Modal syntax cuts short the claim that modern Persian lacks apocopated infinitives

Setayesh Dashti
University of Oxford

Ash Asudeh
University of Rochester

Proceedings of the LFG'23 Conference

Miriam Butt, Jamie Y. Findlay and Ida Toivonen (Editors)

2023

PubliKon

lfg-proceedings.org

Abstract

Persian is usually assumed to lack a nonfinite clause, as its morphosyntactically distinguished infinitival form is a nominal infinitive. This paper takes a closer look at Persian modal verbs, which, in their impersonal use, take a seemingly past stem or third person singular verb as their complement. We show that the modal syntax of Persian reveals the existence of another type of infinitive in Persian which existed in earlier stages of the language as well. This infinitive, which has been traditionally called the apocopated infinitive, is clausal in nature. We propose a Lexical-Functional Grammar (LFG) analysis for modal syntax in Persian, and show how our analysis captures marginal agreement patterns and a related raising phenomenon.

1 Background

Persian is an SOV Indo-European language with '*pro*-drop'.[1] In Persian, verbal morphology is traditionally based on a distinction between two stems. These stems are usually called the *present stem* and the *past stem*. The present stem bears no overt present tense marker, whereas the past stem is regularly marked with *-d* and its allomorphs, modulo suppletive patterns (Anoushe 2018). For instance, the verbal stems of the verb *xordan* 'to eat' are as follows:

(1) Present stem: *xor* 'eat'

(2) Past stem: *xord* 'eat'

Following Anoushe (2018), we break from the traditional/descriptive tradition in treating the *-d* in *xord*, etc., as the actual marker of past tense rather than as part of the stem per se. Thus, the past 'stem' is actually the present stem plus this suffix.[2] The present stem always occurs with either aspectual or mood markers; *mi-* for imperfective aspect (3a) and *be-* for subjunctive mood (3b). The unprefixed past stem with agreement suffixes is used to show the perfective aspect (3c).

Past imperfective, progressive and perfect are also derived from the past stem with agreement suffixes; for example, past imperfective is formed with the same prefix as present imperfective, *mi-* (3d).[3,4]

[1] The dialect reported on here is colloquial spoken Persian, not the written standard.

[2] There are various analytical options for how to capture this, but one option is to treat the present stem as contributing the f-description $\{(\uparrow \text{TENSE}) \mid (\uparrow \text{TENSE}) = \text{PRESENT}\}$, which treats present tense as the default/unmarked value, but allows another tense, e.g. PAST, to be specified instead, if it is contributed by something else, such as the proposed suffix.

[3] Glosses are abbreviated as follows: AUX–auxiliary, COP–copula, DO–direct object, EZ–*ezafe* (nominal linker), IPFV–imperfect, INF–infinitive, NEG–negation, PP–past participle, PRES–present tense, PAST–past tense, SBJV–subjunctive mood, SG–singular, PL–plural, □–necessity modal, ◇–possiblity modal. We use the hyphen (-) to indicate an affix boundary and an equals sign (=) to indicate clitic attachment. What we have glossed as IPFV–imperfective is sometimes glossed as DUR–durative. We prefer to gloss it based on its morphological form rather than its typical morphosyntactic function.

[4] Past imperfective also functions as a 'fake past' to convey counterfactuality, regardless of tense or aspect (Bjorkman and Halpert 2017).

(3) a. Nika be madrese mi-rav-ad.
Nika to school IPFV-go.PRES-3SG
'Nika goes to school.'

b. Nika šāyad be madrese be-rav-ad.
Nika may to school SBJV-go.PRES-3SG
'Nika might go to school.'

c. Nika be madrese raf-t.
Nika to school go-PAST.3SG
'Nika went to school.'

d. bače-hā har ruz be madrese
child-PL every day to school
mi-raf-t-and.
IPFV-go-PAST-3PL
'The kids used to go to school every day.'

Persian has several adverbial and complex predicate modals, but there are only two main simplex modal verbs, *bāyestan* (necessity/□) and *šodan* (possibility/◇).[5] These modals always appear in the default third person singular form: *bāyad* (□.PRES)/*bāyest* (□.PAST) and *mi-še* (IPFV-◇.PRES)/*mi-šod* (IPFV-◇.PAST).[6] They can occur with either:

1. a finite complement (4), marked with subjunctive mood in present tense (4a) or imperfective aspect in past tense (4b); or

2. a complement in which the verb has a simple past stem, which *resembles* the third person singular past inflected form, but is historically an *apocopated infinitive* (short infinitive); importantly, it is interpreted as an impersonal (5).

Note that in (5), the gloss for the verb in the complement is left unspecified, since it could be either a past form or a short infinitive.

(4) a. bāyad be xune be-rav-am.
□.PRES to home SBJV-go.PRES-1SG
'I have to go home.'

b. bāyad bačehā be xune mi-raf-t-and.
□.PRES child-PL to home IPFV-go-PAST-3PL
'The children had to go home.'

[5]There is some debate over the status of *šāyestan*. Some literature, such as Karimi (2005) and Taleghani (2008), treats it as another modal auxiliary, while other literature, such as Labbafankhosh and Darzi (2015), treats it is a modal adverb.

[6]In this paper, we focus on the former modal verb. The only difference between *bāyestan* and *šodan* is that the latter has semantic contribution to the tense while the former does not; this distinction does not affect our discussion. We will mention the lack of tense contribution of *bāyestan* later on.

(5) bāyad zood be xune raf-t.
□.PRES early to home go-??
'It's necessary to go home early.'/
'One must go home early.'

The form we see in the complement of the modal here also appears as the lexical component of the periphrastic future construction:

(6) Ali farda be madrese xāh-ad raf-t.
Ali tomorrow to school want-3SG go-??
'Ali will go to school tomorrow.'

When the modal occurs with a finite complement, as in (4b), it is possible to topicalize the embedded subject to the left:

(7) bačehā bāyad be xune mi-raf-t-and.
child-PL □.PRES to home IPFV-go-PAST-3PL
∼'As for the children, they had to go home.'

A verb that works very similarly to the simplex modal verbs is the complex *be nazar āmad-an* (lit. 'to opinion come-INF'), which is the equivalent of English *seem*. Unlike *bāyad* (□/necessity), which never inflects for agreement, some speakers allow both the non-agreeing/default form (8b) and the agreeing form (8c);[7] all speakers allow the first, non-agreeing form:

(8) a. be nazar mi-ā-d ke bače-hā xaste šo-d-an.
to opinion IPFV-come.PRES-3SG COMP child-PL tired become-PAST-3PL
'It seems that the children have gotten tired.'

b. bače-hā be nazar mi-ā-d ke xaste šo-d-an.
child-**PL** to opinion IPFV-come.PRES-**3SG** COMP tired become-PAST-3PL
'As for the children, it seems that they have gotten tired.'

c. % bače-hā be nazar mi-ā-n ke xaste šo-d-an.
child-**PL** to opinion IPFV-come.PRES-**3PL** COMP tired become-PAST-3PL
'The children seem to have gotten tired.'

Note that in (8) we are emphasizing the colloquialness of the reported example by using certain spoken-only forms, such as the contracted 3PL in (8c): *-(a)n* rather than *-(a)nd*.

2 Puzzles/questions

Based on the background above, we can identify the following puzzles or questions to be addressed:

[7] Storoshenko et al. (2020) have demonstrated experimentally that sentence (8c) is possible for some speakers.

1. How should we account for the complement in (5)?

   (5) bāyad zood be xune raf-t.
   □.PRES early to home go-??
   'It's necessary to go home early.'/ 'One must go home early.'

   Is it a past tense form or a short infinitive (synchronically as well as diachronically)?

2. How can we capture the personal readings of modal constructions, as in (4), versus their impersonal readings, as in (5)?

3. What is the syntactic structure of simplex modal constructions?

4. How should the variable agreement displayed in (8) be explained?

(8) b. bače-hā be nazar mi-ā-d ke
child-PL to opinion IPFV-come.PRES-3SG COMP
xaste šo-d-an.
tired become-PAST-3PL
'As for the children, it seems that they have gotten tired.'

c. % bače-hā be nazar mi-ā-n
child-PL to opinion IPFV-come.PRES-3PL
ke xaste šo-d-an.
COMP tired become-PAST-3PL
'The children seem to have gotten tired.'

In the next section, we will address the first two questions. After giving an overview of the previous literature on Persian modal syntax, we focus on what we can conclude about the nature of the complements of modals based on their syntactic behaviour. Specifically, we show that *bāyad* in Persian is a finite modal verb and that its complement is a nonfinite form, as it cannot be the target of aspectual constraints imposed by the modal. We also discuss the differences between this infinitival form and the nominal *-an* infinitives. Afterwards, we provide an analysis of modal syntax in Persian within Lexical-Functional Grammar that can capture our empirical observations. We then show how our treatment of modal syntax can explain the variation in agreement.

3 The syntax of Persian modals

The syntax of *bāyad* has been extensively discussed by Karimi (2008) within Minimalism. Karimi takes *bāyad*-clauses to be instances of what are often called *impersonal constructions*, a type of arbitrary control.[8] Karimi describes the main verbal

[8] Arbitrary control in Persian also includes constructions with non-apocopated *-an* infinitives.

form in these impersonal constructions (which include the constructions with the two modal verbs *bāyad* and *mi-še*)[9] as a "past stem with no overt inflection, representing the 3rd person singular for the past tense." In order to investigate the nature of the subject in these control constructions, Karimi first provides the following arguments that Persian impersonal constructions can have arbitrary readings:

1. Impersonal constructions are incompatible with a referential antecedent for their subjects; for instance, an impersonal construction cannot be the complement of an obligatory control predicate.

   (9) * say kar-d-am ke bāyad raf-t
   try do-PST-1SG COMP □.PRES go-??

2. Impersonal constructions are incompatible with weather constructions.

   (10) a. bārun umad
   rain come.PST.3SG
   'It rained'

   b. * bārun bāyad umad
   rain □.PRES come.PST.3SG

3. Impersonal constructions are incompatible with emphatic pronominal subjects.

   (11) * bāyad xod in kār ro kard
   □.PRES self DEM task DO do-??
   Intended: 'One has to do this task oneself.'

Karimi also shows that impersonal constructions are incompatible with overt subjects, which cannot be explained by assuming that the clause is not tensed (see example (21) and its discussion):

(12) * un bāyad gof-t ke ...
s/he □.PRES say-?? COMP ...

Karimi argues against the traditional account of control, which has taken the subject in arbitrary control to be PRO, at least since Chomsky and Lasnik (1977). Karimi's reason for rejecting this account is that PRO, depending on the exact framework, is either caseless or receives null Case. However, empirical evidence shows otherwise: in some languages it can appear in nominative Case positions, and in some languages an overt pronoun can appear instead of PRO (for discussion and references, see, e.g., Asudeh 2005). In Persian, PRO and lexical DPs can both appear in the subject position of subjunctive verbs (Ghomeshi 2001, Karimi 2005).

[9] Karimi assumes *bāyad* is a modal adverb and that *mi-še* is a 'semi-auxiliary.'

To account for the these properties of impersonal constructions, Karimi gives an account of these data (and other control constructions in Persian) in terms of feature agreement, following Manzini and Roussou (2000). In this kind of analysis, the DP in control constructions is base-generated. In Manzini and Roussou's analysis, the DP attracts the features of both matrix and control predicates. In Karimi's analysis, the matrix predicate and matrix C head have feature agreement. This agreement satisfies the theta-features of the predicate, and determines the semantic type and interpretation of the clause/phrase.[10]

In this paper, we show that a closer look at the syntax of these constructions motivates an alternative account of the complement of *bāyad*, one that takes it to be an infinitival form. Once we establish that the complement in these constructions is a nonfinite clause, the data that Karimi tries to account for could be captured within various accounts of control, including the traditional account of Chomsky and Lasnik (1977).

The main challenge is the construction with an impersonal interpretation, as in (5), shown here with the complement structure made explicit, or the similar example (13):

(5′) bāyad [zood be xune raf-t].
□.PRES [early to home go-??]
'It's necessary to go home early.'/ 'One must go home early.'

(13) bāyad [šab-hā hašt sā'at xāb-id].
□.PRES night-PL eight hour sleep-??
'It's necessary to sleep for eight hours a night.'/'One must sleep for eight hours a night.'

Let us assume that the complement is a clause, because there is apparent inflection on the embedded predicate. The question is: what kind of clause? Is it finite, as indicated by its shared form with the past tense (see, e.g., Karimi 2008)? Or is it infinitival, despite the shared form (Karimi 2005, Samvelian 2018)? Or perhaps it is not a clause at all, but presumably some kind of nominal? In order to answer these questions, we turn to a brief consideration of infinitives more generally in the language.

3.1 Impersonal constructions are infinitivals

Previous analyses of Persian modal syntax have struggled to explain how the apparently past tense verb in the complement should be analyzed, as it does not contribute any temporal reference. In this section, we will show that there are compelling arguments to treat the complement of the modal verb as an infinitive, not a past tense verb. First, note that the past finite complement of the modal should

[10] In arbitrary control, a generic operator in C gives rise to an impersonal generic interpretation.

bear imperfective marking, but adding this marking to the sort of complement under discussion renders an impersonal reading unavailable and requires it to have a personal reading (thanks to the availability of *pro*-drop in Persian). Contrast (14a), repeated from (13) above, with (14b):

(14) a. bāyad šab-hā hašt sā'at xāb-id.
□.PRES night-PL eight hour sleep-??
'It's necessary to sleep for eight hours a night.'/'One must sleep for eight hours a night.'

b. bāyad šab-hā hašt sā'at mi-xāb-id.
□.PRES night-PL eight hour IPFV-sleep-PAST.3SG
'It's necessary to sleep for eight hours a night.'/'One must sleep for eight hours a night.'
✓ '*pro*.3SG had to sleep for eight hours a night.'

Persian is sometimes assumed to lack a nonfinite clause (Karimi 2008, Darzi and Kwak 2015), exactly because of the similarity in morphological form between the third singular past form, which is unmarked for agreement morphology (e.g., *raf-t* go-PAST.3SG) and the simple stem form in question (e.g., *raf-t* go-??). But, this does not account correctly for the impersonal readings.

In earlier stages of the language, Persian used to have a short/apocopated infinitive, homophonous with the past stem.[11] This infinitival form appeared as the complement of modal verbs (among other things). Moreover, in some linguistics literature it has been claimed or at least implied, in one way or another, that Modern Persian does have short/apocopated infinitives; or, at least, that the morpheme in question is not in fact the past stem with third person singular agreement. Karimi (2005) uses the term *curtailed infinitives* for the verbal forms under discussion, since this 3rd person singular for the past tense form only differs from the long infinitive in not taking the infinitival marker *-an*. But in Karimi's later work (2008), short/curtailed infinitives are instead taken to be "past stem[s] with no overt inflection, representing the 3rd person singular for the past tense." In the only dedicated study of infinitivals in Persian, Samvelian and Mir-Samii (2007) mention the existence of apocopated infinitives, but exclude them from their study. Also, Samvelian (2018) uses the gloss SINF ("short infinitive") for both

1. the complement of *bāyad*; and

[11] The history and diachrony of the apocopated infinitive is somewhat unclear. This form can be found in Classical Persian, Middle Persian (the direct ancestor of Modern Persian), and Parthian (Skjaervø 2009). However, the Middle Persian case is not as straightforward: early Middle Persian texts such as Sassanian inscriptions and Manichaean texts do not contain the apocopated infinitives, and the texts that do have this form are likely to be influenced by Middle Persian (Lenepveu-Hotz 2012). They are abundant in Parthian, but Persian is not directly related to or even regionally close to this Middle Iranian language. There is speculation about how this form came to be an infinitival form: some scholars hypothesize that is the same form as the past stem, others take it as the continuation of an absolutive nominal form in Old Iranian languages. See Lenepveu-Hotz (2012: pp. 117-118) for discussion and references.

2. the lexical component of the future construction.

In Mirrazi (2022), the *-d* marker of past stems is taken as ambiguous between past and perfective. The form in question is assumed to be the perfective form of the verb without a deictic temporal specification, which gives rise to what Mirrazi calls the subjunctive function of the perfective form, namely its use in future and modal constructions. Therefore, according to Mirrazi, the complement of the *bāyad*-clauses is the perfective form of the verb with default third person agreement.

As the form in question does not pattern with finite verbs, we suggest that this form is still (synchronically, as well as diachronically) the apocopated infinitive and is thus unmarked for TENSE/ASPECT/MOOD. The future construction, shown in (15), repeated from (6), provides further evidence for nonfiniteness of this verbal form (now glossed INF), since (15) does not have a future-past reading (∼'Ali will have gone to school tomorrow').

(15) Ali farda be madrese xāh-ad raf-t.
Ali tomorrow to school want-3SG go-INF
'Ali will go to school tomorrow.'

Additionally, Lowe (2019) has observed that nonfinite forms are prototypically involved in periphrastic constructions as the lexical content of the clausal predicate. In the analysis section, §4, a template (Dalrymple et al. 2004, Asudeh et al. 2013) is used to define this defective/infinitival verbal form.

3.2 Infinitivals and nonfinite clauses in Persian

As mentioned above, Persian has a morphosyntactically distinguished infinitival form, which is formed from the past stem and the suffix *-an*. This long infinitive is nominal, which means that it can freely occur in subject and object position:

(16) a. dav-id-an barāye salāmati mofid-e
run-PAST-INF for health beneficial-COP.3SG
'Running has health benefits.'

b. Nika dav-id-an=ro dust dār-e
Nika run-PAST-INF=DO friend have.PRES.3SG
'Nika likes running.'

On the other hand, the apocopated infinitives that occur in the complements of modal verbs cannot appear in subject or object position:

(17) a. * dav-id barāye salāmati mofid-e
run-INF for health beneficial-COP.3SG

b. * Nika dav-id=ro dust dār-e
Nika run-INF=DO friend have.PRES.3SG

Furthermore, long infinitives can take modifiers with the nominal linker *-e* (*ezafe*).

(18) a. xand-id-an-e nowzād ārāmešbaxš-e
laugh-PAST-INF-EZ baby calming-COP.3SG
'Babies' laughing is calming.'

b. Ali dav-id-an-e ārum=ro be šenā kar-d-an-e sari'
Ali run-PAST-INF-EZ slow=DO to swim do-PAST-INF-EZ fast
tarjih mi-d-e
preference IMPF-give.PRES-3SG
'Ali prefers running slowly to swimming fast.'

Note that the infinitive in (18a) has a genitive modifier, and the infinitive in (18b) has an adjectival one.

Also, unbounded dependencies cannot be formed into long infinitives, which is to be expected if long infinitives are nominals:

(19) a. Sarina tanhā be bače šir dā-d-an=ro dust
Sarina alone to baby milk give-PAST-INF=DO friend
na-dār-e
NEG-have.PRES-3SG
'Sarina doesn't like to breast-feed the child alone.'

b. * tanhā Sarina be bače šir dā-d-an=ro dust
alone Sarina to baby milk give-PAST-INF=DO friend
na-dār-e
NEG-have.PRES-3SG
'Sarina doesn't like to breast-feed the child alone.'

c. * be bače Sarina tanhā šir dā-d-an=ro dust
to baby Sarina alone milk give-PAST-INF=DO friend
na-dār-e
NEG-have.PRES-3SG
'Sarina doesn't like to breast-feed the child alone.'

If long infinitives are nominals, it follows that extraction creates a complex NP island violation (Ross 1967). In other words, the ungrammaticality of (19a–c) follows from the treatment of islands more generally (see, e.g., Dalrymple et al. 2019: 656–661 or Kaplan and Zaenen 2023). On the other hand, unbounded dependencies can be formed into the impersonal complement marked by the short infinitive. This is shown in (20):

(20) xoreš=ro bāyad ru šole-ye kam pox-t.
stew-DO □.PRES on flame-EZ low cook-INF
∼'As for the stew, it has to be cooked on low heat.'

These observations show that the long infinitive is a nominal, but that the apocopated infinitive in impersonal constructions is not. This already suggests that the impersonal complement is a nonfinite clause (contra Karimi 2008, Darzi and Kwak 2015), as this is the obvious alternative to its being a nominal. The assumption that the form in the impersonal is a short infinitive that heads a clause explains why it does not take the *-an* suffix: this suffix is a nominalizer and the short infinitive is not a nominal. The fact that this looks superficially like the past form is a fact of morphosyntactic syncretism and nothing more, because the short infinitive does not *function* as a past form or have past meaning.

In sum, we have presented syntactic arguments for why the long infinitive is a nominal and for why the form in the impersonal construction is not. This already suggests that the impersonal complement is a nonfinite clause, as this is the obvious alternative to its being a nominal.

3.3 A possible alternative analysis

There is a possible analysis which assumes that *bāyad* is simply a modal adverb and therefore does not head its own clause or take a clausal complement. However, there is strong evidence against this analysis. In order to show this, let us first see that the *bāyad*-clauses are themselves finite. The most compelling case for the finiteness of *bāyad*-clauses comes from their embedding under the complementizer *ke*:

(21) a. mā motaqed hast-im ke bāyad tavarrom rō kam kar-d
we believer COP.PRES-1PL COMP □.PRES inflation DO little do-INF
'We believe that one must reduce inflation.'

b. mošāver-hā gof-t-and ke bāyad haghighat rō gof-t
consultant-PL say-PST-3PL COMP □.PRES truth DO say-INF
'The consultants said that one must tell the truth.'

Overt complementizers are known to select for a finite clause. Now that we have shown that these clauses are finite, we need to also identify the source of finiteness. Finiteness here might come from either the modal *bāyad* or the main verb. But we have just argued that the non-modal verbs in examples such as these are short infinitives, so it cannot be the source of finiteness.

However, there is also independent evidence that shows that *bāyad* has verbal properties. Most importantly, note that when *bāyad* takes a finite complement, it requires that its complement should be in the subjunctive, as in (4a) above, repeated here as (22).

(22) bāyad be xune be-rav-am.
□.PRES to home SBJV-go.PRES-1SG
'I have to go home.'

No modal adverb in Persian puts such constraints on the predicate. In (23a–c), we see three modal adverbs with respectively past perfect, present imperfective, and

present progressive predicates (these three adverbs can also come with subjunctive mood in addition to the forms below):

(23) a. Maryam šāyad/ehtemalan/ejbāran rāz-e Zahra ro
Maryam maybe/probably/compulsorily secret-EZ Zahra DO
fahmide ast
understand.PPL COP.PRES.3SG
'Maryam has maybe/probably/compulsorily found out about Zahra's secret.'

b. Maryam šāyad/ehtemalan/ejbāran rāz-e Zahra ro
Maryam maybe/probably/compulsorily secret-EZ Zahra DO
mi-dun-e
IPFV-know.PRES-3SG
'Maryam maybe/probably/compulsorily knows about Zahra's secret.'

c. Maryam šāyad/ehtemalan/ejbāran dār-e be London
Maryam maybe/probably/compulsorily have.PRES-3SG to London
ādat mi-kon-e
habit IPFV-do.PRES-3SG
'Maryam is maybe/probably/compulsorily getting used to London.'

Furthermore, as mentioned by Labbafankhosh and Darzi (2015), *bāyad* can be negated, unlike modal adverbs.[12]

(24) a. na-bāyad az šekast tars-id
NEG-□.PRES of failure fear-??
'One should not fear failure.'

b. bače-hā na-bāyad šab dir be-xāb-and
children NEG-□.PRES night late SBJV-sleep.PRES.3PL
'One should not fear failure.'

Another piece of evidence is that although *bāyad* is defective in its inflectional paradigm (appearing only in the impersonal 3rd person singular), its past counterpart *bāyest* is still in use.[13] It is, however, worth mentioning that neither *bāyad* nor *bāyest* denote tense or any temporal reference and are not sensitive to the tense of the main predicate:[14]

[12]Taleghani (2008) has put forward other arguments for taking *bāyad* as a modal verb, but these are mostly rejected in Labbafankhosh and Darzi (2015).

[13]Compare with *šāyad* ('maybe'), which used to be a verb in earlier stages of the language. Since *šāyad* is now fully grammaticalized as a modal adverb in Persian (Labbafankhosh and Darzi 2015), its past form *šāyest* is no longer in use.

[14]There is, indeed, some degree of grammaticalization in *bāyad* itself, in that it rarely occurs with the imperfective marker *mi-*. This contrasts with the usual pattern of present verbs in Persian, which can never occur without either the imperfective marker *mi-* or the subjunctive marker *be-* (the copula and the verb *dāštan* 'to have' are exceptions).

(25) a. Alireza bāyad dirooz emtehān mi-dā-d
Alireza □.PRES yesterday exam IPFV.give.PST-3SG
'Alireza had to take an exam yesterday.'

b. Alireza bāyest fardā emtehān be-dah-ad
Alireza □.PST tomorrow exam IPFV.give.PRES-3SG
'Alireza has to take an exam tomorrow.'

In sum, there is compelling evidence that *bāyad* is a finite verbal head and is therefore the source of finiteness in *bāyad*-clauses. It is not an adverbial.

3.4 Interim summary

The first question posed in §2 above was *whether the morphologically ambiguous form in the impersonal modal construction is a past tense form or a short infinitive (apocopated infinitive)*. We argued that the form in question, both as the complement of modal verbs and as the lexical part of the future construction as in (6)/(15), is an infinitival form *synchronically*, and that the apocopated infinitive is morphologically formed by referral to the past stem, which explains their identity. However, the agreeing past tense form and the non-agreeing short infinitive have different functions and interpretations.

One way to capture the identity of morphological form is through *rules of referral* in Paradigm Function Morphology (Stump 2001, 2016) or their equivalents in other frameworks. For example, in L_RFG (Melchin et al. 2020, Asudeh and Siddiqi 2022, among others; lrfg.online), the required statement could look like this:

(26) $\langle$[v],@AP-INF(__)$\rangle \xrightarrow{\nu} \nu\langle$[v],@PAST$\rangle$

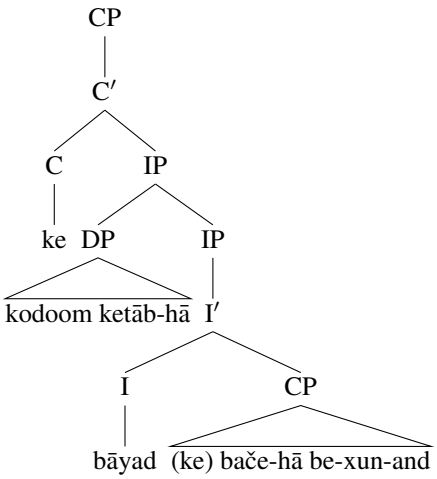
The statement in (26) is intended as a *rule schema* or *meta-rule*. Its lefthand side is underspecified for the parameter of the AP-INF template and can thus match any instance of it in the Vocubulary. The schema states that the exponent of this vocabulary item is the exponent of the PAST template, which controls contribution of [TENSE PAST] to the f-structure. As mentioned above, the past tense exponent is *-d* and its allomorphs (Anoushe 2018). Thus, the schema in (26) elegantly captures the fact that the apocopated infinitive form of 'eat' is *xord*, as in *Bayad xord* ('One must eat.'), that the one of 'go' is *raft*, as in (b) above, and that the one of 'sleep' is *xābid*, as in (14a) above. But it does so without ever referring to any particular form. The details of such an approach remain to be worked out, though.

4 An LFG analysis of Persian modal syntax

The following examples illustrate the c-structure position of the modal and the general structure of the CP and IP in Persian:[15]

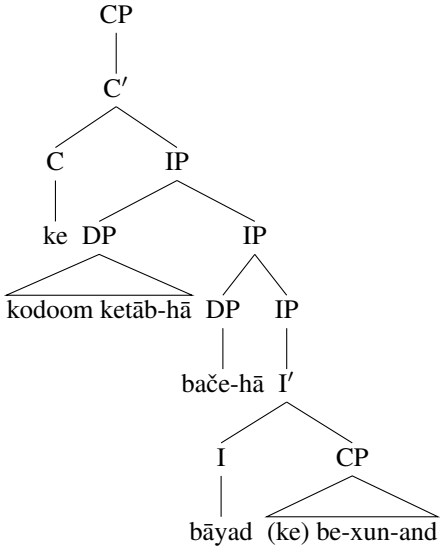
[15]For related LFG work on Iranian verbal paradigms, see Bano et al. (2019) and for LFG work on modals in the closely related language Hindi–Urdu, see Bhatt et al. (2011).

(27) a. Mariam goft
Mariam said
[$_{CP}$ [$_{C'}$ [$_{C}$ ke] [$_{IP}$ kodoom ketāb-hā=ro [$_{IP}$ [$_{I'}$ [$_{I}$ bāyad] [$_{CP}$ (ke)
COMP which book-PL=DO must (COMP)
bače-hā be-xun-and]]]]]]
child-PL SBJV-read-3PL
'Mariam said that the children must read WHICH BOOKS?'

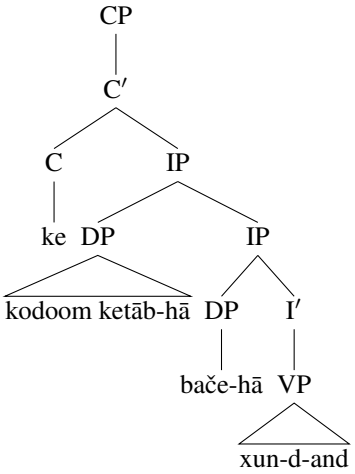


b. Mariam goft
[$_{CP}$ [$_{C'}$ [$_{C}$ ke] [$_{IP}$ kodoom ketāb-hā=ro [$_{IP}$ bače-hā [$_{IP}$ [$_{I'}$ [$_{I}$ bāyad]
COMP which book-PL=DO child-PL must

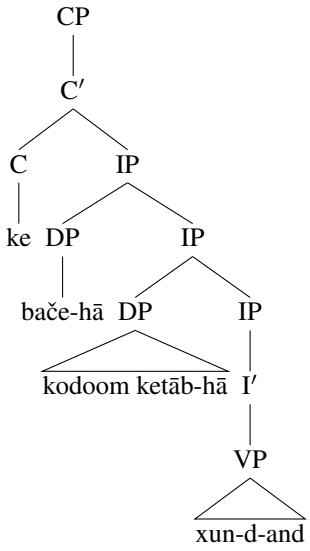
[$_{CP}$ (ke) be-xun-and]]]]]]]
(COMP) SBJV-read-3PL
'Mariam said that the children must read WHICH BOOKS?'



c. Mariam goft
[$_{CP}$ [$_{C'}$ [$_{C}$ ke] [$_{IP}$ kodoom ketāb-hā=ro [$_{IP}$ bače-hā
COMP which book-PL=DO child-PL
[$_{I'}$ [$_{VP}$ xun-d-and]]]]]]
read-PAST-3PL
'Mariam said that the children read WHICH BOOKS?'



d. Mariam goft
[$_{CP}$ [$_{C'}$ [$_{C}$ ke] [$_{IP}$ bače-hā [$_{IP}$ kodoom ketāb-hā=ro
COMP child-PL which book-PL=DO
[$_{IP}$ [$_{I'}$ [$_{VP}$ xun-d-and]]]]]]]
read-PAST-3PL
'Mariam said that, as for the children, they read WHICH BOOKS?'



Example (27a) shows that there is a position for the top of an unbounded dependency below C, since the C position is occupied by an overt complementizer. We

assume that this position is an IP-adjunct, since otherwise the *wh*-phrase would be in regular subject position in SpecIP. Example (27b) shows that the subject of the lower clause can also be fronted/topicalized to an IP-adjoined position. Thus, in (27b), *bačehā* is in a non-agreeing topic position, reflected by the lack of plural agreement on the modal (recall that this modal verb never shows number agreement, cf. 8). This contrasts with example (27c), where there is no modal, and the subject, *bačehā*, is an agreeing subject in the standard SpecIP subject position. Example (27d) shows that the IP-adjunction in (27c) also allows a topicalized *bačehā* occurring adjoined to an IP that itself contains an IP-adjoined *wh*-phrase; this demonstrates that the recursive IP-adjunction rule below is appropriate. Comparing (27b) and (27d), we can similarly observe that the IP-adjuncts can be reversed.[16]

The following rules license the left periphery in the c-structures in (27):[17]

(28) a. $\text{CP} \rightarrow \underset{(\uparrow \text{DISPATH}) \in (\uparrow \text{DIS})}{\text{XP}} \quad \underset{\uparrow = \downarrow}{\text{C}'}$

b. $\text{C}' \rightarrow \underset{\uparrow = \downarrow}{\text{C}} \quad \underset{\uparrow = \downarrow}{\text{IP}}$

c. $\text{I}' \rightarrow \underset{\uparrow = \downarrow}{\text{I}} \quad \left\{ \underset{\uparrow = \downarrow}{\text{VP}} \mid \underset{(\uparrow \text{COMP}) = \downarrow}{\text{CP}} \right\}$

d. $\text{IP} \rightarrow \underset{(\uparrow \text{SUBJ}) = \downarrow}{\text{XP}} \quad \underset{\uparrow = \downarrow}{\text{I}'}$

e. $\text{IP} \rightarrow \underset{\substack{(\uparrow \text{DISPATH}) \in (\uparrow \text{DIS}) \\ (\uparrow \text{DIS})\sigma \in (\uparrow \sigma\iota \{\text{TOPIC} \mid \text{FOCUS})\}}}{\text{XP}} \quad \underset{\uparrow = \downarrow}{\text{IP}}$

Rule (28a) is the normal Indo-European SpecCP rule for the top of an unbounded dependency. Rule (28b) is the normal C′ expansion rule. Rule (28c) is the normal I′ expansion rule (I′ as a co-head with VP), but also allows for the cases in (27a,b) above, in which the c-structure complement to the I′ is a CP which maps to a COMP grammatical function in f-structure. Rule (28d) is the normal SpecIP rule for SUBJ. Rule (28e) accounts for the IP-adjuncts in (27a–d). Lastly, note that given the possibility of multiple fronted discourse functions, as in (27b) and (27d), we assume that the f-structure function DIS (for the top of unbounded dependencies) is set-valued.

[16]It is beyond the scope of this paper to capture the interaction with information structure more precisely than this, but standard LFG methods should allow it.

[17]The equation regarding DIS in (28e) connects the top and bottom of the unbounded dependency in the corresponding f-structure (Dalrymple et al. 2019: 39ff). The grammatical function DIS is a way of unifying the previously distinguished f-structural functions of TOPIC and FOCUS, which are properly elements of i-structure rather than f-structure, as an *overlay function* that captures the abstract f-structural role of the top of unbounded dependencies; this was originally proposed by Asudeh (2004), where the function was named UDF, for *unbounded dependency function*. The set statement ($\in$) regarding DIS and TOPIC/FOCUS in (28e) states that the top of the unbounded dependency encodes a TOPIC or FOCUS at i(nformation)-structure (Dalrymple and Nikolaeva 2011).

We assume the following lexical entry for *bāyad* (□.PRES):

(29) *bayad* I (↑ PRED) = 'must⟨CF⟩SUBJ'
(↑ TENSE) = PRES
{ @EXPL-SUBJ | (↑ SUBJ) = (↑ XCOMP SUBJ) }
{ (↑ COMP MOOD) =$_c$ SUBJUNC | }

This lexical entry is for both the personal and impersonal present modal construction, so some information is shared: The modal in both constructions occupies an identical position, hence the category I is shared. The modal in both constructions is present tense, hence the specification of [TENSE PRES]. The two modals are also forms of the same basic predicate, so have the same PRED value.

However, the information at the lexical entry needs to diverge at some point. The personal and impersonal modals are distinguished by the fact that the personal construction takes a closed sentential complement, COMP, which can realize its own subject, whereas the impersonal is a kind of subject raising construction modelled with functional control; it takes an open sentential complement, XCOMP, which cannot realize its own subject. The personal construction also requires that its complement independently have subjunctive mood.[18]

The lefthand side of (29) calls a template, EXPL-SUBJ. A template call is marked by @. The semantics of template invocation is very simple (Dalrymple et al. 2004): the template just defines a bundle of lexical information and gives it a name; when the template is invoked, the corresponding information it encodes is substituted in. Note that a template may call other templates, so there may be multiple such substitutions; this is also exemplified by EXPL-SUBJ.

(30) EXPL-SUBJ := ¬(↑ SUBJ PRED)
@SUBJ-3SG

(31) SUBJ-3SG := (↑ SUBJ PERS) = 3
(↑ SUBJ NUM) = SG

The righthand case in (29) is for the nonfinite-complements containing apocopated infinitives, e.g. (5). We define the following templates for apocopated infinitives:

(32) APINF(P) := (↑ PRED) = P
@NO-TAM
@IMPERS-SUBJ

(33) NO-TAM := ¬(↑ TENSE)
¬(↑ ASPECT)
¬(↑ MOOD)

(34) IMPERS-SUBJ := (↑ SUBJ PRED) = 'pro'
(↑ SUBJ PRONTYPE) = IMPERSONAL
@SUBJ-3SG

[18]The fact that the subjunctive mood requirement is information that is checked by the modal, rather than information that is actually contributed by it, is modelled by the constraining equation, marked =$_c$ rather than simply =.

Note that the APINF template is one that takes an argument: whatever is passed in as the argument becomes the value of PRED.

The lexical entry for a sample apocopated infinitive, *raft* ('go'), is:

(35) *raft* V @APINF('go⟨SUBJ,OBL⟩')

The f-structures for examples (5), (7a), and (7b) respectively are shown in (36a)–(38a); the corresponding examples are repeated in (36b)–(38b).[19,20]

(36) a.

$$\begin{bmatrix}
\text{PRED} & \text{'must⟨XCOMP⟩SUBJ'} \\
\text{SUBJ} & \begin{bmatrix}
\text{PRED} & \text{'pro'} \\
\text{PERS} & 3 \\
\text{NUM} & \text{SG} \\
\text{PRONTYPE} & \text{IMPERSONAL}
\end{bmatrix} \\
\text{XCOMP} & \begin{bmatrix}
\text{PRED} & \text{'go⟨SUBJ,OBL⟩'} \\
\text{SUBJ} & \\
\text{OBL} & \begin{bmatrix}
\text{PRED} & \text{'to⟨OBJ⟩'} \\
\text{OBJ} & \begin{bmatrix}\text{PRED} & \text{'home'}\end{bmatrix}
\end{bmatrix} \\
\text{ADJ} & \left\{\begin{bmatrix}\text{PRED} & \text{'early'}\end{bmatrix}\right\}
\end{bmatrix} \\
\text{TENSE} & \text{PRES}
\end{bmatrix}$$

b. bāyad zood be xune raf-t.
□.PRES early to home go-INF
'It's necessary to go home early.'/
'One must go home early.'

(37) a.

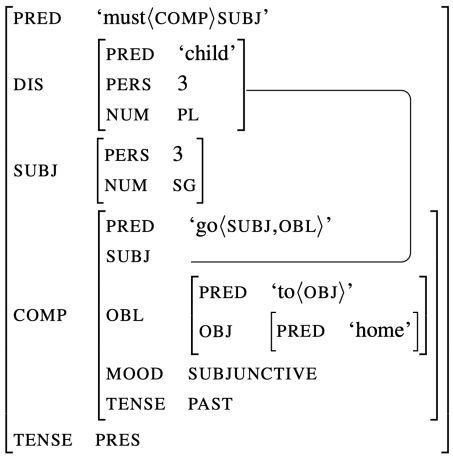
$$\begin{bmatrix}
\text{PRED} & \text{'must⟨COMP⟩SUBJ'} \\
\text{SUBJ} & \begin{bmatrix}\text{PERS} & 3 \\ \text{NUM} & \text{SG}\end{bmatrix} \\
\text{COMP} & \begin{bmatrix}
\text{PRED} & \text{'go⟨SUBJ,OBL⟩'} \\
\text{SUBJ} & \begin{bmatrix}\text{PRED} & \text{'child'} \\ \text{PERS} & 3 \\ \text{NUM} & \text{PL}\end{bmatrix} \\
\text{OBL} & \begin{bmatrix}
\text{PRED} & \text{'to⟨OBJ⟩'} \\
\text{OBJ} & \begin{bmatrix}\text{PRED} & \text{'home'}\end{bmatrix}
\end{bmatrix} \\
\text{MOOD} & \text{SUBJUNCTIVE} \\
\text{TENSE} & \text{PAST}
\end{bmatrix} \\
\text{TENSE} & \text{PRES}
\end{bmatrix}$$

b. bāyad bačehā be xune mi-raf-t-and.
□.PRES child-PL to home IPFV-go-PAST-3PL
'The children had to go home.'

[19]Note that the imperfective and the subjunctive are syncretic in the past tense. Therefore, although for consistency we have always glossed *mi-* as imperfective (IPFV), we assume that it can convey subjunctive mood and hence satisfy the constraining equation in (29). We do not attempt to account for this syncretism here.

[20]See footnote 17 regarding the function DIS in (38).

(38) a.



b. bačehā bāyad be xune mi-raf-t-and.
child-PL □.PRES to home IPFV-go-PAST-3PL
'The children had to go home.'

4.1 Interim summary

We are now in a position to answer the second and third questions in §2. The third question asked, *What is the syntactic structure of simplex modal constructions in Persian?* Persian modals occupy the category I; this is unsurprising from an LFG-theoretic perspective, since it is not unusual for modals to have category I (or C, depending on distribution). This interacts with the general structure of the left periphery that we have provided—see (27) and the c-structure rules in (28)—such that all and only the valid orderings are captured.

The second question asked, *How can we capture the personal and impersonal readings of modals like (4a) vs. (5)?*

(4) a. bāyad be xune be-rav-am.
□.PRES to home SBJV-go.PRES-1SG
'I have to go home.'

(5″) bāyad zood be xune raf-t.
□.PRES early to home go-INF
'It's necessary to go home early.'/
'One must go home early.'

The lexical entry for the modal *bāyad* (□.PRES) in (29) explains the differences by treating the personal as a subcategorized subjunctive COMP and treating the impersonal as a raising predicate, which allows the requirements of the apocopated infinitive, as captured in template (32), to control the reading, with the modal simply wrapping necessity around this.

4.2 Capturing the variation

The fourth question in §2 asked, *How should the variable agreement displayed in (8) be explained?*

(8) b. bače-hā be nazar mi-ā-d ke xaste šo-d-an.
child-PL to opinion IPFV-come.PRES-3SG that tired become-PAST-3PL
'As for the children, it seems that they have gotten tired.'

c. % bače-hā be nazar mi-ā-n ke xaste šo-d-an.
child-PL to opinion IPFV-come.PRES-3PL that tired become-PAST-3PL
'The children seem to have gotten tired.'

Our proposal is based on the contrast between these cases and *bāyad*-clauses, considering our analysis of the latter.

Speakers who only allow the non-agreeing form (8b) maintain an analysis of the preposed nominal, *bačehā* 'children', as a TOPIC in information structure (mapped from a DIS in f-structure). It is a general fact about Persian (and perhaps universally), that topichood is not sufficient to directly trigger agreement. Speakers who do allow the agreeing form (8c) have instead analyzed the preposed nominal as a SUBJ, which robustly triggers agreement in Persian. The verb, *āmadan* 'to come', in this construction, unlike the modals, is a fully agreeing form (*miān*), which allows for (8c). For these speakers, *be nazar āmadan* 'seems', when it shows agreement with a preposed element, is akin to English copy raising (Rogers 1973, Postal 1974):

(39) Harry seems like he is tired.

However, since Persian is *pro*-drop, the embedded pronominal does not surface.[21] When it does not show agreement, as in (8b), the construction is akin to English *seems that* with topicalization; i.e., there is an (in Persian, unrealized) expletive subject with the bare-topicalized nominal occurring in only apparent subject position:

(40) As for Harry, it seems that he is tired.

We hope to have shown that a fairly simple LFG analysis of Persian modal syntax is possible using standard tools of the framework. This analysis lends further support to the view that synchronic Persian grammar indeed does contain an apocopated infinitive, and that this short infinitive's formal resemblance to the past stem/zero-marked PAST.3SG form is misleading.

[21] In fact, one could possibly get it to surface given enough discourse support, but it is difficult because of opposing discourse forces.

5 Conclusion

We have now answered all of the questions that we posed in §2. The first question asked, *How should we account for the complement in the impersonal modal construction?* We argued that it is a (short/apocopated) infinitival which is formally but not functionally identical to the past stem. The formal identity can be captured by standard means, such as rules of referral or their alternatives in other frameworks, as sketched in §3.4 above.

The second and third questions asked, *How can we capture the personal vs. impersonal readings of the modals?* and *What is the syntactic structure of the simplex modal constructions?* With respect to the first of these two questions, we argued that the distinction is governed by the lexical entry for the modal and the templates that it uses. As for the structural question, we showed that the modal is in I. There is a topic position above this, but below C.

The fourth question asked, *How should the variable speaker agreement displayed for the subject of the raising/perception verb* be nazar āmadan *(lit. 'to opinion come'/∼'to seem like/that') be captured?* The variation is enabled by the fact that the verb that anchors this predicate, *āmadan*, is a fully agreeing predicate, unlike the modals. Some speakers have reanalyzed the preposed topic as a subject, since the position it occupies is in many cases string-identical to subject position. On this analysis, the verb must agree with the subject, as is the case overall in Persian grammar. However, the other analysis, in which the preposed nominal is actually a topic, is also available, but does not trigger agreement. Therefore, these speakers display variation in their utterances.

In answering these questions, we have shown that the complement of the modal verb in the relevant Persian impersonal constructions is an apocopated infinitive, and its formal resemblance to the past stem/zero-marked PAST.3SG form is misleading. We also demonstrated that the apocopated counterparts of the Persian long *-an* infinitives are not nominal, which reveals that Persian does, in fact and contrary to common assumptions in the literature, have nonfinite clauses. Moreover, we provided an LFG account of modal syntax in Persian that can account for our data and the empirical generalizations. Finally, we argued that our account can also capture the seemingly puzzling agreement patterns of the verb *be nazar āmadan* 'to seem' by attributing a copy-raising analysis to the agreeing forms.

References

Anoushe, Mazdak. 2018. A Revision of Persian Past Tense Inflection: A Distributed Morphology Approach. *Language Related Research* 9(1): 57–80. In Persian.

Asudeh, Ash. 2004. Resumption as Resource Management. Ph.D. thesis, Stanford University.

Asudeh, Ash. 2005. Control and Semantic Resource Sensitivity. *Journal of Linguistics* 41(3): 465–511.

Asudeh, Ash, Mary Dalrymple, and Ida Toivonen. 2013. Constructions with Lexical Integrity. *Journal of Language Modelling* 1(1): 1–54.

Asudeh, Ash, and Daniel Siddiqi. 2022. Morphology in LFG. To appear in Dalrymple (2023).

Bano, Saira, Miriam Butt, and Ashwini Deo. 2019. Achievement Predicates and Tense Paradigms in Hazaragi. In *Proceedings of the LFG19 Conference*, 394–414.

Bhatt, Rajesh, Tina Bögel, Miriam Butt, Annette Hautli, and Sebastian Sulger. 2011. Urdu/Hindi Modals. In *Proceedings of the LFG11 Conference*, 47–67.

Bjorkman, Bronwyn M., and Claire Halpert. 2017. In an Imperfect World: Deriving the Typology of Counterfactual Marking. In Ana Arregui, María Luisa Rivero, and Andrés Salanova, eds., *Modality Across Syntactic Categories*, 157–178. Oxford University Press.

Chomsky, Noam, and Howard Lasnik. 1977. Filters and Control. *Linguistic Inquiry* 8: 425–504.

Dalrymple, Mary, ed. 2023. *The Handbook of Lexical Functional Grammar*. Language Science Press. Forthcoming.

Dalrymple, Mary, Ronald M. Kaplan, and Tracy Holloway King. 2004. Linguistic Generalizations over Descriptions. In Miriam Butt and Tracy Holloway King, eds., *Proceedings of the LFG04 Conference*, 199–208. Stanford, CA: CSLI Publications.

Dalrymple, Mary, John J. Lowe, and Louise Mycock. 2019. *The Oxford Reference Guide to Lexical Functional Grammar*. Oxford: Oxford University Press.

Dalrymple, Mary, and Irina Nikolaeva. 2011. *Objects and Information Structure*. Cambridge: Cambridge University Press.

Darzi, Ali, and Saera Kwak. 2015. Syntax and Semantics of Subjunctive Clauses in Persian. *Lingua* 153: 1–13.

Ghomeshi, Jila. 2001. Control and Thematic Agreement. *Canadian Journal of Linguistics* 46: 9–40.

Kaplan, Ronald M., and Annie Zaenen. 2023. Unbounded Dependencies. Forthcoming in Dalrymple (2023).

Karimi, Simin. 2005. *A Minimalist Approach to Scrambling: Evidence from Persian*. Berlin: De Gruyter Mouton.

Karimi, Simin. 2008. Raising and Control in Persian. In Simin Karimi, Vida Samiian, and Don Stilo, eds., *Aspects of Iranian Linguistics.*, 177–208. Cambridge Scholars Publishers Ltd.

Labbafankhosh, Zahra, and Ali Darzi. 2015. On the Syntactic Analysis of Persian Modals: Šâyad 'Perhaps' and Bâyad 'Must'. *Journal of Linguistic Research* 5(2): 97–112. In Persian.

Lenepveu-Hotz, Agnès. 2012. Etude diachronique du système verbal persan (XeXVIe siècles) : d'un équilibre à l'autre? Ph.D. thesis, Ecole pratique des hautes études, Paris.

Lowe, John J. 2019. The Syntax and Semantics of Nonfinite Forms. *Annual Review of Linguistics* 5(1): 309–328.

Manzini, M.Rita, and Anna Roussou. 2000. A Minimalist Theory of A-movement and Control. *Lingua* 110(6): 409–447.

Melchin, Paul B., Ash Asudeh, and Dan Siddiqi. 2020. Ojibwe Agreement in Lexical-Realizational Functional Grammar. In Miriam Butt and Ida Toivonen, eds., *Proceedings of the LFG20 Conference*, 268–288. Stanford, CA: CSLI Publications.

Mirrazi, Zahra. 2022. Tense in Conditionals: Ins and Outs. Ph.D. thesis, University of Massachusetts Amherst.

Postal, Paul M. 1974. *On Raising: One Rule of English Grammar and its Theoretical Implications*. Cambridge, MA: MIT Press.

Rogers, Andy. 1973. Physical Perception Verbs in English: A Study in Lexical Relatedness. Ph.D. thesis, UCLA.

Ross, John R. 1967. Constraints on Variables in Syntax. Ph.D. thesis, MIT.

Samvelian, Polet, and Reza Mir-Samii. 2007. Les infinitifs en persan. *Faits de langues* 30: 117–135.

Samvelian, Pollet. 2018. Specific Features of Persian Syntax: The Ezâfe Construction, Differential Object Marking and Complex Predicates. In Anousha Sedighi and Pouneh Shabani-Jadidi, eds., *The Oxford Handbook of Persian Linguistics*, 226–269. Oxford University Press.

Skjaervø, Prods Oktor. 2009. Middle West Iranian. In Gernot Windfuhr, ed., *The Iranian Languages*, 196–278. London: Routledge.

Storoshenko, Dennis Ryan, Elias Abdollahnejad, and Mahyar Nakhaei. 2020. Copy Raising in Persian. In *Proceedings of the 2020 annual conference of the Canadian Linguistic Association*.

Stump, Gregory T. 2001. *Inflectional Morphology: A Theory of Paradigm Structure*. Cambridge: Cambridge University Press.

Stump, Gregory T. 2016. *Inflectional Paradigms: Content and Form at the Syntax-Morphology Interface*. Cambridge: Cambridge University Press.

Taleghani, Azita H. 2008. *Modality, Aspect and Negation in Persian*. Amsterdam: John Benjamins.