Persian
— annotation notes —

June 2016

Shirin Adibifar
University of Bamberg

cite this document as:

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 sabadāhā xāli ast
    # that one of basket.PL empty be.PRS.3SG
    #cc ke ind:pros 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.’

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:gr
b. # ke dāSTE az ān taraf barmigaSTE
    # 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.’
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.’

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 %
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.’
Complex predicates

Complex predicates (CP) in Persian are conventionalized combinations of a non-verbal element with a light verb, which together create the predicate of a clause. Both the non-verbal element and the light verb contribute to the resulting semantics, but CPs are often not semantically compositional. Complex predicates raise a number of problems in connection with GRAID. The first issue is to decide on the transitivity value of the entire expression, as this determines whether we gloss the subject with <:s> or with <:a>. We identify a verb as transitive if it has the ability to govern a direct object marked with the clitic nā. We distinguish four possible coding scenarios, each glossed as follows:

(6) a. intransitive light verb + entire CP is intransitive
    rad šodan ‘pass by’ ➤ subject glossed <:s>

    b. transitive light verb + entire CP is intransitive
    farār kardan ‘escape’ (lit. ‘escaping do’)
    zamin xordan ‘fall down’ (lit. ‘earth eat’)
    ➤ subject glossed <:s_cp>

    c. transitive light verb + entire CP is transitive
    yād gereftan ‘learn’ (lit. ‘memory take’)
    ➤ subject glossed <:a>

    d. intransitive light verb + entire CP is transitive
    balad budan ‘know’ ➤ subject glossed <:a_cp>

For the purposes of cross-corpus comparison, the additional underscores can be ignored and the <:a_cp> and <:s_cp> symbols included in the <:a> and <:s> categories respectively.

The second issue in annotating CPs is the status of the non-verbal element. CPs are typically highly conventionalized, and the non-verbal element is generally not referential, hence could be simply included into the predicate gloss. However, there are also borderline cases, and the class of CPs in Persian cannot be readily distinguished from other expressions involving indefinite or generic objects. We have generally applied a neutral <lvc> ‘light verb complement’ gloss for these elements, which mainly serves the purpose of identifying complex predicates in the annotation in case researchers are particularly interested in their properties.

The example in (7) illustrates the annotation procedure. A special kind of CP involving non-canonical subjects is dealt with in the next section.
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.
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 pesarbaçe ye digar naz diktar istāde
     # three piece little.boy other closer stand.PST.PTCP
     # in_qu in_class np:h:s rn_lex other v:pred

     budand
     AUX.PST.3PL
     aux

     ‘Three boys were standing nearby’

(12) # hameye golāihā mirizad
    # all pear.PL pour.PRS.INDIC.3SG
    # ln.qu np:s v:pred

    ‘All the pears spill out’

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’

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’

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

(15) # bad 0 yeki =§ =nā
    # 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’
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»:

(16) # yeki az ān baçehā bā sut
  # one of that kids with whistle
  # ind.pro.h:a mn_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 lv= v:pred

‘One of the kids calls the boy by whistling’

(17) cand tā az peserhā
    a.few piece of boy.PL
    ln_qu class_np.h:a mn_adp rn_np.h:obl

‘… a few of the boys’

(18) # ke yeki az ān zanbilhā =rā
    # that one of that basket.PL =ACC
    #rc other in_prop mn_adp mn_dem mn_np:obl =acc rn
    gozāšt ruye doçarxe =aş
    put.PST.3SG on bike =POSS.3SG
    v:pred adp np:l =pro.h:poss

‘… he puts one of the baskets on his bike’

(19) yek dāne az sabade golābihāyi
    one piece of basket=EZAFE pear.PL.INDF
    ln_qu class_np:p mn_adp mn_np:obl rn_lex
    # ke 0 çide 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’
**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’

**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**