

### Persian

— annotation notes —

June 2016

Shirin Adibifar

University of Bamberg

cite this document as:

Adibifar, Shirin. 2016.  
Multi-CAST Persian annotation notes.  
In Haig, Geoffrey & Schnell, Stefan (eds.),  
*Multi-CAST (Multilingual Corpus of Annotated Spoken Texts)*.  
(handle) (date accessed.)

## Annotation notes

### Subordinate and relative clauses

In Persian, most subordinate clauses (including relative clauses) are introduced by the all-purpose complementizer *ke*, and involve finite verb forms governing a set of arguments basically identical to those of independent clauses. We therefore count them as clause units containing a normal <:pred>. The example in (1) contains such a complement clause:

- (1) a. # *mard miāyad pāyin*  
 # man come.PRS.INDIC.3SG down  
 # np.h:s v:pred rv
- b. # *0 mibinad*  
 # 0 see.PRS.INDIC.3SG  
 # 0.h:a v:pred
- c. # *ke teki az sabadhā xāli ast*  
 # that one of basket.PL empty be.PRS.3SG  
 #cc ke ind\_pro:s rn\_adp rn\_np:obl other:pred cop  
 ‘The man comes down [from the tree] and finds out that one of the baskets is empty.’

*persian.g2-f-01\_018*

When glossing relative clauses it is important to note that the head noun is usually systematically gapped in the relative clause (i.e. it cannot be overtly expressed). In such cases, we do not gloss a zero in the relative clauses, because speakers have no choice between zero and overt argument expression (following the rationale of glossing zeroes in Bickel 2003), with the result that in a large number of relative clauses there is no representative of a core argument in the GRAID annotation. Example (2) illustrates subject relativization, where overt expression of the subject NP is systematically banned from the relative clause, while (3) illustrates systematic gapping of the object in object relativization:

- (2) a. # *0 mixorad be yek doxtari*  
 # 0 hit.PRS.INDIC.3SG to one girl.INDEF  
 # 0.h:s\_cp v:pred adp ln\_deti np.h:g
- b. # *ke dāšte az ān taraf barmigašte*  
 # that AUX.PST.3SG from that side return.PST.PTCP  
 #rc ke aux adp ln\_dem np:l v:pred  
 ‘(He) runs into a girl who was coming back from the opposite direction.’

*persian.g1-f-08\_011*

- (3) a. *in mivehā =rā*  
 this fruit.PL =ACC  
 ln\_dem np:p =acc\_rn
- b. # *ke 0 jam mikonand*  
 # *that 0 collected do.PRS.INDIC.3PL*  
 #rc *ke 0.h:a lvc v:pred*  
 ‘these fruits that they gather ...’

*persian\_g1-f-05.007*

Relative clauses are frequently centre-embedded, in which case standard GRAID procedure is followed, indicating the right edge of the embedded clause with the symbol <%> (unless it coincides with the right edge of its matrix clause):

- (4) a. # *ān se tā tačeyi*  
 # *that three piece kid.INDEF*  
 # *ln\_dem ln\_qu ln\_class np.h:a*
- b. # *ke dārand miravand %*  
 # *that AUX.PRS.3PL go.PRS.INDIC.3PL %*  
 #rc *ke aux v:pred %*
- c. *kolāh =rā peydā mikonand*  
 hat =ACC found do.PRS.INDIC.3PL  
 np:p =acc\_rn lvc v:pred  
 ‘Those three boys that are just leaving find the hat.’

*persian\_g2-f-04.009*

In a small number of cases centre-embedded structures would have required complex (and controversial) syntactic annotation. In order to avoid undue complications, we treated the relevant strings as <nc>, but annotated the matrix clause – to the extent that it is a syntactically well-formed clause – in the normal way:

- (5) a. # *0 kolāh =aš =rā*  
 # *0 hat =POSS.3SG =ACC*  
 # *0.h:a np:p =pro.h:poss =acc\_rn*
- b. # *ke didand ruye zamin ast %*  
 # *that see.PST.INDIC.3PL on earth is %*  
 #nc *nc nc nc nc nc %*
- c. *be =heš bargardāndand*  
 to =3SG return.PST.INDIC.3PL  
 adp =pro.h:g v:pred  
 ‘(They) returned his hat to him which they saw lying on the ground.’

*persian\_g2-m-11.006*



- (7) a. # *bad čand tā az peserhā*  
 # then a.few piece of boy.PL  
 # other ln\_qu class\_np.h:a rn\_adp rn\_np.h:obl
- b. # *ke az hamān jā dāštand rad*  
 # that from same place AUX.PST.3PL crossing  
 #rc ke adp ln\_lex np:l aux lvc  
*mišodand* %  
 become.PST.INDIC.3PL %  
 v:pred %
- c. *āmadand komak =aš kardand*  
 come.PST.3PL help =PRO.3SG do.PST.3PL  
 aux lvc =pro.h:p v:pred
- d. # *0 golābihā =rā jam kardand*  
 # 0 pear.PL =ACC collecting do.PST.3PL  
 # 0.h:a np:p =acc\_rn lvc v:pred
- e. # *0 dāxele zanbil rixtand 0*  
 # 0 inside basket pour.PST.3PL 0  
 # 0.h:a adp np:g v:pred 0:p
- ‘Some boys who were **passing by** came and **helped** him **gather up** the pears, and put them back in the basket.’

*persian.g2-f-07\_011*

### Non-canonical subjects

In Persian, subjects can be uncontroversially defined in terms of (i) their ability to control agreement suffixes on the verbal predicate, and (ii) their lack of overt case marking. These morphological features also correlate with syntactic features such as the ability to control reflexives, or co-referential deletion. However, a set of predicates in Persian has NPs that show most of the typical properties of subjects, but lack the ability to control agreement suffixes on the verb. We refer to them as **non-canonical subjects** (NCS). Semantically, NCSs are generally EXPERIENCERS, or some kind of external POSSESSOR or BENEFACTIVE. Typically they occur with complex predicates (CP), and the non-verbal element of the CP obligatorily carries a possessive clitic reflecting person and number of the NCS. Functionally, this is evidently a kind of “agreement”, though the exponent of agreement is not a verbal suffix, but a possessive clitic. In this kind of construction, we gloss the possessive clitic in the same manner as other possessive clitics, and the NCS is glossed with the function gloss <ncs>. If the NCS is not present in the clause, it receives a zero gloss in GRAID.

- (8) # 0      *çeşm* =aş      *in*      *sabad-hā* =rā  
 # 0      eye =POSS.3SG    this      basket.PL =ACC  
 # 0.h:ncs lvc    =pro.h:poss    ln\_dem    np:p      =acc\_rn  
*gereft*  
 catch.PST.3SG  
 v:pred

‘(He) caught sight of these baskets (lit. his eye took the baskets)’

*persian\_g1-f-05\_005*

- (9) # *va*    0      *çaşm* =aş      *mioftad*      *be*  
 # and    0      eye =POSS.3SG    fall.PRS.INDIC.3SG    to  
 # other 0.h:ncs lvc    =pro.h:poss    v:pred      adp

*golābihā*  
 pear.PL  
 np:obl

‘(He) caught sight of the pears (lit. his eyes fall on the pears)’

*persian\_g1-m-13\_012*

- (10) # *bad*    *in*      *ham*    *havās*      =aş      *part*  
 # then    3SG      ADD    attention =POSS.3SG    separated  
 # other    pro.h:ncs    other    lvc      =pro.h:poss    lvc

*mišavad*  
 become.PRS.INDIC.3SG  
 v:pred

‘[His hat fell off] and then he got distracted (lit. he his.attention became separated)’

*persian\_g1-f-14\_013*

## Complex noun phrases

### *NP-internal classifiers, quantifiers, and demonstratives*

The speakers make very frequent use of NPs of the type ‘three pieces (of) X’, involving a quantifier (often a numeral, but also indefinite expressions such as ‘one’, ‘some’ etc.), a classifier (e.g. *tā* ‘piece’), and a noun, in some cases linked to the entire expression with the preposition *az* ‘from’. These expressions lead to certain issues in analysis, particular in deciding on the head. Structurally, the classifier expression is the head, while semantically, the complement of the preposition *az* is the head.

When classifiers and quantifiers are combined in the NP, we gloss <ln.class> and <ln.qu> respectively, while treating the lexical noun as the head, and adding the function gloss to it, as in (11–12):

- (11) # *se tā pesarbaçeye digar nazdiktar istāde*  
 # *three piece little.boy other closer stand.PST.PTCP*  
 # *ln\_qu ln\_class np.h:s rn\_lex other v:pred*  
*budand*  
 AUX.PST.3PL  
 aux

‘Three boys were standing nearby’

*persian\_g1-f-02\_015*

- (12) # *hameye golāihā mirizad*  
 # *all pear.PL pour.PRS.INDIC.3SG*  
 # *ln\_qu np:s v:pred*  
 ‘All the pears spill out’

*persian\_g1-f-01\_010*

Analogously, we gloss NP-internal demonstratives with <ln\_dem>, as demonstrated in (13):

- (13) # *bad in āqā dobāre miravad bālāye deraxt*  
 # *then this man again go.PRS.INDIC.3SG top.of tree*  
 # *other ln\_dem np.h:s other v:pred adp np:l*  
 ‘Then he climbs up the tree again’

*persian\_g1-f-01\_006*

In the absence of a lexical head, the classifier or quantifier is treated as the head and receives the appropriate function gloss, as in (14):

- (14) # *in se tā dāštand miraftand*  
 # *this three piece AUX.PST.PL IMPF.go.PST.3PL*  
 # *ln\_dem ln\_qu class\_np.h:s aux v:pred*  
 ‘These three were leaving’

*persian\_g1-m-04\_009*

The same procedure is adopted for indefinite pronouns, where we use the gloss <ind\_pro>:

- (15) # *bad 0 yeki =ş =rā*  
 # *then 0 one =POSS.3SG =ACC*  
 # *other 0.h:a ind\_pro:p =pro:obl =acc\_rn*  
*barmidārad*  
 pick.up.PRS.INDIC.3SG  
 v:pred

‘Then he picks up one of them’

*persian\_g1-f-01\_004*

- (16) # *yeki*            *az*        *ān*        *baḡehā*        *bā*    *sut*  
 # **one**            of        that        kids        with    whistle  
 # **ind\_pro.h:a** rn\_adp rn\_dem rn\_np.h:obl adp np:obl  
  
*pesar*    =*rā*        *sedā*        *mikonad*  
 boy        =ACC        calling do.PRS.INDIC.3SG  
 np.h:p    =acc\_rn lvc        v:pred  
 ‘One of the kids calls the boy by whistling’

*persian.g1-m-13.025*

*Partitive modifiers within the NP*

In several cases we find a lexically light expression (classifier, indefinite pronoun, quantifier etc.) modified by a prepositional phrase, yielding expressions like ‘three of the boys’, and so on. In these cases we have treated classifiers or quantifiers as the head (and hence carrier of the function gloss) in examples such as the following, taken from above. Where partitive expressions within the NP occur, they are considered <:obl>:

- (17) *čand tā*            *az*        *peserhā*  
 a.few piece        of        boy.PL  
 ln\_qu class\_np.h:a rn\_adp rn\_np.h:obl  
 ‘... a few of the boys’

*persian.g2-f-07.011*

- (18) # *ke yeki*        *az*        *ān*        *zanbilhā*    =*rā*  
 # that one        of        that        basket.PL =ACC  
 #rc other in\_pro:p rn\_adp rn\_dem rn\_np:obl =acc\_rn  
  
*gozāšt*        *ruye doḡarxe* =*aš*  
 put.PST.3SG on bike        =POSS.3SG  
 v:pred        adp np:l        =pro.h:poss  
 ‘... he puts one of the baskets on his bike’

*persian.g2-f-07.007*

- (19) *yek dāne*        *az*        *sabade*        *golābihāyi*  
 one piece        of        basket=EZAFE pear.PL.INDEF  
 ln\_qu class\_np:p rn\_adp rn\_np:obl rn\_lex  
  
 # *ke 0*        *čide*        *bud*  
 # that 0        pick.PST.PTCP AUX.PST.3SG  
 #rc ke 0.h:a v:pred        aux  
 ‘one basket of pears that (he) had picked’

*persian.g1-f-02.008*



*Lexical modifiers within the NP*

Where lexical modifiers (adjectives or nouns) are included in the NP (generally linked via the *ezafe* particle), we have glossed them with `<rn.lex>`:

- (20) # *yeki*            *az*        *sabadhāye*        *golābi* =*aš*  
 # one                of        basket.PL=EZAFE    *pear* =POSS.3SG  
 # ind\_pro.h:s   rn\_adp   rn\_np:obl        *rn.lex* =pro.h:poss
- nist*  
 NEG.be.PRS.3SG  
 cop
- ‘One of the baskets is not here’

*persian.g1-f-01.018*

**List of corpus-specific GRAID symbols**

- `<=acc.rn>` object postpositional particle *rā*
- `<rn.lex>` lexical modifier within the NP; traditionally an adjective or some other item of uncertain word class
- `<class_np>` classificatory particle, subtype of NP
- `<ln_class>` classificatory particle, leftward modifier of a NP
- `<qu_np>` quantifier, subtype of NP
- `<ln_qu>` quantifier, leftward modifier of a NP
- `<ln_deti>` indefinite article as determiner within a NP
- `<ln_dem>` demonstrative as determiner within a NP
- `<ind_pro>` indefinite pronoun
- `<lvc>` light verb complement
- `<ke>` complementizer *ke*

**References**

- Bickel, Balthasar. 2003. Referential density in discourse and syntactic typology. *Language* 79(4). 708–736. [p. 1]