An RRG Parser for Middle Constructions in Faroese

SIGMUND VESTERGAARD – B00108126

Institute of Technology Blanchardstown

In this article we describe a parser for Middle Constructions in Faroese built on Role and Reference Grammar (RRG). Before we describe parser we describe morphology and syntax in Faroese, and middle constructions in particular. We analyse middle constructions in terms of the RRG framework and build up a feature based grammar for use with NLTK's parsing functions. We show that by building up the grammar carefully we are able to use NLTK's parsing functions to parse middle constructions in Faroese, particularly middle constructions with passive reading.

Keywords: RRG, Faroese, Middle Constructions, Parsing

1. Introduction

In this paper we describe the architecture and implementation of a Role and Reference Grammar (RRG) parser for middle constructions in Faroese. The parser will be able to take a so-called "middle" sentence and tell us whether it is grammatical or not, and if it is, show an RRG tree for it.

In the section 2 below we will briefly describe the morphology and syntax of Faroese in general. The description is based on (Thráinsson et al. 2012). We will then move on to describe middle constructions in Faroese in the second section. In Section 3 we will also analyse middle constructions in terms of Role and Reference Grammar (RRG). After this we will describe the design of the lexicon in Section 4 and the parser itself in Section 5. We will finish off in Sections 6 by testing the parser on some test sentences.

The report and the code are found in the Git repository at

https://github.com/sigmundv/programming-for-nlp

2. Faroese Morphology and Syntax

Faroese is derived from the Norse language of the primarily Norwegian settlers who moved to the islands in the ninth century. Due to a lack of Viking Age and medieval sources little is known about the development of Faroese into the language we know today. Nothing of substance was written in Faroese until the 1770s, by which time most the features of the modern language must have developed.

2.1. Nominal categories

Like Icelandic, Faroese has three grammatical genders; masculine, feminine, and neuter. Faroese nouns inflect for gender, but also number (singular/plural)

and case (nominative/accusative/dative/genitive). The inflectional endings vary according to gender. Definiteness of nouns is indicated by a suffixed article.

Just like nouns, adjectives also inflect for number, case, and gender. Additionally they also inflect for degree (positive/comparative/superlative). Adjectives typically have two forms of inflections - strong or weak - depending on the definiteness of the noun phrase they are a part of.

Finally, articles, pronouns, the cardinal numbers 1 to 3, and the ordinal numbers, inflect for number, case, and gender.

Faroese noun phrases show extensive agreement, including number agreement between nouns and the adjectives that modify them. This holds for both attributive and predictive adjectives. We illustrate this using the noun *drongur* 'boy' with the adjective *klókur* 'smart' (Thráinsson et al. 2012: 61):

(1) (a) ein klókur drongur klókir drongir
a smart boy.SG smart boy.PL
'a smart boy' 'smart boys'

(b) drongurin er klókur dreingirnir eru klókir
boy-DET is smart boy-DET,PL is.PL smart.PL
'the boy is smart' 'the boys are smart'

The gender of nouns is reflected in the different forms of the personal pronouns used to refer to them and the gender of adjectives and articles used to modify said nouns. We illustrate this with the following example:

(2) Hetta er ein klókur drongur
this is a.M smart.M boy.M
Hann. er klókur
he.M is smart.M

Hetta er ein klók genta
this is a.F smart.F girl.F
Hon er klók
she.F is smart.F

Hetta er eitt klókt barn
this is a.N smart.N child.N
Tað er klókt
it.N is smart.N

Old Norse and older Faroes had four morphologically distinctive cases - nominative, accusative, dative, and genitive - but only the three first are productive in modern spoken Faroese (modern *written* Faroese still retains the genitive to a degree). We illustrate this with a small example:

(3) Gentan svav
girl-DET.NOM sleep.PST
'The girl slept'

Eg sá gentuna
I see.PST girl-DET.ACC

'I saw the girl'

Hetta er hundurin hjá gentuni
This is dog-DET with girl-DET.DAT

'This is the girl's dog'

Despite this a genitive form can be produced for nouns and personal pronouns, but less so for adjectives. The genitive form of personal pronouns is widely used while the genitive form of many nouns is found in fixed expressions and as the first part of certain compounds, but it is uncertain whether speakers intuitively interpret these forms as genitive (Thráinsson et al. 2012: 62).

Instead of genitive, modern spoken Faroese prefers prepositional constructions involving a dative form of the noun as illustrated in these examples:

(4) Her eru húsini hjá einum ríkum manni
here is.PL house-DET.PL with a.DAT rich.DAT man.DAT
'Here is a rich man's house/home'

Kettlingurin hjá kettuni hjá mær er vakur
kitten-DET with cat-DET.DAT with I.DAT is beautiful

'My cat's kitten is beautiful'

Above we saw examples with the preposition *hjá* 'with', but Faroese speakers also use other prepositions with the dative, depending on the semantic function. We'll look at some examples (Thráinsson et al. 2012):

(5) takið á húsinum motorurin í bilinum
roof-DET.NOM.N on house-DET.DAT.N motor-DET.NOM.M in car-DET.DAT.M
'the roof of the house' 'the car's engine'

abbi at dreinginum aldurin á kirkjuni
grandfather to boy-DET.M.DAT age-DETon church-DET.DAT

'the boy's grandfather' 'the age of the church'

halin á kúnni tenninar í hundinum
tail-DETon cow-DET.DAT tooth-DET.PL in dog-DET.DAT

’the cow’s tail’ ’the dog’s teeth’

høvdið á mær eyguni í honum
head-DETon I.DAT eyes-DET in he.DAT

’my head’ ’his eyes’

With nouns denoting family relationships an accusative form is normally used instead of genitive or a prepositional phrase as illustrated here:

(6) pápi dreingin mamma gentuna beiggi Jógvan
father boy-DET.ACC mum girl-DET.ACC brother Jógvan.ACC
’the boy’s father’ ’the girl’s mother’ ’Jógvan’s brother’

As mentioned above, adjectives can be grouped into two categories: *strong* adjectives and *weak* adjectives. Which category they belong to depends on the definiteness of the noun phrase they form a part of. The general rule is that the adjective takes the weak form if the noun phrase is definite, and the strong form if the noun phrase is indefinite. We illustrate this with two examples (Thráinsson et al. 2012: 65):

(7) (a) Hetta er ein stórur bilur og ein lítil
this is a big.NOM.SG.M car.NOM.SG.M and a small.NOM.SG.F
bók.
book.NOM.SG.F
’This is a big car and a small book’

(b) Hetta er tann stóri bilurin og tann
this is the big.NOM.SG.M car-DET.NOM.SG.M and the
lítla bókin.
small-DET.NOM.SG.F book
’This is the big car and the small book’

Finally, most adjectives can be inflected for degree by adding the suffixes *-(a)r* and *-(a)st* in comparative and superlative, respectively. Indeclinable adjectives express difference in degree by using the auxiliary verbs *meiri* ’more’ (comparative) and *mest* ’most’ (superlative). We’ll finish by showing a couple of examples of this:

(8) (a) gulur gul-a-ri gul-ast-ur
yellow yellower yellowest

(b) hóskandi meiri hóskandi mest hóskandi
appropriate more appropriate most appropriate

2.2. Verbal categories

Faroese verbs are inflected by person, number, and tense, with the two following characteristics with respec to person inflection (Thráinsson et al. 2012: 67):

1. Faroese verbs do not show any person distinctions in the plural and regular (weak).
2. Faroese verbs do not show any person distinctions neither in the singular nor in the past tense.

This can be illustrated with the following example:

(9) (a) eg kalli tú kallar hann/hon/tað kallar
I call.1SG.PRS you call.2SG.PRS he/she/it call.3SG.PRS
'I call' 'you call' 'he/she/it calls'

(b) vit kalla tit kalla teir/tær/tey kalla
we call.1PL.PRS you.PL call.2PL.PRS they.M/F/N call.3PL.PRS
'we call' 'you call' 'they call'

(c) eg kallaði tú kallaði hann/hon/tað kallaði
I call.1SG.PST you call.2SG.PST he/she/it call.3SG.PST
'I called' 'you called' 'he/she/it called'

Faroese has two distinct imperative forms, plural and singular, as illustrated here:

(10) Gev/gevið hesum manninum gætur!
give.SG/PL this man-DET.3SGDAT attention
'Give attention to this man!'

The singular *gev* would be used if addressing one person, and the plural *gevið* if more than one person is being addressed. There was no distinctive plural imperative in Old Norse, where the 2PL indicative had this role, but in Faroese there is a distinction between the default finite forms and the imperative forms. We illustrate this below, where the non-imperative forms are referred to as *indicative*, although it is uncertain that one can speak of indicative in Faroese, because there is no productive contrasting subjunctive in Faroese (Thráinsson et al. 2012: 67-68):

(11) (a) Tú fert til hús.
you.2SG go.2SG.PRS.IND to house
'You go home'

(b) Far til hús!
go.2SG.IMP to house
'Go home!'

(c) Tit fara til hús.
you.2PL go.2PL.PRS.IND to house
'You go home'

(d) Farið til hús!
go.2PL.IMP to house
'Go home!'

We mentioned above that the subjunctive is not productive in Faroese anymore, and it should be added that only a few relic forms exist in main clauses in relatively fixed expressions and in religious language. And where they exist they almost exclusively express *optative modality*, as illustrated below (Thráinsson et al. 2012: 68):

(12) (a) Jesus fylgir tær
Jesus follow.3SG.IND you
'Jesus is with you'

(b) Jesus fylgi tær
Jesus follow.3SG.SUBJ you
'Jesus be with you'

(c) Gud signar Føroyar
God bless.3PL.IND Faroes
'God blesses the Faroes'

(d) Gud signi Føroyar
God bless.3PL.SUBJ Faroes
'God bless the Faroes'

Examples of other, relatively fixed optative forms, which aren't religious expressions, are:

(13) (a) Hann leingi livi!
He long live.3SG.SUBJ
'Long live he!'

(b) Gævi at tað skjótt varð heystfrí!
Give.3SG.PST.SUBJ that it soon become.SG.PST autumn break
'I wish we had autumn break soon!'

(c) Hevði tað nú bara gingist henni væl.
Have.3SG.PST.SUBJ it now just go.SG.PST her well
'I wish things would go well for her'

Of these the first one uses the present subjunctive while the two others use the past subjunctive (or what was the past subjunctive in older Faroese).

Typically the passive is formed with the auxiliary verbs *verða* 'be, become' and *blíva* 'be, become'. The participle agree in case, gender, and number with a nominative subject, and the agent is more frequently left out than in English. If the agent is included, it is with the auxiliary *av* 'by', which takes a dative form. We illustrate this with some examples (Thráinsson et al. 2012: 69):

(14) (a) Hann kysti hana
he.3SG.NOM kiss.SG.PST her.3SGACC
'He kissed her'

(b) Hon varð/bleiv kyst (av honum)
she.3SG.NOM be.SG.PST.AUX kiss.NOM.SG.F.PST.PTCP (by him.DAT)
'She was kissed by him'

(15) (a) Hon kysti teir
she.3SGNOM kiss.SG.PST.IND them.ACC.PL.M
'She kissed them'

(b) Teir vórðu/blivu kystir (av henni)
they.3PL.NOM.M be.PL.PST.AUX kiss.NOM.PL.M (by her.DAT)
'They were kissed by her'

(16) (a) Teir smurdu hann av
they smear he.3SG.ACC off
'They beat him up'

(b) Hann varð/bleiv avsmurdur
he.3SG.NOM be offsmear.NOM.SG.M.PST.PTCP
'He was beaten up'

In Faroese it is frequently possible to form so-called *-st*-forms, or *middle* forms, by adding the suffix *-st* to various inflectional forms of the verb. The meaning of the middle forms varies widely in Faroese, but the usages most frequently mentioned in discussions of the middle forms are reflexive, reciprocal, or passive. Here we give some examples, but we will discuss the middle forms in more detail in the next section:

(17) (a) Eg settist niður
I sit down
'I sat down' (reflexive meaning)

(b) Teir berjast altíð
they fight always
'They always fight' (reciprocal meaning)

(c) Oyggin kallast Nólsoy
island-DET call Nólsoy
'The island is called Nólsoy' (passive meaning)

Perfect tense is either formed with the auxiliary *hava* 'have' and the supine (SGN of PSTPTCP) of the main verb, or it is formed with the auxiliary *vera* 'be' and the inflected and agreeing past participle. *Hava* is used with all transitive verbs and most intransitive verbs (Thráinsson et al. 2012: 72):

(18) (a) Hon hevur lisið bókina.
she have.3SG.PST read.SUP book-DET.ACC
'She has read the book'

(b) Teir hava sovið leingi.
they have.3PL.PST sleep.SUP long
'They have slept for long'

(c) Hann hevur verið ríkur.
he have3SG.PST be.SUP rich
'He has been rich'

(d) Hann er vorðin ríkur.
he.NOM.SG.M is become.NOM.SG.M rich
'He has become rich'

Past perfect is formed with past tense of the relevant auxiliary (*vera/hava*as mentioned above), and the perfect passive is formed with the auxiliary *vera* 'be', not *hava* 'have'. We illustrate this with a couple of examples:

(19) (a) Hann hevði verið ríkur.
he have.PST be.SUP rich.NOM.M
'He had been rich'

(b) Hann var vorðin ríkur.
he be.3SG.PST become.PST.PTCP rich.NOM.M
'He had become rich'

(20) Hann er/*hevur ofta vorðin/blivin avsmurdur.
he be.3SG.PRS/*have often become.PST.PTCP off-smear.NOM.M
'He has often been beaten up'

The indicative-subjunctive distinction is not productive in Faroese and past subjunctive forms generally do not exist. Past subjunctive was commonly used in Old Norse (and still is in Modern Icelandic) to indicate a counterfactual or hypthetical situation. The regular past tense can have this function is Faroese, but the meaning of such forms are typically ambiguous. We illustrate this with a few examples:

(21) (a) Eg gjørdi tað fegin.
I do.PST it gladly
'I did it gladly.' or 'I would gladly do it.'

(b) Hann hevði dripið hundin.
he have.PST kill.PST.PTCP dog-DET.ACC.M
'He had killed the dog.' or 'He would have killed the dog.'

(c) Hann tók bókina.
he take.PST book-DET.NOM.F
'He took the book.' or 'He would gladly take the book if...'

(d) Eg hevði fegin gjørt tað, um eg fekk pengar fyri
I have.PST gladly do.PST.PTCP it if I get.PST money for
tað.
it

'I would gladly have done it if I was paid for it.'

(e) Hann drap hundin, um hann fekk hendur á honum.
he kill.PST dog-DET if he get.PST hand.NOM.PL. on it
'He would kill the dog if he got his hands on it.'

(f) Hann hevði tikið bókina frá mær, um hann
he have.PST take.PST.PTCP book-DET.NOM from me if he
hevði sæð meg lisið í henni.
have.PST see.PST.PTCP me read.PST.PTCP in it
'He would have taken the book from me if he had seen me reading it.'

The default word order in Faroese is subject-verb-object (SVO) or subject-auxiliary-main verb-object (SAVO), both in main clauses and embedded ones. We look at some examples (Thráinsson et al. 2012: 236):

(22) (a) Jógvan las bókina.
Jógvan read.PST book-DET.ACC.SG.F
'Jógvan read the book.'

(b) Jógvan hevur lisið bókina.
Jógvan have.PST read.PST.PTCP book-DET.ACC.SG.F
'Jógvan has read the book.'

(c) Eg haldi, at Jógvan hevur lisið bókina.
I think.PRS that Jógvan have.PST read.PST.PTCP book-DET
'I think that Jógvan has read the book.'

As a rule, the indirect object precedes the direct object and typically appears in the dative, although indirect objects in the accusative form also appear. Lets look at some examples:

(23) (a) Turið gav Hjalmari nógvar bøkur.
Turið.NOM give.PST Hjalmar.DAT many book.NOM.PL
'Turið gave Hjalmar many books.'

(b) Eg spurdi, um Zakaris seldi Eivindi tann gamla
I ask if Zakaris.NOM sell.PST Eivind.DAT the old
bilin.
car-DET.ACC
'I asked if Zakaris sold the old car to Eivind.'

(c) Hon lærdi meg niðurlagið.
she teach.PST me.ACCrefrain-DET.ACC
'She taught me the refrain.'

Should we move an object or a prepositional phrase, or some other non-subject, to the front of a sentence, as is done in Topicalisation, the finite verb

shows up in second place followed by the subject, i.e. Faroese is a "verb-second" (V2) language like the other Germanic languages except English. We take a look at some examples (Thráinsson et al. 2012: 238-239):

(24) (a) *Hesa bókina* hevur Jógvan lisið.
this book-DET.ACC have.PST Jógvan.NOM read.PST.PTCP
'This book has Jógvan read.'

(b) *Tann gamla bilin* seldi Zakaris Eivindi.
the old car-DET.ACC sell.3.SG.PST Zakaris.NOM Eivind.DAT.
'The old car Zakaris sold to Eivind.'

(c) *Jóannes* haldi eg eigur hesa bókina.
Jóannes.NOM think I own.3SG.PRS this book-DET.ACC
'Jóannes, I think, owns this book.'

As is typical for modern Germanic languages, adjectives precede the noun they modify:

(25) (a) ein *vøkur* genta, ein *bláur* bilur
a beautiful.NOM.F girl.NOM.F, a blueNOM.M car.NOM.M
'a beautiful girl', 'a blue car'

(b) Hann kom súkklandi á einari *gamlari* súkklu.
he come.PST ride.PRS.PTCP on an old.DAT.F bicycle.DAT.F
'He came riding on an old bicycle.'

When it comes to adverbs we can broadly distinguish between three basic adverbial positions (Thráinsson et al. 2012: 241):

1. the medial position, following the finite verb
2. the verb phrase position, following a possible object and other elements of the verb phrase
3. the modifying position, when the adverb is modifying and adjective and other adverbs

We illustrate this with a few examples:

(26) (a) Tey hava *ikki/ivaleyst/jú/aldri* lisið bókina.
they have not/undoubtedly/actually/never.ADV read book-DET.ACC
'They have not/undoubtedly/actually/never read the book.'

(b) Tey hava lisið bókina *tá/har/væl og virðiliga.*
they have read book-DET.ACC then/there/well and thoroughly.ADV
'They have read the book then/there/well and thoroughly.'

(c) Tey hava lisið hesa ógvuliga longu bókina sera væl.
they have read this extremely.ADV long book-DET.ACC very.ADV well
'They have read this extremely long book very well.'

Let's end by looking at the case of subject, object, and indirect object. The regular *subject* case in Faroese is *nominative*, as we already have seen in previous examples:

(27) (a) *Hann* skrivar.
he.NOM.SG write.3SG
'He writes.'

(b) *Hon* arbeiðir.
she.NOM.SG work.3SG
'She works.'

(c) *Børnini* spæla.
child-DET.NOM.PLplay.3PL
'The children play'

(d) *Vit* settu niður epli í gjár.
we.NOM.PL put down potato.PL yesterday
'We planted potatoes yesterday.'

Before we continue, it should be mentioned that some verbs take non-nominative subjects in modern Faroese, but we won't look further into that here.

Looking at *direct object* case we can say that the *accusative* is the default case, as we have already seen in previous examples:

(28) (a) Hon keypti bókina.
she buy.PST book-DET.ACC
'She bought the book.'

(b) Hann seldi telduna.
he sell.PST computer-DET.ACC
'He sold the computer.'

It should be mentioned, however, that a number of verbs take a direct object in dative case, which we also touched upon earlier, but we will restrict ourselves here to saying that among the verbs that take a direct object in the dative are verbs of helping, ordering, praising, thanking, welcoming, etc., leading one to associate the dative objects with thematic roles such as recipients and experiencers. The semantic verb classes and thematic roles we have mentioned are roughly the same as the ones taking a dative object in Icelandic, although dative objects are getting more rare in modern Faroese while still being prevalent in Icelandic (Thráinsson et al. 2012: 257-258).

At last we have the *indirect object*, which takes the *dative* as its default case, as we have seen in previous examples. Among the verbs taking an indirect object are verbs meaning 'sell', 'lend', 'give', 'send', etc. Let's look at three examples:

(29) (a) Hann beyð henni starv.
he offer.PST she.DAT job.ACC
'He offered her a job.'

(b) Tey fingu sær bil.
they get.PST themselves.DAT car.ACC
'They got themselves a car.'

(c) Fyrigev honum syndir hansara.
forgive.IMP him.DAT sin.PL.ACC he.GEN
'Forgive him his sins.'

We notice from the examples that these verbs, *bjóða* 'offer', *fáa* 'get', *fyrigeva* 'forgive', are examples of verbs that take two objects, one dative (indirect) object and one accusative (direct) object. We call these verbs *ditransitive*. This dative–accusative pattern is also the most common case marking pattern of ditransitive verbs in Icelandic, but there we also have many other patterns, such as dative–dative, dative–genitive, accusative–dative, accusative–genitive, and accusative–accusative, most of which are not found in Faroese (Thráinsson et al. 2012: 262-263).

3. Middle Constructions in Faroese

Above we briefly touched on middle forms in Faroese, but since they are the main subject of this paper we need to look closer at them.

We mentioned that the middle constructions are formed by adding *-st* to various inflectional forms of the verb, but let's now look further at the origin and semantics of the middle constructions.

The origin of the *-st* suffix is considered to be the reflexive pronoun *sik* in Old Norse, which corresponds to the reflexive *seg* in modern Faroese. Thus Old Norse *setja sik* became *setjast* 'sit down', possibly via the intermediate form *setjask*. In modern Faroese this evolved to *setast* (the Old Norse verb *setja* became *seta* in modern Faroese) (Thráinsson et al. 2012: 277).

As we mentioned in the previous section, the *-st*-forms can have a wide variety of meanings, i.e. the semantics vary according to what verb is used, and we'll now look at the various meanings while giving some glossed examples (we add *+st* to the gloss to indicate an *-st*-form) (Thráinsson et al. 2012: 277-278).

Reflexive meaning

(30) (a) Eg settist niður.
I sit+st.PST down
'I sat down.'

(b) Tey vandust skjótt við hitan.
They get-used+st.PST soon with heat-DEF
'They soon got used to the heat.'

Reciprocal meaning

(31) (a) Maðurin og konan heilsaðust.
Man-DEF and woman-DEF greet+st.PST
'The man and the woman greeted each other.'

(b) Tey bítast og klórast.
They bit+st.PRS and scratch+st.PRS
'They bite and scratch each other.' (i.e. 'They fight.')

Modal meaning

(32) (a) Ikki slepst uppaftur uttan hjálp.
not get+st.PRS upp again without help
'One cannot get back up without help.'

(b) Ikki kemst uttanum, at málið hevur týdning.
not get+st.PRS around that language-DEF have.PRS importance
'One cannot get around the fact that language is important.'

Middle (or passive) meaning

(33) (a) Íbúð ynskist til leigu.
apartment wish+st.PRS to rent
'Apartment for rent is sought.'

(b) Bókin seldist væl.
book-DEF sell+st.PST well
'The book sold well.'

(c) Her skal eitt hús byggjast.
here shall a house build+st.PRS
'A house is to be build here.'

(d) Hon hoyrdist syngja langa leið.
she hear+st.PST sing long way
'She was heard singing from a distance.'

(e) Hann brendist illa.
he burn+st.PST badly
'He was badly burnt.'

(f) Bókin fæst ikki.
book-DEF get+st.PRS not
'It is impossible to get the book.'

(g) Dyrnar opnaðust knappliga.
door-DEF.PL open+st.PL.PST suddenly
'The doors suddenly opened.'

These last examples is what corresponds most closely to middles in other languages including English. An important difference to note between middle forms like the ones in example (33) and regular passives is that the former are non-agentive while the latter are agentive. The agent is normally understood (and can be expressed with a prepositional phrase, but this is frequently not done) from a regular passive, but not from the middle. It is even considered ungrammatical to express an agent in middle constructions. We'll mention a couple of examples illustrating the contrast between a middle and a regular passive:

(34) (a) * Bókin seldist av H. N. Jacobsens Bókahandli.
book-DEF sell+st.PST by H. N. Jacobsen's bookstore

(b) Bókin varð seld av H. N. Jacobsens Bókahandli.
book-DEF was sell.PST by H. N. Jacobsen's bookstore
'The book was sold by H. N. Jacobsen's bookstore.'

(35) (a) * Jóhanna brendist av óvinum sínum.
Johanna burn+st.PST by enemy.PL her.REFL

(b) Jóhanna varð brend av óvinum sínum.
Johanna was burn.PST.PTCP by enemy.PL her.REFL
'Johanna was burned by her enemies.'

Here we see that it is not possible to specify an agent in a middle form in the way we can do with a regular passive construction.

Finally we can mention that it can be difficult to predict what type of meaning an *-st*-form of a given verb will have. For instance the same *-st*-form can have two different meanings:

(36) (a) Tey kennast væl.
they know+st.PRS well
'They know each other well.'

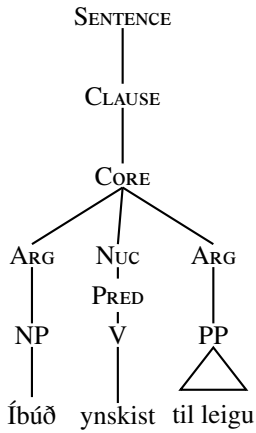
(b) Tey kennast á málinum.
they know+st.PRS on language-DEF
'They can be recognised by the way they speak.'

3.1. Role and Reference Grammar

Before we continue on to the structure of the lexicon let's analyse the middle constructions in example (33) in terms of Role and Reference Grammar (RRG), because this will be the basis for the rules in our parser.

Example (33a): Íbúð ynskist til leigu
This sentence only consists of a single clause. The main verb, the predicate, in the clause is *-st*-form *ynskist*, which is based on the verb *ynskja* 'wish'. In RRG terms we call this the nucleus (NUC) of the clause. The nucleus has two arguments (ARG), the noun *íbúð* 'apartment' and the prepositional phrase *til leigu* 'for rent'. The nucleus together with these two arguments form the core (CORE) of the clause. We use a tree to give a quick overview of the sentence structure:

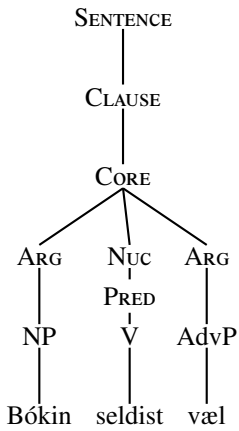
(37)



For the subsequent examples we won't write out the analysis like we have done above, but just illustrate the structure with a diagram, i.e. an RRG tree.

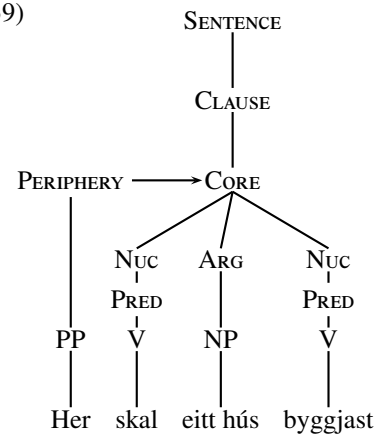
Example (33b): Bókin seldist væl

(38)



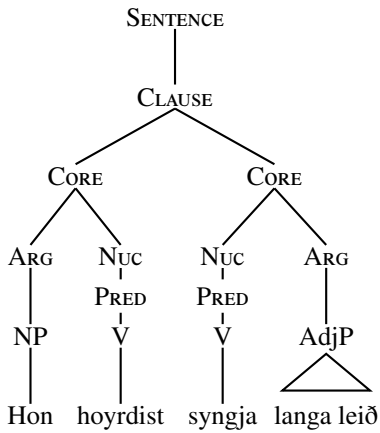
Example (33c): Her skal eitt hús byggjast

(39)



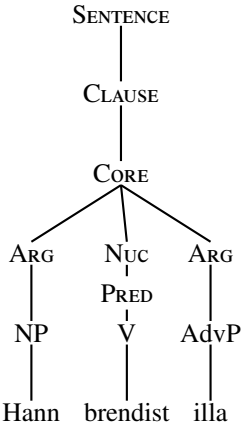
Example (33d): Hon hoyrdist syngja langa leið

(40)



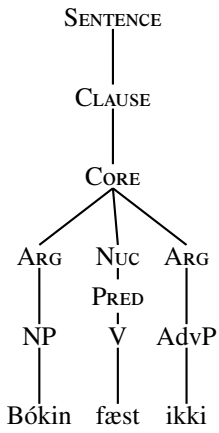
Example (33e): Hann brendist illa

(41)



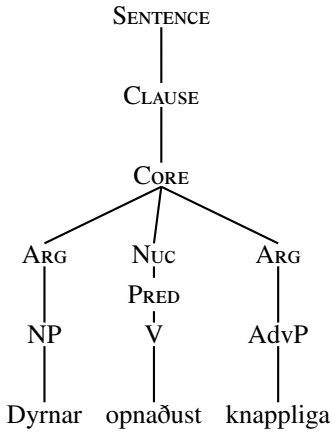
Example (33f): Bókin fæst ikki

(42)



Example (33g): Dyrnar opnaðust knappliga

(43)



4. Implementation of the Lexicon and Grammar

In this section we will first specify a lexicon to be used with the parser, and based on this lexicon we build up a feature based grammar that we can use with the parsers in NLTK (Natural Language ToolKit).

4.1. Lexicon

For our parser we need to define a number of features for nouns, pronouns, verbs, prepositions, adjectives, determiners and adverbs. The features we will record are as follows:

(44) (a) word; the word itself in the form it is found in the text
(b) part of speech type; noun (n), verb (v), etc.

(c) definiteness; DEF- / DEF+
(d) "middleness"; MID- / MID+
(e) person; 1 (first), 2 (second), or 3 (third)
(f) gender; M (masculine), F (feminine), or N (neuter)
(g) number; SG (singular) or PL (plural)
(h) case; NOM (nominative), ACC (accusative), DAT (dative), or GEN (genitive)
(i) tense; PST (past) or PRS (present)
(j) logical structure

Obviously, not all of these features will be relevant for all types of words, and in that case we will write *NA* for that feature. In the lexicon we give here we'll include the words appearing in the examples in (33), and before we get to the lexicon itself we list the words in their base form:

nouns	íbúð, leiga, bók, hús, leið, dyr
determiners	ein
pronouns	hon, hann
verbs	ynskja, selja, skula, byggja, hoyra, syngja, brenna, fáa, opna
prepositions	til
adjectives	langur
adverbs	væl, illa, ikki, her, knappliga

Now we will move on and give each word as it appears in the examples in (33) and define the features as stated above in Example (44).

4.1.1. Nouns

Table 1
The lexical features of the noun *íbúð*.

Word	íbúð
POS type	n
DEF	DEF-
MID	NA
Person	3
Gender	F
Number	SG
Case	NOM/ACC/DAT
Tense	NA
Logical structure	NA

Table 2
The lexical features of the noun *leigu*.

Word	leigu
POS type	n
DEF	DEF-
MID	NA
Person	3
Gender	F
Number	SG
Case	ACC/DAT/GEN
Tense	NA
Logical structure	NA

Table 3
The lexical features of the noun *bókin*.

Word	bókin
POS type	n
DEF	DEF+
MID	NA
Person	3
Gender	F
Number	SG
Case	NOM
Tense	NA
Logical structure	NA

Table 4
The lexical features of the noun *hús*.

Word	hús
POS type	n
DEF	DEF-
MID	NA
Person	3
Gender	N
Number	SG
Case	NOM/ACC
Tense	NA
Logical structure	NA

Table 5
The lexical features of the noun *leið*.

Word	leið
POS type	n
DEF	DEF-
MID	NA
Person	3
Gender	F
Number	SG
Case	NOM/ACC/DAT
Tense	NA
Logical structure	NA

Table 6
The lexical features of the noun *dyrnar*.

Word	dyrnar
POS type	n
DEF	DEF+
MID	NA
Person	3
Gender	F
Number	PL
Case	NOM/ACC
Tense	NA
Logical structure	NA

4.1.2. Determiners

Table 7
The lexical features of the determiner *eitt*.

Word	eitt
POS type	DET
DEF	DEF-
MID	NA
Person	3
Gender	N
Number	SG
Case	NOM/ACC
Tense	NA
Logical structure	NA

4.1.3. Pronouns

Table 8
The lexical features of the pronoun *hon*.

Word	hon
POS type	PRO
DEF	DEF+/-
MID	NA
Person	3
Gender	F
Number	SG
Case	NOM
Tense	NA
Logical structure	NA

Table 9
The lexical features of the pronoun *hann*.

Word	hann
POS type	PRO
DEF	DEF+/-
MID	NA
Person	3
Gender	M
Number	SG
Case	NOM/ACC
Noun type	NA
Verb type	NA
Tense	NA
Logical structure	NA

4.1.4. Verbs

Table 10
The lexical features of the verb *ynskist*.

Word	ynskist
POS type	V
DEF	NA
MID	MID+
Person	3
Gender	NA
Number	SG
Case	NA
Tense	PRS
Logical structure	**wish′**(∅,y)

Table 11
The lexical features of the verb *seldist*.

Word	seldist
POS type	V
DEF	NA
MID	MID+
Person	3
Gender	NA
Number	SG
Case	NA
Tense	PST
Logical structure	**do′**(∅,[**sell′**(y)])

Table 12
The lexical features of the verb *skal*.

Word	skal
POS type	AUX
DEF	NA
MID	MID-
Person	3
Gender	NA
Number	SG
Case	NA
Noun type	NA
Verb type	FIN+
Tense	PRS
Logical structure	**be′**(x,y)

Table 13
The lexical features of the verb *byggjast*.

Word	byggjast
POS type	V
DEF	NA
MID	MID+
Person	3
Gender	NA
Number	SG
Case	NA
Noun type	NA
Verb type	FIN+
Tense	PRS
Logical structure	**do′**(∅,[**build′**(y)])

Table 14
The lexical features of the verb *hoyrdist*.

Word	hoyrdist
POS type	V
DEF	NA
MID	MID+
Person	3
Gender	NA
Number	SG
Case	NA
Tense	PST
Logical structure	**hear′**(∅,y)

Table 15
The lexical features of the verb *syngja*.

Word	syngja
POS type	V
DEF	NA
MID	MID-
Person	3
Gender	NA
Number	SG
Case	NA
Tense	PRS
Logical structure	**do′**(x,[**sing′**(x,(y))])

Table 16
The lexical features of the verb *brendist*.

Word	brendist
POS type	V
DEF	NA
MID	MID+
Person	3
Gender	NA
Number	SG
Case	NA
Tense	PST
Logical structure	**do′**(∅,[**burn′**(y)])

Table 17
The lexical features of the verb *fæst*.

Word	fæst
POS type	V
DEF	NA
MID	MID+
Person	3
Gender	NA
Number	SG
Case	NA
Tense	PRS
Logical structure	**get′**(x,y)

Table 18
The lexical features of the verb *opnaðust*.

Word	opnaðust
POS type	V
DEF	NA
MID	MID+
Person	3
Gender	NA
Number	PL
Case	NA
Noun type	NA
Verb type	FIN+
Tense	PST
Logical structure	**do′**(∅,[**open′**(y)])

4.1.5. Prepositions

Table 19
The lexical features of the preposition *til*.

Word	til
POS type	PREP
DEF	NA
MID	NA
Person	NA
Gender	NA
Number	NA
Case	NA
Tense	NA
Logical structure	**to′**(x)

4.1.6. Adjectives

Table 20
The lexical features of the adjective *langa*.

Word	langa
POS type	ADJ
DEF	DEF-
MID	NA
Person	3
Gender	F
Number	SG
Case	ACC
Tense	NA
Logical structure	NA

4.1.7. Adverbs

Table 21
The lexical features of the adverb *væl*.

Word	væl
POS type	ADV
DEF	NA
MID	NA
Person	NA
Gender	NA
Number	NA
Case	NA
Tense	NA
Logical structure	NA

Table 22
The lexical features of the adverb *illa*.

Word	illa
POS type	ADV
DEF	NA
MID	NA
Person	NA
Gender	NA
Number	NA
Case	NA
Tense	NA
Logical structure	NA

Table 23
The lexical features of the adverb *ikki*.

Word	ikki
POS type	ADV
DEF	NA
MID	NA
Person	NA
Gender	NA
Number	NA
Case	NA
Tense	NA
Logical structure	NA

Table 24
The lexical features of the adverb *her*.

Word	her
POS type	ADV
DEF	NA
MID	NA
Person	NA
Gender	NA
Number	NA
Case	NA
Tense	NA
Logical structure	NA

Table 25
The lexical features of the adverb *knappliga*.

Word	knappliga
POS type	ADV
DEF	NA
MID	NA
Person	NA
Gender	NA
Number	NA
Case	NA
Tense	NA
Logical structure	NA

4.2. Grammar

We will now build up a feature based grammar for use with the parser. For this we will use a so-called feature structure in NLTK, which basically is a Python dictionary with the features as keys and the feature values as values, e.g.:

Listing 1
A feature structure showing lexical features for the word *íbúð*.

```
import nltk

fs = nltk.FeatStruct(
    word='íbúð',
    pos='n',
    definiteness="def-",
    middle="na",
    person=3,
    gender='f',
    number='sg',
    case='nom',
    tense='na',
    ls='na')

print(fs)

[ case         = 'nom'   ]
[ definiteness = 'def-'  ]
[ gender       = 'f'     ]
[ ls           = 'na'    ]
[ middle       = 'na'    ]
[ number       = 'sg'    ]
[ person       = 3       ]
[ pos          = 'n'     ]
[ tense        = 'na'    ]
[ word         = 'íbúð'  ]
```

Faroese nouns agree with verbs in gender, person and number and with other nouns in gender, person, number and case. Nouns agree with adjectives in case, number and gender. To cover these agreement patterns we define an agreement feature like this:

Listing 2
A feature structure showing the word *íbúð* with an *agreement* feature.

```
    import nltk

    fs1 = nltk.FeatStruct(gender='f', person=3, number='sg')

    fs2 = nltk.FeatStruct(
        word='íbúð',
        pos='n',
        definiteness="def-",
        middle="na",
        agreement=fs1,
        case='nom',
        tense='na',
```

```
            ls='na')

        print(fs2)

        [                  [ gender = 'f'  ] ]
        [ agreement    = [ number = 'sg' ] ]
        [                  [ person = 3    ] ]
        [                                    ]
        [ case         = 'nom'               ]
        [ definiteness = 'def-'              ]
        [ ls           = 'na'                ]
        [ middle       = 'na'                ]
        [ pos          = 'n'                 ]
        [ tense        = 'na'                ]
        [ word         = 'íbúð'              ]
```

Now we will move on to define the grammar itself in a file stating first the rules for the middle constructions we have discussed in previous sections, then the lexicon in terms of feature structures as discussed in this section. Below we show the finalised grammar, complete will all grammar productions and lexical productions that cover the middle constructions listed in Example (33) on page 13.

Listing 3
A feature based grammar for parsing of middle constructions in Faroese

```
% start SENT

## Grammar Productions
SENT -> CLAUSE
CLAUSE -> CORE
CLAUSE -> Periphery CORE
CLAUSE -> CORE CORE

CORE[case=?c, agr=?a] -> NP[case=?c, agr=?a] NUC[agr=?a] AdvP
CORE[case=?c, agr=?a] -> AdvP NUC[agr=?a] NP[case=?c, agr=?a]
CORE[case=?c, agr=?a] -> NP[case=?c, agr=?a] NUC[agr=?a] PP[case=?c
    , agr=?a]
CORE[case=?c, agr=?a] -> NUC[agr=?a] NP[case=?c, agr=?a] NUC[agr=?a
    ]
CORE[case=?c, agr=?a] -> NP[case=?c, agr=?a] NUC[agr=?a]
CORE[case=?c, agr=?a] -> NUC[agr=?a] NP[case=?c, agr=?a]
CORE[case=?c, agr=?a] -> NUC[agr=?a] AdjP[case=?c, agr=?a]

NP[case=?c, agr=?a] -> Det[case=?c, agr=?a] N[case=?c, agr=?a]
NP[case=?c, agr=?a] -> N[case=?c, agr=?a]
NP[case=?c, agr=?a] -> PropN[case=?c]
NP[case=?c, agr=?a] -> PRO[case=?c, agr=?a]

NUC[agr=?a] -> PRED[agr=?a]
PRED[agr=?a] -> V[agr=?a]

PP[case=?c, agr=?a] -> Prep NP[case=?c, agr=?a]

AdjP[case=?c, agr=?a] -> Adj[case=?c, agr=?a] NP[case=?c, agr=?a]

Periphery -> AdvP
```

```
30
31
32 ## Lexical Productions
33
34 # Prepositions
35 Prep -> 'til' | 'av'
36
37 # Determiners
38 Det[case=nom, agr=[gdr=n, per=3, num=sg]] -> 'eitt'
39 Det[case=acc, agr=[gdr=n, per=3, num=sg]] -> 'eitt'
40
41 # Pronouns
42 PRO[case=nom, agr=[gdr=f, per=3, num=sg]] -> 'hon'
43 PRO[case=acc, agr=[gdr=f, per=3, num=sg]] -> 'hana'
44 PRO[case=dat, agr=[gdr=f, per=3, num=sg]] -> 'henni'
45 PRO[case=gen, agr=[gdr=f, per=3, num=sg]] -> 'hennara'
46 PRO[case=nom, agr=[gdr=m, per=3, num=sg]] -> 'hann'
47 PRO[case=acc, agr=[gdr=m, per=3, num=sg]] -> 'hann'
48 PRO[case=dat, agr=[gdr=m, per=3, num=sg]] -> 'honum'
49 PRO[case=gen, agr=[gdr=m, per=3, num=sg]] -> 'hansara'
50 PRO[case=dat, agr=[gdr=m, per=3, num=sg]] -> 'sínum'
51
52 # Nouns
53 N[case=nom, agr=[gdr=f, per=3, num=sg]] -> 'íbúð' | 'leið'
54 N[case=acc, agr=[gdr=f, per=3, num=sg]] -> 'íbúð' | 'leigu' | 'leið
       '
55 N[case=dat, agr=[gdr=f, per=3, num=sg]] -> 'íbúð' | 'leigu' | 'leið
       '
56 N[case=nom, agr=[gdr=f, per=3, num=sg]] -> 'bókin'
57 N[case=nom, agr=[gdr=n, per=3, num=sg]] -> 'hús'
58 N[case=acc, agr=[gdr=n, per=3, num=sg]] -> 'hús'
59 N[case=nom, agr=[gdr=f, per=3, num=pl]] -> 'dyrnar'
60 N[case=acc, agr=[gdr=f, per=3, num=pl]] -> 'dyrnar'
61 N[case=dat, agr=[gdr=m, per=3, num=pl]] -> 'óvinum'
62
63 # Proper nouns
64 PropN[case=nom] -> 'jóhanna'
65
66 # Verbs
67 V[agr=[num='sg', per=3], tns='prs'] -> 'ynskist'
68 V[agr=[num='sg', per=3], tns='pst'] -> 'seldist'
69 V[agr=[num='sg', per=3], tns='prs'] -> 'skal'
70 V[agr=[num='sg', per=3], tns='prs'] -> 'byggjast'
71 V[agr=[num='sg', per=3], tns='pst'] -> 'hoyrdist'
72 V[agr=[num='sg', per=3], tns='prs'] -> 'syngja'
73 V[agr=[num='sg', per=3], tns='pst'] -> 'brendist'
74 V[agr=[num='sg', per=3], tns='prs'] -> 'fæst'
75 V[agr=[num='pl', per=3], tns='pst'] -> 'opnaðust'
76 V[agr=[num='sg', per=3], tns='pst'] -> 'brendist'
77
78 # Adverbs
79 AdvP -> 'ikki' | 'illa' | 'væl' | 'her' | 'knappliga'
80
81 # Adjectives
82 Adj[case=acc, agr=[gdr=f, per=3, num='sg']] -> 'langa'
```

5. Implementation of the Parser

The parser itself will utilise the chart parser that comes with NLTK as described in (Bird et al. 2014: Chapter 9). Since we use the NLTK parser with the feature based grammar from the previous section, the amount of code is minimal, hence we display it all here in Listing (4).

Listing 4
An RRG parser for middle constructions in Faroese implemented in Python.

```
# -*- coding: utf-8 -*-

import nltk

def read_sentences(_file):
    """
    Read sentences to parse from a text file and store them in a
    list.
    The sentences are given one per line in the file.

    Input: file name
    Return: list of sentences
    """

    with open(_file, 'r') as sentences:
        return sentences.read().splitlines()


def parse_sentence(sentence, trace=0):
    """
    Using the chart parser from NLTK, parse a given sentence and
    return an RRG parse tree.

    Input: Sentence, trace (if 0, return only the parse tree,
    otherwise return intermediate steps)
    Output: Parse tree showing the various constituents of a
    sentence.
    """

    hline = 30 * '='
    print("\n{0}\n{1}\n{0}".format(hline, sentence))

    tokens = sentence.lower().split()

    cp = nltk.load_parser("faroese.fcfg", trace=trace)

    for tree in cp.parse(tokens):
        print(tree)


if __name__ == "__main__":
    trace = 0  # define the trace level; we set it to 0 to only
    return the final parse tree
    sentences = read_sentences("sentences.txt")  # read in the
    sentences from the file sentences.txt
    ungrammatical_sentences = read_sentences("
    sentences_ungrammatical.txt")

```

```
    for sentence in sentences:  # loop through the sentences and
    return the parse tree
        parse_sentence(sentence, trace)

    for sentence in ungrammatical_sentences:
        parse_sentence(sentence, trace)
```

6. Testing of the Parser

We list the sentences mentioned in Example (33) on page 13 and run them through the parser shown in Listing 4. The sentences and the result of the parsing are shown below.

Listing 5
The sentences used to test the parser

```
Íbúð ynskist til leigu
Bókin seldist væl
Her skal eitt hús byggjast
Hon hoyrdist syngja langa leið
Hann brendist illa
Bókin fæst ikki
Dyrnar opnaðust knappliga
```

Listing 6
The parse trees returned by the parser.

```
    ==============================
    Íbúð ynskist til leigu
    ==============================
    (SENT[]
      (CLAUSE[]
        (CORE[agr=[gdr='f', num='sg', per=3], case='dat']
          (NP[agr=[gdr='f', num='sg', per=3], case='dat']
            (N[agr=[gdr='f', num='sg', per=3], case='dat'] íbúð))
          (NUC[agr=[num='sg', per=3]]
            (PRED[agr=[num='sg', per=3]]
              (V[agr=[num='sg', per=3], tns='prs'] ynskist)))
          (PP[agr=[gdr='f', num='sg', per=3], case='dat']
            (Prep[] til)
            (NP[agr=[gdr='f', num='sg', per=3], case='dat']
              (N[agr=[gdr='f', num='sg', per=3], case='dat'] leigu)
    )))))
    (SENT[]
      (CLAUSE[]
        (CORE[agr=[gdr='f', num='sg', per=3], case='acc']
          (NP[agr=[gdr='f', num='sg', per=3], case='acc']
            (N[agr=[gdr='f', num='sg', per=3], case='acc'] íbúð))
          (NUC[agr=[num='sg', per=3]]
            (PRED[agr=[num='sg', per=3]]
              (V[agr=[num='sg', per=3], tns='prs'] ynskist)))
          (PP[agr=[gdr='f', num='sg', per=3], case='acc']
            (Prep[] til)
```

```
            (NP[agr=[gdr='f', num='sg', per=3], case='acc']
              (N[agr=[gdr='f', num='sg', per=3], case='acc'] leigu)
))))))

==============================
Bókin seldist væl
==============================
(SENT[]
  (CLAUSE[]
    (CORE[agr=[gdr='f', num='sg', per=3], case='nom']
      (NP[agr=[gdr='f', num='sg', per=3], case='nom']
        (N[agr=[gdr='f', num='sg', per=3], case='nom'] bókin))
      (NUC[agr=[num='sg', per=3]]
        (PRED[agr=[num='sg', per=3]]
          (V[agr=[num='sg', per=3], tns='pst'] seldist)))
      (AdvP[] væl))))

==============================
Her skal eitt hús byggjast
==============================
(SENT[]
  (CLAUSE[]
    (Periphery[] (AdvP[] her))
    (CORE[agr=[gdr='n', num='sg', per=3], case='nom']
      (NUC[agr=[num='sg', per=3]]
        (PRED[agr=[num='sg', per=3]]
          (V[agr=[num='sg', per=3], tns='prs'] skal)))
      (NP[agr=[gdr='n', num='sg', per=3], case='nom']
        (Det[agr=[gdr='n', num='sg', per=3], case='nom'] eitt)
        (N[agr=[gdr='n', num='sg', per=3], case='nom'] hús))
      (NUC[agr=[num='sg', per=3]]
        (PRED[agr=[num='sg', per=3]]
          (V[agr=[num='sg', per=3], tns='prs'] byggjast))))))
(SENT[]
  (CLAUSE[]
    (Periphery[] (AdvP[] her))
    (CORE[agr=[gdr='n', num='sg', per=3], case='acc']
      (NUC[agr=[num='sg', per=3]]
        (PRED[agr=[num='sg', per=3]]
          (V[agr=[num='sg', per=3], tns='prs'] skal)))
      (NP[agr=[gdr='n', num='sg', per=3], case='acc']
        (Det[agr=[gdr='n', num='sg', per=3], case='acc'] eitt)
        (N[agr=[gdr='n', num='sg', per=3], case='acc'] hús))
      (NUC[agr=[num='sg', per=3]]
        (PRED[agr=[num='sg', per=3]]
          (V[agr=[num='sg', per=3], tns='prs'] byggjast))))))

==============================
Hon hoyrdist syngja langa leið
==============================
(SENT[]
  (CLAUSE[]
    (CORE[agr=[gdr='f', num='sg', per=3], case='nom']
      (NP[agr=[gdr='f', num='sg', per=3], case='nom']
        (PRO[agr=[gdr='f', num='sg', per=3], case='nom'] hon))
      (NUC[agr=[num='sg', per=3]]
        (PRED[agr=[num='sg', per=3]]
          (V[agr=[num='sg', per=3], tns='pst'] hoyrdist))))
```

```
      (CORE[agr=[gdr='f', num='sg', per=3], case='acc']
        (NUC[agr=[num='sg', per=3]]
          (PRED[agr=[num='sg', per=3]]
            (V[agr=[num='sg', per=3], tns='prs'] syngja)))
        (AdjP[agr=[gdr='f', num='sg', per=3], case='acc']
          (Adj[agr=[gdr='f', num='sg', per=3], case='acc'] langa)
          (NP[agr=[gdr='f', num='sg', per=3], case='acc']
            (N[agr=[gdr='f', num='sg', per=3], case='acc'] leið))
))))

==============================
Hann brendist illa
==============================
(SENT[]
  (CLAUSE[]
    (CORE[agr=[gdr='m', num='sg', per=3], case='nom']
      (NP[agr=[gdr='m', num='sg', per=3], case='nom']
        (PRO[agr=[gdr='m', num='sg', per=3], case='nom'] hann))
      (NUC[agr=[num='sg', per=3]]
        (PRED[agr=[num='sg', per=3]]
          (V[agr=[num='sg', per=3], tns='pst'] brendist)))
      (AdvP[] illa))))
(SENT[]
  (CLAUSE[]
    (CORE[agr=[gdr='m', num='sg', per=3], case='acc']
      (NP[agr=[gdr='m', num='sg', per=3], case='acc']
        (PRO[agr=[gdr='m', num='sg', per=3], case='acc'] hann))
      (NUC[agr=[num='sg', per=3]]
        (PRED[agr=[num='sg', per=3]]
          (V[agr=[num='sg', per=3], tns='pst'] brendist)))
      (AdvP[] illa))))

==============================
Bókin fæst ikki
==============================
(SENT[]
  (CLAUSE[]
    (CORE[agr=[gdr='f', num='sg', per=3], case='nom']
      (NP[agr=[gdr='f', num='sg', per=3], case='nom']
        (N[agr=[gdr='f', num='sg', per=3], case='nom'] bókin))
      (NUC[agr=[num='sg', per=3]]
        (PRED[agr=[num='sg', per=3]]
          (V[agr=[num='sg', per=3], tns='prs'] fæst)))
      (AdvP[] ikki))))

==============================
Dyrnar opnaðust knappliga
==============================
(SENT[]
  (CLAUSE[]
    (CORE[agr=[gdr='f', num='pl', per=3], case='nom']
      (NP[agr=[gdr='f', num='pl', per=3], case='nom']
        (N[agr=[gdr='f', num='pl', per=3], case='nom'] dyrnar))
      (NUC[agr=[num='pl', per=3]]
        (PRED[agr=[num='pl', per=3]]
          (V[agr=[num='pl', per=3], tns='pst'] opnaðust)))
      (AdvP[] knappliga))))
(SENT[]
```

```
      (CLAUSE[]
        (CORE[agr=[gdr='f', num='pl', per=3], case='acc']
          (NP[agr=[gdr='f', num='pl', per=3], case='acc']
            (N[agr=[gdr='f', num='pl', per=3], case='acc'] dyrnar))
          (NUC[agr=[num='pl', per=3]]
            (PRED[agr=[num='pl', per=3]]
              (V[agr=[num='pl', per=3], tns='pst'] opnaðust)))
          (AdvP[] knappliga))))
```

Here we see that the sentences have been correctly parsed, albeit some of them return two parse trees, because the object noun can be found in two different cases; we do not enough information to determine the case of the noun.

Since all of the sentences were parsed correctly we conclude that they are all grammatical. Let's now add a sentence that be know is ungrammatical, e.g the sentence in (35a) on page 14. In order to do this we need to add the words occurring in the sentence to our feature based grammar. The words are the following:

Table 26
Categorisation of the words appearing in (35a).

Noun	óvinum
Proper noun	Jóhanna
Preposition	av
Verb	brendist
Pronoun	sínum

We add this sentence to a new list of ungrammatical sentences:

Listing 7
The ungrammatical sentences used to test the parser

```
Jóhanna brendist av óvinum sínum
```

Then we add the following lexical productions to our grammar:

```
    N[case=dat, agr=[gdr=m, per=3, num=pl]] -> 'óvinum'
    PropN[case=nom] -> 'jóhanna'
    Prep -> 'til' | 'av'
    V[agr=[num='sg', per=3], tns='pst'] -> 'brendist'
    PRO[case=dat, agr=[gdr=m, per=3, num=sg]] -> 'sínum'
```

And the following lines to our Python code:

```
    ungrammatical_sentences = read_sentences("
    sentences_ungrammatical.txt")
    for sentence in ungrammatical_sentences:
        parse_sentence(sentence, trace)
```

When we run this we see that no parse tree is returned, which means that the sentence is not grammatical (we have cut out the output from parsing of all the grammatical sentences):

```
[sigmund@oc8351072275 code]$ python3 parse.py
..............
..............
..............
==============================
Jóhanna brendist av óvinum sínum
==============================
[sigmund@oc8351072275 code]$
```

7. Conclusion

We have given an overview of Faroese morphology and syntax, then looked closer at middle constructions and analysed middle forms with a passive reading using the Role and Reference Grammar (RRG) framework. The RRG framework has then been used to build a feature based grammar, which can be plugged into NLTK's chart parser, and we finally saw that it is able to parse the middle constructions with a passive reading correctly. We have also shown that ungrammatical sentences are not parsed correctly, hence our parser is doing what we set out to do.

7.1. Future Work

Looking a future work, a first step could be to expand the parser to handle the other readings of middle in Faroese, and ultimately parse any Faroese sentence construction into its constituent parts.

REFERENCES

Bird, Steven, Ewan Klein & Edward Loper. 2014. Natural language processing with python – analyzing text with the natural language toolkit. http://www.nltk.org/book/. Accessed: 2018-04-28.

Thráinsson, Höskuldur, Hjalmar P. Petersen, Jógvan í Lon Jacobsen & Zakaris Svabo Hansen. 2012. *Faroese – an overview and reference grammar*. Faroe University Press and Linguistic Institute, University of Iceland 2nd edn.

Author's address: *Institute of Technology Blanchardstown, Blanchardstown Road North, Dublin 15, Ireland*
B00108126@student.itb.ie