Chapter 1

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 **??**) 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

		V-Goal		V-non-Goal		V-X	
Genre	Clause	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 ??), 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 preverbal (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.PST.3SG mehrdad RA
'When the car hit Mehrdad.'

(29) Colloquial New Persian (Izadi 2022: Q, 1919)
faqat did-am ***yek daste mu***
Just see.PST-1SG 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=EZ talk-PL=3SG return.PST.3SG from=1SG ask.PST.3SG name-PL RA
'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.PST.3SG I RA
'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.PST=3SG man=DEF
'The man took it.'

(33) Colloquial New Persian (Izadi 2022: F, 0754)
hiči um-ad ***doktor=e***
anyway come.PRS-3SG doctor=DEF
'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 ??).

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 $e^{\beta} = 2.52$, 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 $e^{\beta} = 0.05$). 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 ($e^{\beta} = 8.12$ times higher odds of post-verbal positioning for private register, with $p < 0.05$) and flagging for other obliques ($e^{\beta} = 0.02$ 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.

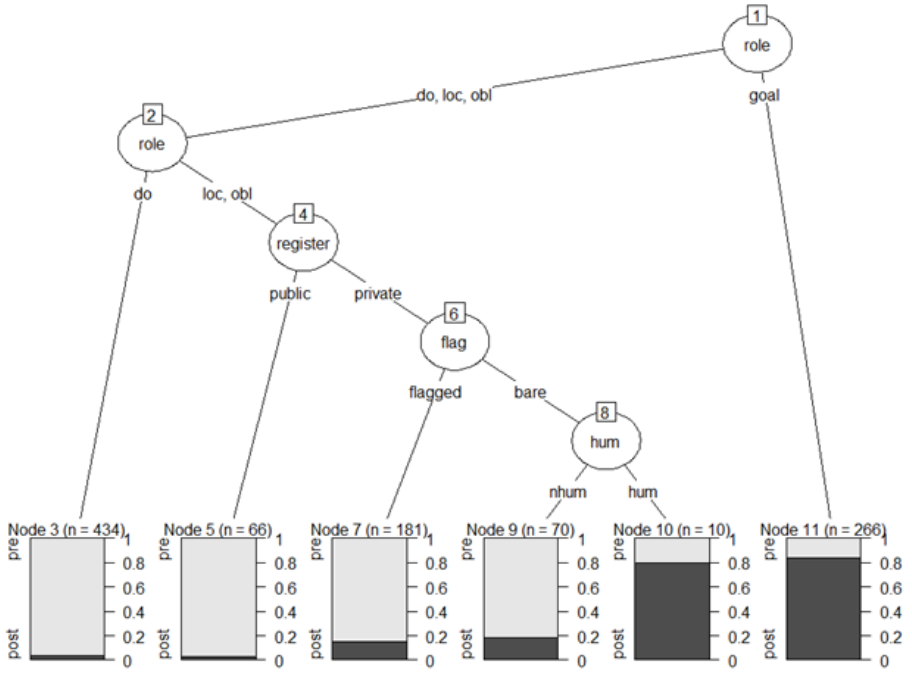

Figure 1: Binary classification tree.

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.

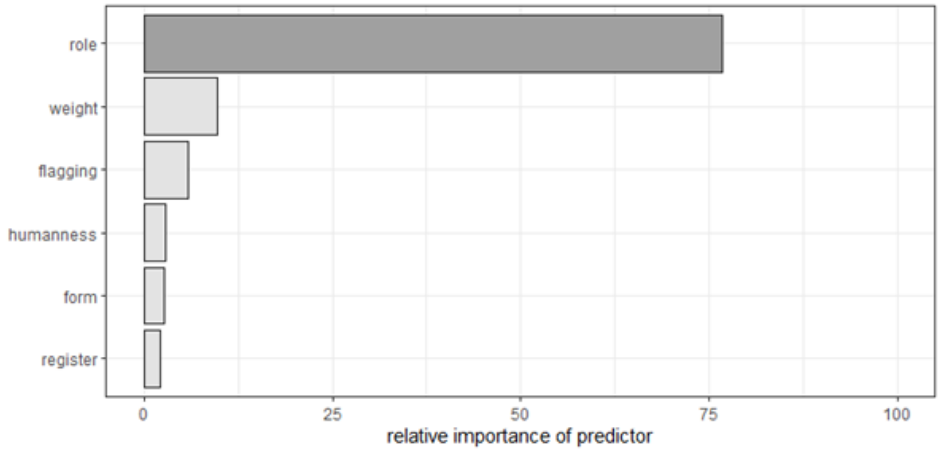

Figure 2: Relative importance of the predictors in the gradient boosting model.

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.

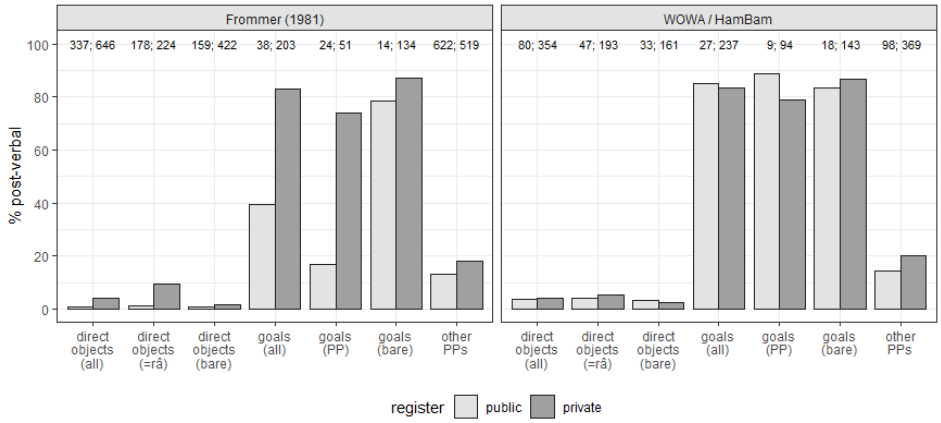

Figure 3: Register differences per role in spoken Persian 1981 and the 2020's respectively

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-economic 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 constituant
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, 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, Rasekh-Mahend, Stilo, Schreiber & 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*, 3–53. Berlin: Language Science Press. DOI: ??.

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*, 243–277. Berlin: Language Science Press. DOI: ??.

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*, 119–152. Berlin: Language Science Press. DOI: ??.

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*, 229–242. Berlin: Language Science Press. DOI: ??.

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. Pleşca, 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*, 87–116. Berlin: Language Science Press. DOI: ??.

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.