Atlas of the Languages of Iran (ALI)

Morphosyntax explanation and glossing key

(version: 2020/03/27)

The purpose of the ALI questionnaire is to give an overall picture of linguistic structures found in the languages of Iran, and how these structures are distributed geographically. Conversely, it addresses key grammatical structures only, and although it may be a useful tool for initial field research, it is not intended to provide a comprehensive account of any single language.

The previous version (Version 1) of the questionnaire was part of the multi-section ALI pilot questionnaire developed for the languages of Iran by Erik Anonby in 2014, much of which was based on the list of areally salient features in Don Stilo's *Atlas of the Iran–Araxes Linguistic Area* (in preparation) and several other questionnaires (see the file "ALI Questionnaire Intro and Instructions").

Testing and revising the morphosyntax section of the questionnaire

Over a three-year period, various researchers carried this pilot questionnaire out in about 50 locations in Iran. In 2017, for the ALI Questionnaire Workshop held in Bamberg, Germany (https://www.uni-bamberg.de/en/aspra/workshop-questionnaire-languages-of-iran-2017/), participating scholars carried out the questionnaire in additional locations and provided a detailed critique based on their experiences.

In general, the purpose and value of the questionnaire were affirmed. However, a number of important changes suggested by this group of scholars in relation to the morphosyntax section of the questionnaire were as follows:

- Shorten the morphosyntax section to ensure that the entire questionnaire can be easily completed in half a day.

- Focus on morphosyntactic functions without asking for or expecting particular morphological categories, even within a single language family.
- Separate analysis from the data elicitation process, replacing questions that require expert knowledge of a topic in a given language (e.g., "How does this language mark gender?") with simple translations that can be elicited by any researcher.
- Expand the range of morphosyntactic functions covered by the questionnaire to include typologically important topics such as verbal alignment.
- Reduce redundant questionnaire items, that is, items that test the same function as multiple other items.
- In order to increase the contribution of each question, and further reduce the number of questions, test multiple functions in the same question a) if the functions are unlikely to interact or b) if the interaction between functions is typologically relevant.
- For morphosyntactic functions which only make sense in the context of a longer utterance (for example: definiteness), provide the necessary discourse context.

Morphosyntax in the current version of the questionnaire

This revised morphosyntax section in the current version of the ALI questionnaire has been developed by Geoffrey Haig and Erik Anonby, in consultation with the participants of the 2017 ALI Questionnaire Workshop. In keeping with the guidance of the scholars at the workshop, it represents a compromise between the demands of comprehensive coverage, and practicability of administration. It is intended for use in a maximally large number of research locations – an experienced field researcher should be able to complete this section in about 90 minutes – and for this reason the questions are concise, carefully designed, and carefully selected.

As in the first version, topics have been selected with Iranic languages in mind, but are of general typological importance and the questions are well-suited for handling salient morphosyntactic features of languages from other families (Turkic, Semitic, Indic, etc.) as well.

Rather than eliciting isolated words and phrases, most of the questions here are based on a sentence format, along with several verb paradigms. The items dealing with definiteness constitute a coherent paragraph. The rationale is that it is easier for speakers to understand and produce a meaningful utterance, than isolated words. In addition, through careful sentence design, it is possible to capture several aspects of morphosyntax (for example alignment, flagging, word order, etc.) in a single sentence.

As with all the questionnaire sections, an English version and a Persian version are available.

Annotation (glossing of morphosyntactic functions)

For each item in this morphosyntax section, an annotation "tag" (i.e., code) is provided to identify relevant functions and to facilitate later searches for these functions. The annotation system is based primarily on the GRAID system developed by Haig & Schnell (2014)[1] and adapted for the present questionnaire.

[1] Available at: https://lac.uni-koeln.de/corpora/Multi-CAST/_multicast_background/Annotations/HaigSchnell2014_GRAID-Manual7.pdf

In the annotation used here, each clause constituent is represented by a combination of tags which give information about the kind of constituent, its animacy features, number, syntactic function, and so on. The purpose of this system is that researchers who are specifically interested in certain morphosyntactic features (e.g., how objects are marked, whether adjectives agree in gender with head nouns, whether there is any systematic marking of definiteness, etc.) can perform a search of the annotations to find the relevant questionnaire items. The conventions and abbreviations and used in the annotations are briefly outlined below.

Please note that the annotation system is primarily designed to enable standardized queries across the data once they have been entered into the data bank. The details of annotation are therefore not of primary importance for the fieldworker who is carrying out the questionnaire.

Comments, questions or further suggestions on annotations can be directed to Geoffrey Haig (geoffrey.haig@uni-bamberg.de).

Basic annotation conventions

- Individual clause constituents are separated with commas.
- If a constituent includes (or, in the target languages, is likely to include) more than one grammatical word, the tag sequences are linked by a <+> sign.
- Tags that are part of the same word are separated by a full stop <.>
- The sequence of tags within a word is:
  - semantic class (e.g., human, animate, inanimate)
  - number (e.g., singular, plural)
  - syntactic function (e.g., verbal predicate, possessor)
  - In addition, some other optional tags may be added to provide more detailed specifications.
- Syntactic functions are only annotated when they can be reasonably predicted from the content of the elicitation items (in English or Persian). For nouns elicited in isolation (e.g., 1.1–1.10), or when the nature of the syntactic function is disputable or ambiguous, no syntactic function is provided.

Note: where necessary, tags in this document are placed in angle brackets (= < … >) to distinguish them from other letter combinations, but the brackets are not part of the tag inventory.

Full tag inventory

Tag	Explanation
#	optional tag indicating clause type (e.g., <#rc> relative clause; <#q> interrogative clause)
1	first person
2	second person
a	subject of a transitive verb
abl	ablative; also: standard of comparison with comparative adjectives
adj	adjective

adv	adverb
anim	animate (but not human) noun
can	modal expression of ability
com	comitative
comp	comparative grade of an adjective
cop	copula element linking a subject to a non-verbal predicate
def	definite
dem	demonstrative
exp	experiencer
f	human noun, female
g	goal, recipient, or addressee
gen	generic
h	human noun, either unspecified for gender, or male
indef	indefinite
inf	infinitive
inst	instrument
kin	kinship term
loc	locative
n	inanimate noun
neg	negation
num	numeral
p	direct object
pl	plural
prop	proper name
poss	possessor (with nouns and pronouns); predicative possession (with verbs)
pred	predicate
predex	existential predicate
pro	third person pronoun, either unspecified for gender, or male
pro1	first person pronoun
pro2	second person pronoun
prop	proper name
pst	past tense
q	interrogative (clause)
rc	head of a relative clause
s	subject of an intransitive predicate
sa	subject of a predicate with uncertain transitivity value
super	superlative grade of adjective
t	transitive
vpred	verbal predicate

want	modal expression of desire
wh	interrogative pronoun

Examples of tags and tag sequences

Illustrative examples of typical tags and tag sequences used in the annotations are as follows:

Gloss	**Explanation**
<n>	an inanimate noun (e.g., 1.7 'house')
<anim.pl>	an animate noun, plural number (e.g., 1.6 'goats')
<adj+h.pl.s>	an adjective modifying a human noun, plural number, and subject of an intransitive clause (e.g., 3.3 'the tall boys …')
<h.pl.s, cop, n.loc>	a human noun, plural number, subject of an intransitive clause, accompanied by a copula, and an inanimate locative noun (e.g., 5.9 'the children are in the kitchen')
<pro.h.poss+h.kin.f.g>	a noun, expressing a female human kinship term, possessed by a pronominal human possessor, and in the function of goal, addressee, or recipient (e.g., '…to his mother', cf. 5.7).

Using the annotations

The syntactic annotations do not appear on the questionnaire papers that are administered in the field. However, each item on the questionnaire is listed there with a unique number (1.1, 1.2, 2.1, etc.), and the tags are inventoried as a set at the end of this document using the same item numbering. The annotations themselves provide a computer-searchable text, linked to the item numbering just mentioned, so that researchers can easily find all items on the questionnaire that contain a particular functional or semantic element (e.g., a human plural noun, or a transitive verb, or a direct object). When the search function is implemented online, it will also enable the use of regular expressions[2] to make queries more efficient and powerful.

If researchers are interested in individual lexical items (e.g., 'goat', 'rope', 'girl', etc.), searches can also be initiated based on the English or Persian questionnare items. Consequently, the morphosyntax questionnaire can be treated as a repository of lexical data, as used in particular grammatical contexts.

A more detailed description of the tagging system is in preparation (Haig forthcoming). For the moment, an illustrative example may be helpful:

Suppose you are interested in the way direct objects are marked, and whether the marking is sensitive to definiteness, or animacy, or person distinctions. In this case, you can search for all tag sets containing the sequence <.p> (direct objects). You would want to exclude those items

[2] See, for example, https://www.regular-expressions.info/quickstart.html for a brief introduction to regular expressions.

where <.p> is directly followed by another letter; that is, you need to rule out <.poss> (possessor) and <.pl> (plural). This can be done manually, by deleting these items from the search result; or automatically, using regular expressions, e.g., by searching for "p.\b" (where "\b" is the regular expression that signifies a word boundary); or in other ways. The way this is implemented depends how the query interface is ultimately set up online, but the point is that refining a search can be accomplished quite easily because of the format of the annotations.

A search of this nature would yield the following items which could contain a direct object in the target language:

Items containing <.p> (direct objects)

2.5, 3.3–3.5, 5.1–5.7, 5.15–5.17, 6.1, 6.2, 6.8, 6.11–6.14, 9.2–9.17

These clauses contain various kinds of direct objects (singular/plural, 1st/2nd/3rd person, etc.), so an analysis of these items, and how they pattern in different tenses, would provide considerable insight into the system of object marking in the language.

Of course, many questions will remain open – all possible combinations of person, animacy, tense, and definiteness cannot be included in the questionnaire, since this would require around 50 different forms. However, within the constraints of a short questionnaire, the main features of such systems can be efficiently obtained for a large set of languages.

Once an overview of direct object systems is available, and systems of various types have been identified, a small number can be targeted for more detailed investigation.

Inventory of morphosyntactic annotation tags for questionnaire items

1	**Number: plurality and related functions**
1.1	f
1.2	f.pl
1.3	h
1.4	h.pl
1.5	anim
1.6	anim.pl
1.7	n
1.8	n.pl
1.9	n
1.10	n.pl

2	**Definiteness and related functions**
2.1	n.loc+pro1.pl.poss, h.indef.s, predex, n+pro.h.poss, prop.h
2.2	prop.h.def.sa, predposs, num+h.pl.indef, h.indef, f.indef
2.3	h.def.sa, vpred, f.def.g
2.4	pro1.pl.sa, vpred, dem+n.loc
2.5	#q, pro2.a, want, vpred, adj+n.indef.p
2.6	h.def.poss+f.kin.def.sa, vpred
2.7	pro.h.sa, vpred, n.def.rc, #rc, pro.1.pl.sa, pro.loc, vpred

3	**Nouns and adjectives: agreement and comparison**
3.1	adj+h.s, vpred, dem+adj+n.g
3.2	adj+f.s, cop, n.loc
3.3	adj+h.pl.a, vpred, n.pl.p, prop.h.g
3.4	adj+f.pl.a, vpred, dem+n.pl.def.p
3.5	adj+f.pl.exp, vpred, n.indef.p
3.6	dem+n.def.s, cop, adj.pred
3.7	dem+n.def.s, cop, adj.comp.pred, dem+n.abl
3.8	dem+n.def.s, cop, adj.super+n.pred, n.loc

4	**Noun phrase structure, possessive pronouns, constituent word order**
4.1	pro1.poss+anim.s, cop, adj.pred
4.2	pro2.poss+anim.s, cop, adj.pred
4.3	prop.h, pro.h.poss+anim.s, cop, adj.pred
4.4	prop.f, pro.f.poss+anim.s, cop, adj.pred
4.5	pro1.pl.poss+anim.s, cop, adj.pred
4.6	pro2.pl.poss+anim.s, cop, adj.pred
4.7	pro.pl.poss+anim.s, cop, adj.pred

4.8	prop.f.poss+num+adj+h.kin.pl.s, vpred, n.g
4.9	prop.h.poss+h.kin.sa, vpred, n.loc
4.10	pro1.poss+num+adj+anim.pl.s, vpred
4.11	prop.h.a, vpred, pro.h.poss+anim.pl.p
4.12	h.a, vpred.t.pst, pro.poss+anim.p
4.13	h.poss+anim.s, vpred.pst
4.14	h.pl.poss+anim.pl.s, vpred.pst

5	**Flagging: orientation and relations**
5.1	h.a, vpred.t.pst, n.p, n.inst
5.2	f.a, vpred.t.pst, n.pl.p, n.inst
5.3	pro.h.a, vpred.t.pst, n.p, f.g
5.4	prop.h.a, vpred.t.pst, pro.p, f.pl.g
5.5	f.pl.a, vpred.t.pst, n.p, h.g
5.6	prop.f.a, vpred.t.pst, n.p, pro1.g
5.7	prop.h.sa, vpred, n.p, pro.h.poss+f.kin.g
5.8	h.pl.s, cop.neg, prop.loc
5.9	h.pl.s, cop, n.loc
5.10	prop.h.s, vpred, n.g, prop.h.com
5.11	pro1.s, vpred, n.abl
5.12	pro1.sa, pred.poss, f.kin.indef
5.13	pro1.sa, pred.poss, n.indef
5.14	dem, cop, pro1.poss.pred
5.15	pro1.pl.a, vpred.neg.t.pst, prop.h.p, n.loc
5.16	pro1.a, vpred.t, n.pl.indef.p, n.loc
5.17	pro1.a, vpred.t, n.p
5.18	prop.h.s, vpred.pst, n.g, pro1.com
5.19	prop.f.s, vpred.pst, n.g, pro2.com

6	**Other functions: modality, experiencer predicates, questions**
6.1	prop.f.sa, can, vpred.t, n.pl.p
6.2	prop.f.a, can, vpred.t, n.pl.p, adv
6.3	pro2.s, cop, adj.pred
6.4	pro1.s, cop, adj.pred
6.5	pro1.s.exp, cop, adj.pred
6.6	pro1.s.exp, cop, adj.pred
6.7	pro1.s.exp, cop.neg, adj.pred
6.8	prop.f.a, wh.p, vpred.t
6.9	anim.gen, predex, prop.loc
6.10	#q, anim.gen, predex, prop.loc
6.11	prop.f.a, vpred.t.pst, n.p

6.12	#q, prop.f.a, vpred.t.pst, n.p
6.13	wh.h.a, vpred.t.pst, n.p
6.14	wh, pro2.a, vpred.t, n.p
6.15	pro1.sa, vpred.neg

7	**Forms of the verb 'come'**
7.1	vpred.inf
7.2	pro1.s, vpred.pst
7.3	pro2.s, vpred.pst
7.4	pro.h.s, vpred.pst
7.5	pro.f.s, vpred.pst
7.6	pro1.pl.s, vpred.pst
7.7	pro2.pl.s, vpred.pst
7.8	pro.h.pl.s, vpred.pst
7.9	pro1.s, vpred
7.10	pro2.s, vpred
7.11	pro.h.s, vpred
7.12	pro.f.s, vpred
7.13	pro1.pl.s, vpred
7.14	pro2.pl.s, vpred
7.15	pro.h.pl.s, vpred
7.16	pro2.s, vpred.pst
7.17	pro2.s, vpred
7.18	pro2.sa, want, vpred
7.19	pro.h.sa, want, vpred
7.20	pro2.s, vpred.neg.pst
7.21	pro2.s, vpred.neg
7.22	pro2.s, vpred
7.23	pro2.pl.s, vpred

8	**Forms of the verb 'eat'**
8.1	vpred.t.inf
8.2	pro1.a, vpred.t.pst
8.3	pro2.a, vpred.t.pst
8.4	pro.h.a, vpred.t.pst
8.5	pro.f.a, vpred.t.pst
8.6	pro1.pl.a, vpred.t.pst
8.7	pro2.pl.a, vpred.t.pst
8.8	pro.h.pl.a, vpred.t.pst
8.9	pro1.a, vpred.t
8.10	pro2.a, vpred.t

8.11	pro.h.a, vpred.t
8.12	pro.f.a, vpred.t
8.13	pro1.pl.a, vpred.t
8.14	pro2.pl.a, vpred.t
8.15	pro.h.pl.a, vpred.t
8.16	pro2.a, vpred.t.pst
8.17	pro2.a, vpred.t
8.18	pro2.sa, want, vpred.t
8.19	pro.h.sa, want, vpred.t
8.20	pro2.a, vpred.neg.t.pst
8.21	pro2.a, vpred.neg.t
8.22	pro2.a, vpred.t
8.23	pro2.pl.a, vpred.t

9	**Forms of the verb 'see' with pronominal verbal objects**
9.1	vpred.t.inf
9.2	prop.h.a, vpred.t.pst, pro1.p
9.3	prop.h.a, vpred.t.pst, pro2.p
9.4	prop.h.a, vpred.t.pst, pro.h.p
9.5	prop.h.a, vpred.t.pst, pro.f.p
9.6	prop.h.a, vpred.t.pst, pro.n.p
9.7	prop.h.a, vpred.t.pst, pro1.pl.p
9.8	prop.h.a, vpred.t.pst, pro2.pl.p
9.9	prop.h.a, vpred.t.pst, pro.h.pl.p
9.10	prop.h.a, vpred.t, pro1.p
9.11	prop.h.a, vpred.t, pro2.p
9.12	prop.h.a, vpred.t, pro.h.p
9.13	prop.h.a, vpred.t, pro.f.p
9.14	prop.h.a, vpred.t, pro.n.p
9.15	prop.h.a, vpred.t, pro1.pl.p
9.16	prop.h.a, vpred.t, pro2.pl.p
9.17	prop.h.a, vpred.t, pro.h.pl.p
9.18	pro2.a, vpred.neg.pst, pro.n
9.19	pro.h.a, vpred.neg.pst, pro.n