic linguistic convergence. In Éva Ágnes Csató, Bo Isaksson & Carina Jahani (eds.), *Linguistic convergence and areal diffusion: Case studies from Iranian, Semitic and Turkic*, 133–189. London/New York: Routledge. DOI: 10.4324/9780203327715.

Jastrow, Otto. 2006. Arabic dialects in Turkey: Towards a comparative typology. *Türk Dilleri Araştırmaları* 16. 153–164.

Jastrow, Otto. 2007. Where do we stand in the research on the Anatolian qəltu dialects? *Romano-Arabic* 7. 63–69.

Jastrow, Otto. 2022. *Der arabische Dialekt der Christen von Kaʿbīye (Diyarbakir)*. Wiesbaden: Harrassowitz.

Khan, Geoffrey. 2002. *The Neo-Aramaic dialect of Qaraqosh*. Leiden: Brill.

Khan, Geoffrey. 2008a. *The Jewish Neo-Aramaic dialect of Urmi*. Piscataway, NJ: Gorgias Press.

Khan, Geoffrey. 2008b. *The Neo-Aramaic dialect of Barwar*. Leiden: Brill. DOI: 10.1163/ej.9789004167650.i-2198.

Khan, Geoffrey. 2019. The Neo-Aramaic dialects of eastern Anatolia and north-western Iran. In Geoffrey Haig & Geoffrey Khan (eds.), *The languages and linguistics of Western Asia: An areal perspective*, 190–236. Berlin: De Gruyter Mouton.

Khan, Geoffrey. 2020. Contact and change in Neo-Aramaic dialects. In Bridget Drinka (ed.), *Historical linguistics 2017: Selected papers from the 23rd international conference on historical linguistics, San Antonio, Texas, 31 July – 4 August 2017*, 387–407. Amsterdam: Benjamins.

Lahdo, Ablahad. 2009. *The Arabic dialect of Tillo in the region of Siirt (south-eastern Turkey)*. Uppsala: Uppsala University.

Leitner, Bettina. 2024. Khuzestani Arabic. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 413–430. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266357.

Lyavdansky, Alexey K., Nikita Kuzin, Yulia Furman, Sergey Loesov, Maksim Kalinin, Sergey Koval, Eugene Barsky & Y. Kirpianovich. 2020. *Ṭuroyo corpus.* http://neo-aramaic.web-corpora.net/turoyo_corpus/search (5 July, 2022).

Matras, Yaron. 2009. *Language Contact*. Cambridge: Cambridge University Press.

Mohammadirad, Masoud. 2024. Zagros region: The Kurdish-Gorani continuum. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 245–279. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266347.

Molin, Dorota. 2021. *The Jewish Neo-Aramaic dialect of Dohok: A comparative grammar*. Cambridge: University of Cambridge. (Doctoral dissertation).

Molin, Dorota. 2022. NE Neo-Aramaic (Jewish Dohok). In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA — Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation*. Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa (5 July, 2023).

Molin, Dorota. 2024. *The Neo-Aramaic dialect of the Jews of Dohok: A comparative-typological grammar*. Leiden: Brill.

Nakano, Aki'o. 1973. *Conversational texts in Eastern Neo-Aramaic (Gzira dialect)*. Tokyo: Institute for the Study of Languages, Cultures of Asia & Africa.

Noorlander, Paul M. 2014. Diversity in convergence: Kurdish and Aramaic variation entangled. *Kurdish Studies* 2(2). 201–224.

Noorlander, Paul M. 2018. *Alignment in Eastern Neo-Aramaic languages from a typological perspective*. Leiden: Leiden University. (Doctoral dissertation).

Noorlander, Paul M. 2021. *Ergativity and other alignment types in Neo-Aramaic: Investigating morphosyntactic microvariation.* Leiden: Brill.

Noorlander, Paul M. 2022a. Arabic (Christian, Kaʿbiye). In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA – Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

Noorlander, Paul M. 2022b. Arabic (Jewish, Baghdad). In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA – Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

Noorlander, Paul M. 2022c. Central Neo-Aramaic (Mlaḥso). In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA – Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

Noorlander, Paul M. 2022d. Central Neo-Aramaic (Ṭuroyo, Midyat). In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA – Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation.* Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/resources/wowa/.

Noorlander, Paul M. 2023. Object verb orderings in Arabic dialects of Anatolia: A quantitative approach to contact-induced word order change. *Zeitschrift für arabische Linguistik* 77. 32–52. DOI: https://doi.org/10.13173/ZAL.77.2.032.

Noorlander, Paul M. 2024. Neo-Aramaic in Iran and northeastern Iraq. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*, 431–469. Berlin: Language Science Press. DOI: 10.5281/zenodo.14266359.

Noorlander, Paul M. & Dorota Molin. 2022. Word order typology in North Eastern-Neo-Aramaic: Towards a corpus-based approach. In Hiwa Asadpour & Thomas Jügel (eds.), *Word order variation Semitic, Turkic and Indo-European languages in contact*, 235–258. Berlin: De Gruyter Mouton.

Noorlander, Paul M. & Donald Stilo. 2015. On the convergence of verbal systems of Aramaic and its neighbours. Part I: Present-based paradigms. In Geoffrey Khan & Lidia Napiorkowska (eds.), *Neo-Aramaic in its linguistic context*, 426–452. Piscataway, NJ: Gorgias Press. DOI: https://doi.org/10.31826/9781463236489-027.

Procházka, Stephan. 2002. *Die arabischen Dialekte der Çukurova (Südtürkei)*. Wiesbaden: Harrassowitz.

Procházka, Stephan. 2003. The Bedouin Arabic dialects of Urfa. In Ignacio Ferrando & Sánchez Sandoval. Juan José (eds.), *Association internationale de dialectologie Arabe (AIDA) 5th conference proceedings Cádiz, September 2002*, 75–88.

Procházka, Stephan. 2019. The Arabic dialects of Eastern Anatolia. In Geoffrey Haig & Geoffrey Khan (eds.), *The languages and linguistics of Western Asia: An areal perspective*, 159–189. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110421682-005.

Procházka, Stephan. 2020. Arabic in Iraq, Syria, and southern Turkey. In Christopher Lucas & Stefano Manfredi (eds.), *Arabic and contact-induced change*, 83–114. Berlin: Language Science Press.

Ratcliffe, Robert R. 2005. Bukhara Arabic: A metatypized dialect of Arabic in Central Asia. In Éva Csató, Bo Isaksson & Carina Jahani (eds.), *Linguistic convergence and areal diffusion. Case studies from Iranian, Semitic and Turkic*, 141–160. Routledge Curzon. DOI: 10.4324/9780203327715.

Retsö, Jan. 1987. Copula and double pronominal objects in some Semitic languages. *Zeitschrift der Deutschen Morgenländischen Gesellschaft* 137. 219–245.

Ritter, Helmut. 1967. *Ṭūrōyō: Die Volksprache der syrischen Christen des Ṭūr ʿAbdîn A. Texte I–III*. Beirut: Steiner.

Seeger, Ulrich. 2002. Zwei Texte im Dialekt der Araber von Chorasan. In Werner Arnold & Hartmut Bobzin (eds.), *„Sprich doch mit deinen Knechten aramäisch, wir verstehen es!" 60 Beiträge zur Semitistik Festschrift für Otto Jastrow zum 60. Geburtstag*, 629–646. Wiesbaden: Harrassowitz.

Silva-Corvalán, Carmen. 1994. *Language contact and change: Spanish in Los Angeles*. Oxford: Clarendon.

Silva-Corvalán, Carmen. 2008. The limits of convergence in language contact. *Journal of Language Contact* 2(1). 213–224.

Sinha, Jasmin. 2000. *Der neuostaramäische Dialekt von Bēṣpǝn*. Wiesbaden: Harrassowitz.

Souag, Lameen. 2017. Clitic doubling and language contact in Arabic. *Zeitschrift für Arabische Linguistik* (66). 45–70. DOI: 10.13173/zeitarabling.66.0045.

Stilo, Donald. 2022. NE Neo-Aramaic (Christian) Barwar. In Geoffrey Haig, Donald Stilo, Mahîr C. Doğan & Nils N. Schiborr (eds.), *WOWA – Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation*. Bamberg: Bemberg University. multicast.aspra.uni-bamberg.de/resources/wowa/.

Talay, Shabo. 1999. *Der arabische Dialekt der Khawētna. Vol. 1: Grammatik*. Wiesbaden: Harrassowitz.

Talay, Shabo. 2002. Der arabische Dialekt von Hasköy (Dēr Khāṣ), Ostanatolien, II: Texte und Glossar. *ZAL* 41. 46–86.

Talay, Shabo. 2007. The influence of Turkish, Kurdish and other neighbouring languages on Anatolian Arabic. *Romano-Arabica* 7. 179–189.

Talay, Shabo. 2008. *Die neuaramäischen Dialekte der Khabur-Assyrer in Nordostsyrien: Einführung, Phonologie und Morphologie*. Wiesbaden: Harrassowitz.

Talay, Shabo. 2009. *Neuaramäische Texte in den Dialekten der Khabur-Assyrer in Nordostsyrien*. Wiesbaden: Harrassowitz.

Talay, Shabo. 2012. Arabic dialects of Mesopotamia. In Stefan Weninger, Geoffrey Khan, Michael P. Streck & Janet C. E. Watson (eds.), *The Semitic languages: An international handbook*, 909–920. Berlin: De Gruyter. DOI: 10.1515/9783110251586.909.

Talay, Shabo. 2014. The Mesopotamian-Levantine dialect continuum. In Tal Davidovich, Ablahad Lahdo & Torkel Lindquist (eds.), *From Tur Abdin to Hadramawt: Semitic studies festschrift in honour of Bo Isaksson on the occasion of his retirement*, 179–188. Wiesbaden: Harrassowitz. DOI: 10.2307/j.ctvc2rm3d.17.

Tsereteli, Konstantin. 1963. Образцы современной ассирийской речи (ванские тексты). In G. Sh. Sharbatov (ed.), *Семитские языки: сборник статей I*, 202–218. Moskva: Nauka.

Versteegh, Kees. 1984. Word order in Uzbekistan Arabic and universal grammar. *Orientalia Suecania* 33. 443–453.

Waltisberg, Michael. 2016. *Syntax des Ṭuroyo*. Wiesbaden: Harrassowitz. DOI: 10.2307/j.ctv11qdtvj.

Weninger, Stefan. 2012. Aramaic-Arabic language contact. In Stefan Weninger, Geoffrey Khan, Michael P. Streck & Janet C. E. Watson (eds.), *The Semitic languages: An international handbook*, 747–755. Berlin: De Gruyter. DOI: 10.1515/9783110251586.747.

Wittrich, Michaela. 2001. *Der arabische Dialekt von Āzəx*. Wiesbaden: Harrassowitz.

Name index

Language index

Subject index

Post-predicate elements in the Western Asian Transition Zone

This volume explores word-order phenomena across a phylogenetically diverse sample of languages covering a region loosely referred to as the Western Asian Transition Zone, approximately corresponding to western Iran, northern Iraq, eastern Turkey and the Caucasus. The sample includes representatives from four branches of Indo-European (Iranian, Hellenic, Armenian, Indo-Aryan) as well as Turkic, Semitic, Kartvelian, Northwest Caucasian and Northeast Caucasian. Methodologically, we apply a corpus-based approach to word-order, building on two purpose-built and fully accessible data-bases of spoken language corpora, WOWA (Word Order in Western Asia), and HamBam (Hamedan-Bamberg Corpus of Contemporary Spoken Persian). The majority of the languages are historically OV, yet exhibit high rates of post-verbal elements, and these constitute the primary focus of the volume. One of the major findings is the importance of semantic role in determining pre- versus post-verbal placement of clausal constituents: We identify a consistent bias towards post-verbal placement of spatial Goals, which is amplified by increasing areal proximity to the VO languages of the southwestern periphery of the region (Semitic). In the languages in and adjacent to the Caucasus, on the other hand, we find stronger effects of information structure in triggering post-verbal position. Along with contributions on individual languages and varieties, the volume includes an overview chapter outlining the theoretical background and the data sources, summary chapters on sub-regions, as well as contributions from an experimental and psycholinguistic perspective.