

Annotation guidelines

for the Nijmegen Parsed Corpus of Modern Chechen (NPCMC)

Erwin R. Komen

Last update: Wednesday, April 21, 2021 (11:17:00)

Contents

1	About the NPCMC	6.7.2	Post-verbal parenthetical NPs
2	Download the NPCMC	6.7.3	Infinitival NPs (mazdars)
3	Citation	6.7.4	Case marking within NPs
4	General information	6.7.5	Function marking of NPs
5	Chechen syntax phenomena	6.7.6	Proper nouns
5.1	Ergativity	6.8	Verbs
5.2	Leaving out the subject	6.8.1	Tense
5.3	Extrapolation	6.8.2	Mood
6	Annotation guidelines	6.8.3	Modality
6.1	General philosophy	6.9	Adverbs derived from verbs
6.2	Heads of phrases and clauses	6.9.1	Adverbs derived from finite verbs
6.3	Clause-level constituents	6.9.2	Adverbs derived from participles
6.4	Logical connectors	6.9.3	Adverbs by category
6.5	Relative clauses	6.10	Particles
6.5.1	Standard relative clauses	6.10.1	The particle <i>a</i>
6.5.2	Whose name is RCs	6.11	Empty categories
6.5.3	Free relative NPs	6.12	Left dislocation
6.5.4	Free relative locative PPs	6.13	Right dislocation (or parenthetical?)
6.5.5	Extraposed relative clauses	6.14	Topicalizing polar questions
6.5.6	Cleft (like) relative clauses	6.15	Extrapolation (dislocation)
6.6	Adverbial clauses	6.15.1	Extraposed possessors
6.6.1	Purpose clauses	6.16	Extraposed relative clauses
6.7	Noun phrases	7	Issues and beware
6.7.1	Complement NPs	7.1	Double-headed phrases?

1 About the NPCMC

The annotation guidelines for the **Nijmegen Parsed Corpus of Modern Chechen (NPCMC)** (The Chechen Treebank) describe the annotation standard that is being used for a diverse audience. The first audience consists of the current (and possibly future) *annotators* of the corpus, while the second audience consists of the *users* of the corpus. The annotation scheme is based largely on the tradition that has begun with [the Penn historical corpora](#). The current guidelines are meant to form an exhaustive description of all phrase and clause types for the Chechen treebank.

2 Download the NPCMC

The Chechen corpus can be obtained free of charge from the [download page](#). There is no registration wall, which means that anyone can use the corpus at his or her own discretion. The current release is version 1.0 and it consists of XX words. The corpus consists of texts in three different stages of their annotation: (a) bare texts, (b) part-of-speech tagged, (c) dependency-parsed, and (d) constituency-parsed.

While the downloadable texts are in one particular format (the *psdx* format), they can easily be converted to a number of other formats through the freely available "[Cesax](#)" program. Conversion is currently available for: (a) bracketed labelling, (b) FoLiA *xml*, (c) CONLL-X.

3 Citation

Please use the following format when citing the Parsed corpus of modern Chechen, version 1.0 of the May 2014 release.

Komen, Erwin R., Koudrinski, Maksime. 2014.
Parsed corpus of modern Chechen (PCMC). Version 1.0.
<http://erwinkomen.ruhosting.nl/che/crp>

4 General information

Information:

- Chechen syntax phenomena
- [Texts](#)

Annotation:

- Annotation philosophy
- Annotation [process](#)

Treebank team:

- Erwin Komen (PI) (e.komen@let.ru.nl)
- Maksime Koudrinski (kudrinski@gmail.com)

5 Chechen syntax phenomena

A few words about Chechen syntax should be made here, in order to understand the design decisions that have been made more clearly, and in order to define corpus annotation design decisions in the future in a way that does justice to the language.

5.1 Ergativity

Chechen is a morphologically ergative language. Morphological ergativity in its strictest sense implies that the subject of an intransitive sentence (such as "John" in *John walks*) and the object in a transitive sentence (such as "the dog" in *John hits the dog*) are morphologically marked with the same case (we will refer to this case as the "nominative"). The subject of the latter transitive sentence, "John", is marked with a different case, the "ergative" one.

Chechen is almost always morphologically ergative, but there are a few notable exceptions. The first one occurs in the present continuous, which is realised as a periphrastic tense in Chechen, and which consists of a bi-absolutive construction, as demonstrated in (1).

- (1) a. *Muusa buolx biesh vu.*
Musa work doing is
"Musa is working."
b. *Juxa beerasha iza juoxuosh xilla.* [p34-00033:16]
then children-ERG that destroying were
"Then children would be destroying it."

Both the agent *Muusa* 'Musa' in (1a) as well as the theme/patient *buolx* 'work' are in the nominative case. The morphological ergativity is not dependant upon the continuous aspect, since it reappears in the past continuous, as in (1b).

One more situation where morphological ergativity is not indicated is where the agent consists of "we" including the addressee, as in X.

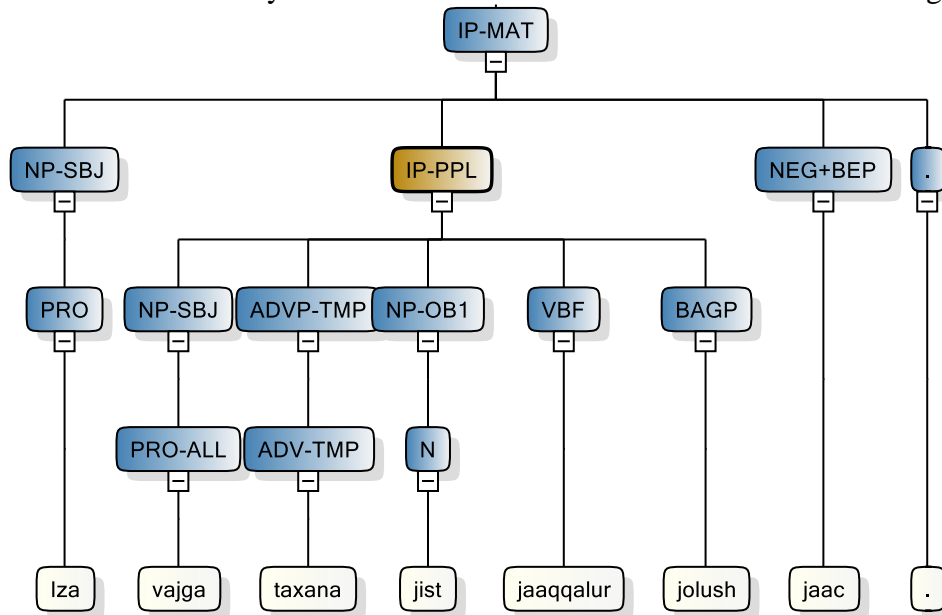
TODO: example of *vaj* as a subject

- (2) a. *Muusa buolx biesh vu.*
 Musa work doing is
 "Musa is working."

There is another possible bi-absolutive construction:

- (3) a. *Iza vajga taxana jist jaaqqalur jolush jaac.* [p34-00002.514]
 3S 1P.INC.ALL today possibility J.draw.POT.FUT J.PTC.PRS J.PRS.NEG
 "We will not be able to resolve it today."

However, it could be that the syntax of this sentence should be like below in the figure



The first absolutive *Iza* 'it' would then be the subject, and the present participle would head a sub clause in which *jaaqqalur jolush* would be the main (compound) verb that has *jist* 'possibility' as object.

5.2 Leaving out the subject

In many instances the subject is not obligatory in Chechen, and sometimes it is not even possible to determine what exactly the subject is.

5.3 Extraposition

Chechen offers a number of dislocations that we will refer to as extrapositions.

TODO: treat infinitival complement extraposition

- c. *Etot chelovjek moj otjec.*
 that person my father
 'That person is my father.'

- d. *Mylxa du vajn literaturan kyerta gerz? Muott bu-q.* [p34-00002:117,120]
 which is our literature's main weapon language is
 'What is the main weapon of our literature? It is language.'

6 Annotation guidelines

6.1 General philosophy

The syntactically parsed corpus of Chechen attempts to reconcile a number of requirements that can never all be met at the same time:

- Theory neutrality
- Compactness
- Accuracy
- Consistency

6.2 Heads of phrases and clauses

Each phrase, such as a Noun Phrase, an Adverbial Phrase, needs to contain a head. Each clause too needs to be properly headed. Possible heads for different phrase and clause types are the following:

Phrase type	Possible heads
ADJP	ADJ, *+ADJ, CONJP, NUM, D
ADVP	ADV, *+ADV, ADJ, CONJP, N, P, VAGP, VANP, RP, Q, CONJ
CONJP	CONJ, CONJP, VB*, ADJ, BE*, NUMP, P, *
CP	C, IP, IP-*
INTJP	INTJ, P, *+P
IP	VB*, *+VB*, VA*, *+VA*, BA*, *+BA*, BE*, *+BE*, AX*, *+AX*, *OB1, CP, WH
NP	N, NS, NPR, NUM, NUMS, Q, QP, NP, CONJP, CP, D, FW, PRO\$
NUMP	NUM, N
PP	P, *+P, ADV
QP	Q, NUM, IP*
WADJP	ADJ
WADVP	WADV*, WPRO*, W*, *
WNP	WPRO
WPP	P

Heads preferences are from left to right in the table above.

6.3 Clause-level constituents

The kinds of clause-level constituents that are allowed for any type of IP, be it IP-MAT, IP-SUB, IP-PPL or another type, are restricted. Here is a list of allowables:

Allowable	Example
Noun phrases	NP-ADV, NP-SBJ, NP-OB1, NP-OB1, NP-POS
Free relatives	CP-FRL, CP-FRL-SBJ
Verbal infinitives	IP-INF-NOM, IP-INF-OB2
PPs	PP-LOC, PP-OB1
Finite verbs	VBD-R, VBD-N, VBF, VBP
Non-finite verb forms	
Adverbs (!)	ADV
Adverbial phrases	ADVP-TMP
Adjectival phrases	ADJP

6.4 Logical connectors

English has logical connectors that are sometimes morphologically made up of two parts (e.g. because = by + cause, therefore = there + for). The same goes for Chechen. Here is a list of logical connectors and how they should be viewed on the phrase-level.

Connector	Meaning	POS + Phrase
<i>cundeela</i>	therefore	(PP (D+P cundeela))
<i>hunda aelcha</i>	because	(CP-ADV (IP-SUB (WADV hunda) (VBD+P aelcha)))
<i>delahw a</i>	however	(CP-ADV (IP-SUB (BEP+P Delahw) (FP a)))
<i>nagahw sanna</i>	if	(CONJP (CONJ Nagahw) (ADJP (ADJ sanna)))
<i>aelcha a</i>	differently	(IP-SUB (VBD+P aelcha) (FP a))
<i>ishtta dwaq qin [a]</i>	and so on	(ADVP (ADJP (ADJ i)) (ADV dw.) (ADV q.))
<i>deera</i>	of course	(ADVP (FP deera))
<i>masalaa</i>	for example	(ADVP (ADV masalaa))
<i>cq'a hwalxa</i>	only when	(ADVP (ADVP (ADV Cq'a)) (ADV-TMP hwalxa))

6.5 Relative clauses

6.5.1 Standard relative clauses

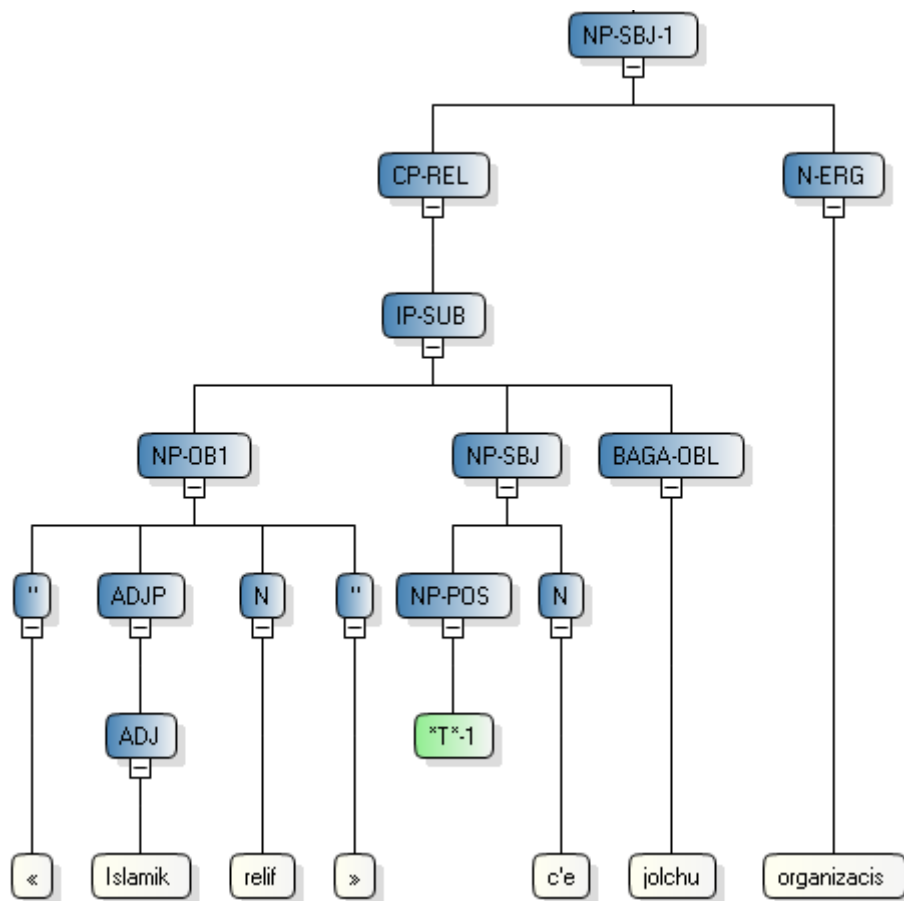
Relative clauses preceding a head-noun must have a number of elements: a participle form of a verb headed by a subordinate IP, which has a CP-REL as parent and an NP-head as sister.

```
(NP-OBI
  (CP-REL
    (IP-SUB
      (ADV laqahw)
      (VANA bilgalbaexna)
    )
  )
  (NS bielxash)
)
```

The example above has *bielxash* ‘works’ as NP-head, and the CP-REL is sister of it. Under the CP-REL there is the required IP-SUB, and this subordinate clause here consists only of two elements: an adverb *laqahw* and a past participle *bilgalbaexna* ‘determined/defined’.

6.5.2 Whose name is RCs

Example: the organization, whose name is *Islamic relief*:



6.5.3 Free relative NPs

A free relative NP consists of a nominalized form of a participle and this verb may have any number of arguments. The free relative is generally coded as CP-FRL or when it is declined CP-FRL-SBJ or CP-FRL-OB1 and so forth.

```
(CP-FRL-OB2
 (IP-SUB
  (NP-SBJ (N Buoxam)
   (AXNA-DAT xillachaarna)))
```

The free relative may, in turn, be placed into an NP, in which case that NP gets the functional label:

```
(NP-SBJ-1
 (CP-FRL
  (IP-SUB
   (NP-SBJ *T*-1)
   (PP-LOC (ADV t'iehw))
   (VBD-NOM bisinarsh))))
```

6.5.4 Free relative locative PPs

Chechen has extended the free relative NP construction to one that can carry a number of locative case endings. These endings include: the (shortened form of the) allative, the locative, the source and the through cases.

Free relative with shortened allative:

Beerash dolchu dwaahwodura iza 'he would run to where the children were'

Free relative 'through' PP:

```
(NP-LOC
  (CP-FRL
    (IP-SUB
      (NP-OB1 (NP-POS (N-GEN xin)) (NS ajmanash)) 'at the water playgrounds'
      (BAGA-LOC jolchuxula))))
```

6.5.5 Extraposed relative clauses

6.5.6 Cleft (like) relative clauses

6.6 Adverbial clauses

There are a number of adverbial affixes that can be attached to different forms of the verb. Many times these affixes turn the verb plus its associated clause into an adverbial clause. The general principle, then, is to end up with a CP-ADV.

6.6.1 Purpose clauses

An example of a purpose clause which centers around verbal Mazdar + *-hwama*:

```
(CP-ADV
  (IP-INF
    (CONJP (CONJ dyhhwaldwaa))
    (NP-OB2 (NP-POS (N-GEN ghaalin)) (NS-DAT marshrutashna))
    (NP-OB1 (N qaachuo))
    (VBI+P jarhwama))))
```

Note that the verb is infinitive + P, where the *hwama* is considered to be a special kind of postposition.

6.7 Noun phrases

6.7.1 Complement NPs

Complex noun phrases can have a head Noun that is modified by a complementary NP, such as the (shortened) dative case NP *husama* 'house' in this clause:

```
(IP-MAT
  (ADVP (ADV T'aaqqa))
  (NP-SRC (NS-MAT lajx))
  (NP-OB1 (ADJP (ADJ joqqa)) (NP-CPL (N-DAT husama)) (- --) (N tajpa))
  (VAGP jiesh)
  (NS-ERG naaxa)
  (. .))
```

Translation: Then people would make a large kind of house from snow.

6.7.2 Post-verbal parenthetical NPs

An NP can have a parenthetical, a further specification of itself. This parenthetical can be dislocated to the postverbal position, as *professor* in this clause:

```
(IP-MAT
```

```

(NP-SBJ
 (CP-FRL
  (IP-SUB
   (NP-OB2 (PRO-DAT Suuna))
   (ADVP-TMP (ADV-TMP hetahw))
   (ADVP (ADV ullie))
   (VANA-NOM vaellarg))))
 (NP-OB1 (NUMP (NUM cwa)) (NPR Arsaxanov) (NP-PRN *T*-1))
 (BED vara)
 (, ,)
 (NP-PRN-1 (N professor))
 (. .))

```

6.7.3 Infinitival NPs (mazdars)

The un-tensed verbal noun (which is sometimes referred to as the “mazdar”) may head a noun phrase. If this happens, the NP is labelled as “IP-INF-...” It *must* have a functional label extension (IP-INF-LOC, IP-INF-SBJ) or if there is no valid functional label, it should carry a case label extension (IP-INF-NOM, IP-INF-DAT) in order to distinguish it from true infinitival clauses, which are headed by IP-INF without any extension.

```

(IP-INF-LOC
 (NP-SBJ (N-ERG Ministra))
 (VB+N-LOC diicariehw))

```

6.7.4 Case marking within NPs

Nouns and pronouns are marked for *case* by extending the part-of-speech tag (N, NS, PRO, D, NPR) with a hyphen and a [case tag](#):

	Case		Case		Case
-NOM	Nominative (Absolutive)	-MAT	Material (Lative)	-ALL	Allative (goal)
-ERG	Ergative	-GEN	Genitive	-DIR	Direction
-DAT	Dative (and Short variant)	-SRC	Source (Ablative)	-THR	Through/via
-INS	Instrumental	-CMP	Comparative	-LOC	Locative

The **nominative** -NOM marker is normally not written, except where confusion would otherwise result. The Noun Phrases heading nouns are almost *never* marked with a case label, since the case of an NP is derivable by looking at the case of the NP's head.

Pronouns in the genitive case, which are **possessive pronouns**, are not marked with a -GEN extension, but are marked as PRO\$, in compliance with the annotation tradition.

Adjectives inside an NP are not marked for case when they are in the nominative. When the NP is in a different case, the adjectives are marked as **oblique**: -OBL.

6.7.5 Function marking of NPs

Every NP must have a functional label in principle, unless the NP is part of a conjunction phrase; in that case its function is determined by the functional label of the NP on top of the conjunctions. The NP *beecan* 'of grass' does not have a functional label, for instance, but the overriding NP is marked as NP-POS:


```
(NP-POS (NP (N-GEN beecan))
  (CONJ a)
  (, ,)
  (CONJP
    (NP (NS-GEN hwaezhkiin))
    (CONJ a))
  (, ,)
  (CONJP
    (NP (D-OBL qiechu) (NS-GEN jaltiin))))
```

Here is a list of functional markers with examples:

	Function	Example
-ADR	Addressee	<i>ghullaq xir du</i> [NP-ADR <i>hwuuna</i>] 'something will come out of this, you know'
-ADV	Adverbial	
-ATT	Attributive	[NP-ATT <i>massuo a tajpa</i>] <i>passazhirash</i> 'all types of passengers'
-COM	Comitative	
-CPL	Complement	
-LFD	Left dislocation	
-OB	PP object	
-OB1	Direct object	
-OB2	non-direct object	
-OB3	locative object	
-POS	Possessive	
-SBJ	Subject	
-TMP	Temporal	<i>ocu diinahw</i> 'on that day'
-VOC	Vocative	<i>Muusa!</i> 'Musa!'

6.7.6 Proper nouns

Uninflected names and surnames are marked **NPR**. Many Russian surnames end in *-aev*, e.g. *Timaev*, *Arsanukaev*. These are inflections within the Russian language, but they are not within Chechen. This can be seen from the fact that Chechen inflection may occur *after* the Russian *-aev* ending:

<i>aliev-na</i>	NPR-DAT	<i>elzhurkaev-n</i>	NPR-GEN
<i>dudaev-ich-na</i>	NPR-DAT	<i>gacaev-n</i>	NPR-GEN
<i>kazancev-na</i>	NPR-DAT	<i>mamakajev-n</i>	NPR-GEN
<i>abdulaev-s</i>	NPR-ERG	<i>munaev-</i>	NPR-GEN
<i>bakajev-s</i>	NPR-ERG	<i>walaroj-ev-lahw</i>	NPR-LOC
<i>seleznev-s</i>	NPR-ERG	<i>dadaev-x</i>	NPR-MAT
<i>tashalajev-s</i>	NPR-ERG	<i>eesambajev-x</i>	NPR-MAT
<i>visaev-s</i>	NPR-ERG		

6.8 Verbs

6.8.1 Tense

Chechen marks tense, aspect, mood and agreement on the verb. The annotation recovers the tenses in the following way:

Tense	Chechen	Marker	Example
Present (generic)	<i>karara xaan</i>	VBP, BEP	<i>dyeshu</i> 'read(s)'
Present continuous		VAGP + AXP, VAGP + BEP	<i>dyeshush du</i> 'is reading'
Past generic	<i>jaxana xaan</i>	VBD-N, BED	<i>dieshna</i> '(has) read'
Past witnessed	<i>gush-jaxana xaan</i>	VBD-R	<i>diishira</i> 'read'
Past recent	<i>hinc-txana xaan</i>	VBD-REC	<i>diishi</i> 'just read'
Past remote	<i>hwalx-jaxana xaan</i>	VBD-REM	<i>dieshniera</i> 'has/had read'
Past imperfective	<i>ixna xaan</i>	VBDI	<i>dyeshura</i> 'was reading'
Past continuous		VAGP + AXD-N, VAGP + BED	<i>dyeshush xilla/dara</i> 'was reading'
Future	<i>xila mega xaan</i>	VBF	<i>dyeshur</i> 'will (probably) read'
Future combined	<i>(xin) jooghu xaan</i>	VBF + BEP	<i>dyeshur du</i> 'will read'
Future past		VBF + AXD-N, VBF + BED	<i>dyeshur dara</i> 'would read'

6.8.2 Mood

Chechen morphologically distinguishes between different moods. Default is the declarative mood (no additional markings). Moods that are marked are: subjunctive, desiderative, imperative, conditional.

Chechen has a rich set of mild, strong and polite **imperatives**. Most of the distinctions need to be gleaned from the verb's morphology, and are not coded in the part-of-speech tag, nor in the function of the clause. Imperative clauses—of whatever degree—are marked as **IP-IMV**. The different possible imperatives and desideratives are these:

Name	Chechen	Tag	Singular	Plural
	<i>t'eduožharan sattam</i>		<i>cwallin</i>	<i>duqallin</i>
Imperative	<i>cwalxie forma</i>	VBI	<i>diesha</i> 'read!'	
Strong imperative	<i>hwemboocu forma</i>	VBI	<i>dieshal</i>	
	<i>t'aehwatattaran forma</i>	VBI	<i>dieshalahw</i>	<i>dieshalash</i>
Polite imperative	<i>ghillaqie forma</i>	VBI	<i>dieshahwa</i>	<i>dieshiisha</i>
	<i>diexaran forma</i>	VBI	<i>dieshahwaara</i>	
Desiderative	<i>t'dillaran-t'eduožharan forma</i>	VBI	<i>dyeshiila</i>	

6.8.2.1 Subjunctive mood

TODO: speak about **subjunctive**

6.8.2.2 Conditional mood

TODO: explain **conditionals**

6.8.3 Modality

The verb *dieza* 'must' could be seen as a modal auxiliary, since it always combines with another more lexical verb (do, be). For example:

(IP-MAT

(NP-SBJ (D I) (N programma))

```
(NP-OBI (NP-POS (N-GEN pachwalqan)) (N programma))
(AX xila)
(MDP jieza)
(. .))
```

Recognition of *dieza* as modal is still on the way.

6.9 Adverbs derived from verbs

NM1:229-232 has an extensive section on how adverbs are derived from all kinds of word categories. There are two categories of adverbs that are derived from verbs: one from finite verbs and one from participles.

6.9.1 Adverbs derived from finite verbs

6.9.2 Adverbs derived from participles

6.9.3 Adverbs by category

Here is a list of adverbs divided into categories (semantic).

Category		
interrogative	<i>maca? micha? michara?</i>	when? where? from where?
negative	<i>cq'a a, cwanwaa a, humanna a</i>	never, nowhere, nothing
without distinction (any)	<i>muxxa a, maccax, macca a</i>	however, anytime
distinguishing (every)	<i>guttar a, daa'im a, massanwaa a</i>	always, always, everywhere
initial	<i>xezzashiehw, xyettushiehw</i>	as soon as/when heard, asked
demonstrative	<i>ishtta, ciga, quzahw</i>	thus, there, here
emphatic negative	<i>aallanie a, xa''anie a, la''anie a</i>	without at all saying / knowing / wanting
temporal	<i>xezzashiehw, oollushiehw¹</i>	at the moment of hearing / saying

The emphatic negatives are annotated this way: (ADVP-NEG (ADV xa''anie) (FP a))

6.10 Particles

6.10.1 The particle *a*

The particle *a* can function as a conjunction or as a kind of (additive) focus marker (translatable as 'also' or 'too' in English). When *a* is clearly a conjunction, its part-of-speech tag is "CONJ". In all other cases it receives the part-of-speech tag "FP" (focus particle), even though it may not have the traditional meaning of a (constituent/narrow) focus particle.

6.11 Empty categories

We try to keep the empty categories to a minimum, yet we have felt a need to use them in a limited number of cases.

Morpheme Part Here: *MPH*-n

¹ NM adds: *c' ehhwana xylush a, oolush a du ysh* 'these are occurring or being said suddenly'.

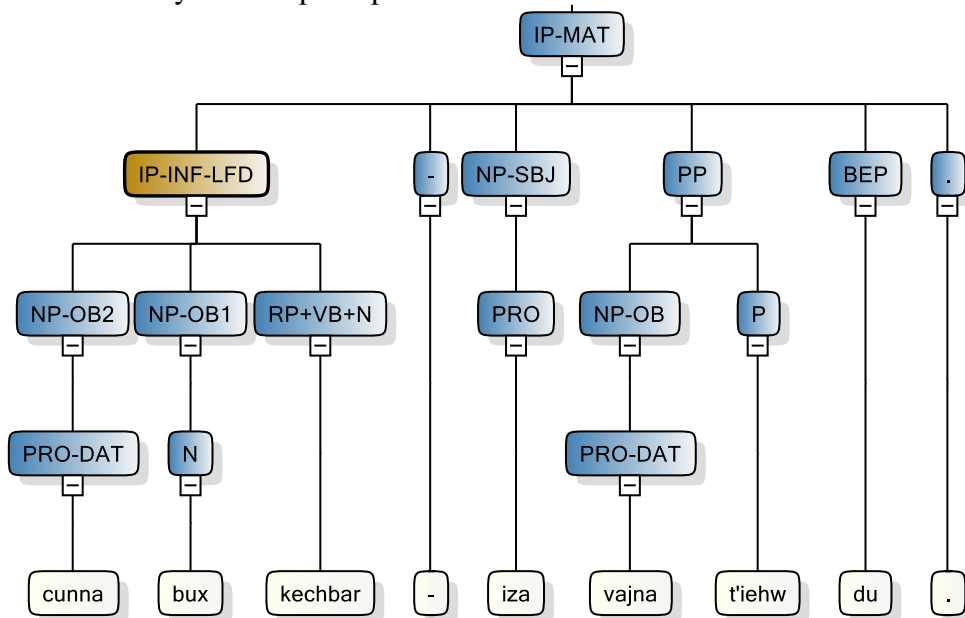
Chechen sometimes combines morphemes into one word that are of different categories, and these categories are needed for the syntactic interpretation. The morpheme *-chuohw* 'inside', for instance, can be added to the past *-na* tense of the verb:

```
(PP-LOC
 (CP-ADV
  (IP-SUB
   (NP-SBJ (PRO Iza))
   (AXD+P-2 xillachuohw))) (P *MPH*-2))
```

It is clear that the morpheme *-chuohw* is functioning as a postposition here, so that *xillachuohw* is labelled as *AXD+P*. A *P* node is added with the empty category **MPH*-n* in order to indicate where the *P* is to be interpreted.

6.12 Left dislocation

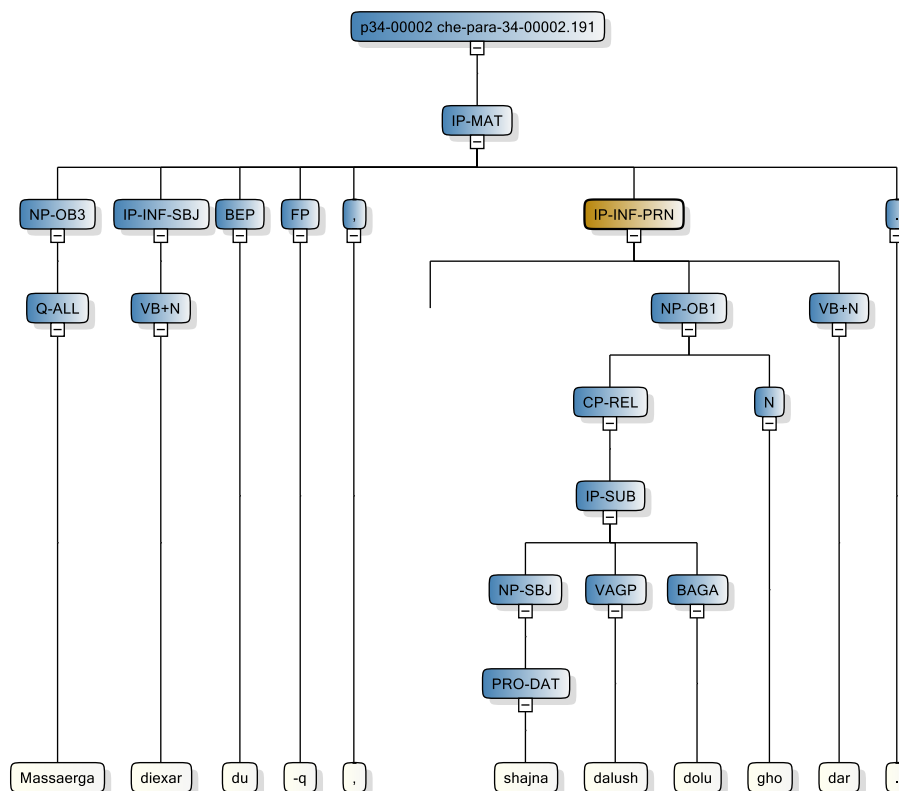
Chechen sometimes uses left dislocations of a complex NP that is then summarized in the main clause by a resumptive pronoun:



The left dislocation receives the functional category *LFD*.

6.13 Right dislocation (or parenthetical?)

I'm not sure whether this should be called right dislocation or parenthetical, but here's an example of the right dislocated *IP-INF-PRN*:



In fact, I'm not at all sure that I understand what the syntactic role of the *IP-INF-PRN* is in the sentence. Syntactically, one would expect a *CP-THT*:

There's a request for everyone, *that* they should help the best they can.

But the *gho dar* heading the phrase really turns it into a *mazdar*, an *IP-INF-SBJ* or *IP-INF-OB1*. Yet the main verb here is *du* 'is', which takes *diexar* 'request' as subject, and can do no more than add a goal in the form of *massaerga* 'all'. So there are no roles left for the *mazdar*.

6.14 Topicalizing polar questions

Sometimes a polar question is used in spoken discourse to set a topic. And sometimes such a construction is clearly marking this topic as part of a larger main clause. Example:

```
(IP-MAT
  (IP-MAT-TOP
    (NP-OB1 (CP-REL (IP-SUB
      (IP-PPL
        (CP-ADV (IP-SUB (NP-SBJ (N Beer)) (RP+VBP+P dwaaduolalusshiehw)))
        (NP-SBJ (PRO-ERG cuo))
        (NP-OB1 (CONJP (NP (" «) (N naana) (" »)) (, ,)) (" «) (N daada) (" »))
        (VANP aella))
        (RP+VANP dwaabuoliina)
        (BAGA bolu))) (N muott))
      (BEP+WH buj))
    (, ,)
  (IP-INF
    (NP-OB1 (PRO i))
    (NP-LOC (N-LOC ishkoliehw))
    (VB biiciita))
  (VBP bieza)
  (. .))
```

The back translation made by a Chechen native speaker has this as:

We have to make a child speak at school the language in which he said ‘mother’, ‘father’ when he first started speaking.

6.15 Extraposition (dislocation)

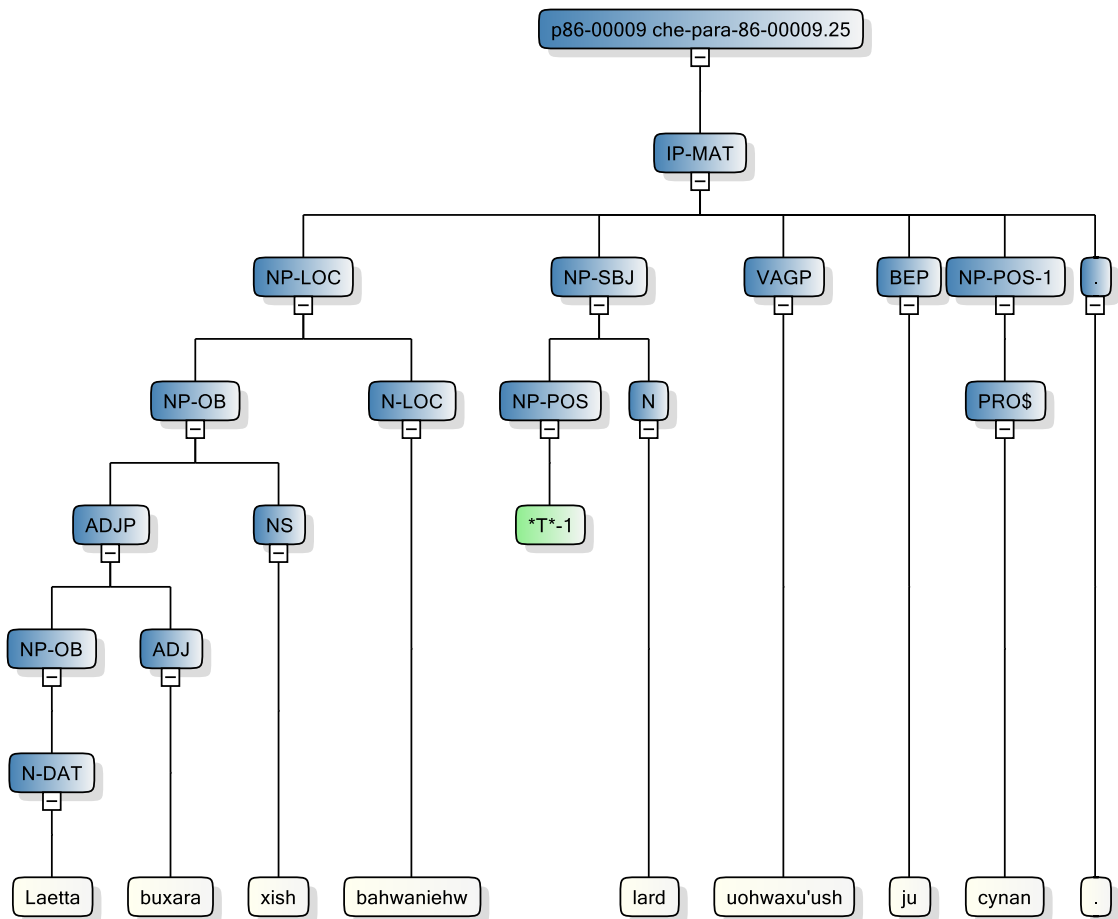
The Chechen language allows (and uses) extraposition (dislocation) in a number of situations that may seem odd to an outsider. This section discusses the different extraposition situations.

6.15.1 Extraposed possessors

Chechen clauses may contain a subject with a modifying possessor, where this possessor is dislocated to the clause-final position. As all kinds of dislocation, the syntactic position of the possessor within the subject NP should be indicated by an *ICH*-n element. The dislocated possessive NP itself must have a numeric extension of its label, such as NP-POS-n, where the number coindexes with the *ICH* element’s number.

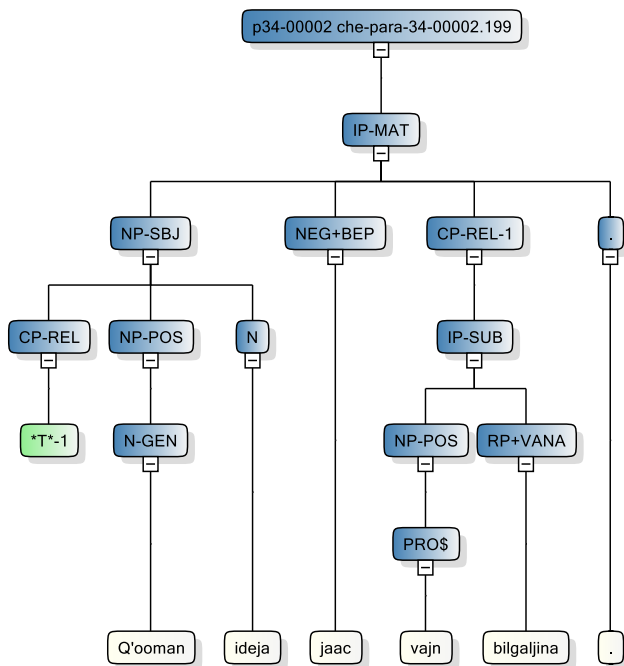
```
(IP-MAT
  (NP-TMP (D-OBL Qu) (NS-LOC dienoskahn))
  (NP-COM (N-INS zhurnalistashca))
  (NP-SBJ (NP-POS *ICH*-1) (N cwaenaqietar))
  (AXD-R xilira)
  (NP-POS-1
    (NP-POS
      (NP-POS (NS-GEN Noxchiin))
      (N-GEN Respublikan))
      (N-GEN Praviteljstvon))
      (N-GEN vice-premjeran)
    (CONJP
      (, ,) (NP-POS (NP-POS (NS-GEN finansiin)) (N-GEN ministran)
        (NP-PRN (NP-POS (NPR-GEN Isaev))
          (NPR-GEN Weelin))))))
  (. .))
```

Another example, as tree:



6.16 Extraposed relative clauses

They don't occur very often, but here are some examples.



Translation: 'We have not defined a national idea.'

7 Issues and bewares

7.1 Double-headed phrases?

Some phrases appear to be double headed, as in the example here:

```
(ADVP
  (ADV Qin)
  (ADV21 Cq'a) (ADV22 A)
)
```

However, the first adverb *qin* ‘again’ in fact modifies the second adverb *cq'a a* ‘never’ (resulting in the meaning ‘never again’). All this is to say that in such ambiguous cases, where two (or perhaps even more) potential head constituents occur adjacent, the last one of them is to be taken as the actual head, since Chechen is a predominantly head-final language. In the annotation of the corpus, we have not used indicators of headedness.

7.2 Information structure

This section contains examples of sentences that are interesting from the point of view of information structure, e.g. constituent focus.

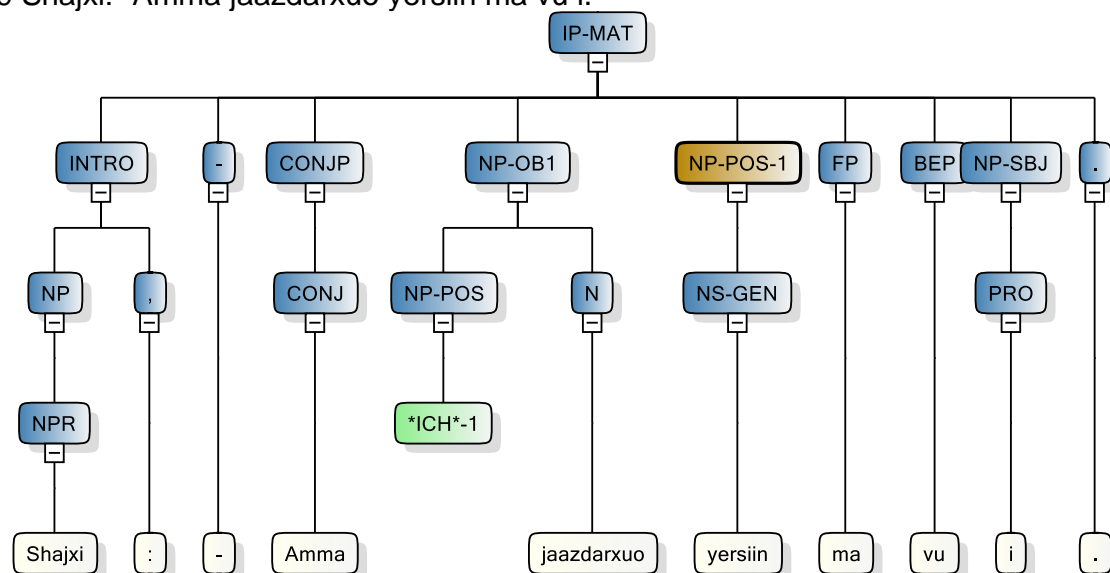
7.2.1 Constituent focus on a possessive modifier

If a noun phrase has a possessive that modifies that NP, this possessive can be focused by moving it out of its default position (i.e. before the head noun) into the position preceding the finite verb. See the discourse in p34-00002:547-549:

547 Muusa, Gogolj hweenan jaazdarxuo vu?

548 Muusa: -Ukra'inec vu.

549 Shajxi: -Amma jaazdarxuo yersiin ma vu i.



Note that the possessive modifier *yersiin* ‘Russian’ is moved from its location under the direct object (the NP-OB1) into the slot directly preceding the finite verb *vu* ‘is’. (The assertive particle *ma* ‘prt’ must still be there too.)

The response without focus would have been:

Amma [SBJ *iza*] [OB1 *yersiin jaazdarxuo*] *vu*
'But he is a Russian author.'

I'm not sure how this would be realized intonationally, that would be interesting to research.
Note too that the *yersiin* 'Russian' is the constituent response to the question *hweenan*
'whose' in sentence 547.