Documentation for the
Faroese PAROLE PoS-tagset

Edited: 03-05-2022

Table of Contents

Preface

This manual is a guide to the Faroese PAROLE POS-tagging system. The tagset has been inspired by the Danish PAROLE tagset (Keson 1999), as well as some tagging systems for other languages, e.g. Icelandic.

The manual serves the purpose of guiding anyone who has interest in the Faroese PAROLE tagset and the details of it. This document therefore describes the system behind PAROLE and goes into detail about all the linguistic attributes of every word class and provides examples for every tag.

The Faroese PAROLE tagset is a very detailed PoS-system and represents much of the complexity of Faroese grammar. The tagset includes the vast majority of grammatical categories and marks, but it is not exhaustive.

Questions or other can be directed at

Iben Nyholm Debess
ibendebess@gmail.com

Sandra S. Lamhauge
sandralamhauge@gmail.com

Annika Simonsen
annika.simonsen@hotmail.com

the Faroe Islands, ultimo 2020

Introduction and overview

This document will introduce the Faroese PAROLE tag system. All of its features and values will be presented and there will be linguistic examples to demonstrate the use of the tag system.

The table below is an overview of the PAROLE tag set adjusted to Faroese (inspired by Volz & Lenz 1996). Positions 1-7 contain the morphosyntactical features that are the same across all PAROLE-languages. Positions 8-12 are language specific and here they have been created for Faroese.

In table 1, the grey cells represent PAROLE features that are not used in the Faroese tagset (*inactive* attributes), and the white cells represent features that are used in the Faroese tagset (*active* attributes).

CatGram	SsCatGram	3	4	5	6	7	8	9	10	11	12
Noun		Gender	Number	Case			Definiteness	Register	Possessiveness		
Verb		Mood	Tense	Person	Number	Gender	Definiteness	TrCat	Voice	Case	Register
Adj		Degree	Gender	Number	Case		Definiteness	Register	TrCat		
Pron		Person	Gender	Number	Case	Possessor	Reflexive	Register			
Det		Person	Gender	Number	Case	Possessor					
Art		Gender	Number	Case							
Adv		Degree	Function	Wh-ness							
Adpos		Formation	Gender	Number			Case	Register			
Conj		C-type	Coord-posit								
Num		Gender	Number	Case			Register				
Interj											
Residual											
Unique											

Table 1: Overview of the Faroese PAROLE tag system.

CatGram is short for *grammatical category*. For Faroese, this is essentially equivalent of word class.
SsCatGram is short for *subsections of the grammatical categories*. This function varies across languages.

CatGram (1st position) and SsCatGram (2nd position) indicate the classification of a word, and any lemma can only have one of the values in CatGram and one of the values in SsCatGram. Values in the 3rd position and all the way through to the 12th position indicate inflectional attributes, and a lemma can in theory be tagged with every value in these positions.

All the word classes (CatGram) in the table will be explained in chapters of their own.

All the different features (e.g. Gender, Case, Person, Register) have values behind them to mark the feature correctly. The values are made of letters and numbers. The values of each feature will be explained in their respective chapters.

The morphosyntactic analysis of a word is represented by a string of letters, numbers and symbols (*values*), corresponding to one of the features in the table. This string of values is what we call the *tag* or the *PoS-tag*. The key to the information in this string is the placement and sequence of the letters and numbers, again corresponding to the table. To make sure that the sequence is always consistent, we use these symbols as place holders for the *inactive* values:

The equal sign (=) serves as a place holder for all the features that are inactive for a whole word class. This sign is only used in places 3-7, where the translinguistic features are inactive or where no translinguistic values has been assigned. This means, for example, that the Adpos word class has two letters, then five equal signs, and then two more letters: SP=====AU and the string for Nouns will have five letters, two equal signs and two more letters: NCF3A==DU.
The equal sign can only be written in places that are grey in the table.

The dash (-) serves as a place holder for when a feature is not being used in a certain subdivision of the word class, but is relevant in other subdivisions of the word class. The dash can only be written in places that are white and active in the table.

The last active feature of a string concludes the string**:** Anytime the last active attribute of a CatGram or an SsCatGram has been marked, the place holders (= and -) are not added anymore. This means that the word class Adv is marked with three letters only and **nothing** more (e.g. RGU) and Nouns do **not** have equal signs all the way to the 12th place.

Square brackets ([]) in a tag string mean that more than one value is relevant for a feature, e.g. a noun with the same word form in nominative and accusative case would be represented as [AN].

In this manual, all PAROLE tag strings will be indicated in **angle brackets** (<>), e.g. <VADA2S---M-U>. This practice serves only as a practical visual representation. When working with PAROLE there is no need for brackets.

In this manual examples will be in the following format:

dømi <TAG-STRING> 'example'

<u>Her</u> *er eitt dømi á* føroyskum <RGU> 'here the example is translated'

Her <u>er</u> eitt dømi á føroyskum <VADR3S---A-U> 'here the example is translated'

Her er <u>eitt</u> dømi á føroyskum <PI-NSN--U> 'here the example is translated'

Her er eitt <u>dømi</u> á føroyskum <NCNSN==IU> 'here the example is translated'

Her er eitt dømi <u>á</u> føroyskum <SP=====DU> 'here the example is translated'

Her er eitt dømi á <u>føroyskum</u> <ANP[MN]SD=IU[RU]> 'here the example is translated'

----

bilur <NCMSN==IU> 'car'

Hatturin <u>fleyg</u> vekk <VADA3S---A-U> 'the hat flew away'

If you wish to work with PAROLE and tag your own matieral or other, keep in mind the following:

Technically, all features and values are independent, and one could *theoretically* mix and match any feature value with any other feature value in the tag string. Any user of of PAROLE is free to use it in any way they see fit, as long as it does not defy or change the tag system. But seeing as a language specific linguistic tag system is naturally anchored in the grammar of the language, natural restrictions in the combinations of values arise.

This document presents guidelines to which features and values are relevant for Faroese based on the linguistic material and background literature of Faroese Grammar. The document also presents guidelines to which tags are suitable for all the different categories and small details of the language. The user is free to disagree, and give a specific word form another tag, than the one we propose. But note, that behind these guidelines lies extensive deliberation and practical experience with PAROLE.

We recommend using the PAROLE tag system with these following guidelines in mind to maintain a sensible, structural and comparable tag system for your linguistic material.

Nouns

CatGram 1	SsCatGram 2	3	4	5	6	7	8	9	10	11	12
Noun	Noun category	Gender	Number	Case	=	=	Definiteness	Register	Possessiveness		
<N>	Common <C>	<M, F, N, U>	<S, P>	<N, A, D, G>	=	=	<I, D>	<U, O>	<P, U>		
	Proper <P>	<M, F, N, U>	<S, P>	<N, A, D, G>	=	=	<I, D>	<U, O>	<P, U>		
	Indeclineable Common <I>	<M, F, N, U>	-	-	=	=	-	-	<P, U>		
	Indeclineable Proper <O>	<M, F, N, U>	-	-	=	=	-	-	<P, U>		

Table 2: Overview of the Faroese PAROLE tag system for nouns

Nouns are divided into *common nouns, proper nouns, indeclineable common nouns* and *indeclinable proper nouns*. The following section will go through the details of these four subcategories.

The section on Nouns will be concluded with some examples of all the different tagging possibilities.

Common Nouns

Noun	Noun category	Gender	Number	Case	=	=	Definiteness	Register	Possessiveness	
<N>	Common <C>	<M, F, N, U>	<S, P>	<N, A, D, G>	=	=	<I, D>	<U, O>	<P, U>	

As shown in the table above, common nouns are tagged for gender, number, case, definiteness, register and possessiveness.

Gender: masculine, feminine, neuter and unmarked (M, F, N, U)

All common nouns have grammatical gender. The gender is reflected in the inflection patterns, which varies across the genders (as well as gender-internally). Unmarked gender is mainly used for acronyms.

Number: singular and plural (S, P)

Common nouns are always written in either singular or plural. Some nouns have the same form in singular and plural (some nouns in neuter as an example: *bál* (S) - *bál* (P) 'bonfire', *skap* (S) - *skap*/*skøp* (P) 'shape').

Case: nominative, accusative, dative and genitive (N, A, D, G)

Common nouns are inflected in four cases. Some case inflections might have the same form, for example *bát* (N) - *bát* (A) 'boat' and *kettu* (A) - *kettu* (D) 'cat'. When the form of the noun can indicate more than one case, all the possible cases are listed in square brackets. For the examples mentioned here, the analysis in a lexicon entry would be *bát* [NA] and *kettu* [AD]. Whether a noun is in accusative or dative case can only be indicated by the context, e.g. in *Eg sá eina kettu,* 'I saw a cat(A)' and *Eg møtti einari kettu,* 'I met a cat(D)' the inflection on the indefinite pronoun *eina(A)/einari(D)* shows a tagger which case the noun is in.

Definiteness: definite and indefinite (D, I)

Common nouns are analysed as being either definite or indefinite. Definiteness is typically indicated with a suffix, e.g. *maður* (I) and *maðurin* (D), 'man' and 'the man'.

Register: **obsolete and unmarked (O, U)**

Common nouns are analysed as having either obsolete or unmarked register. Genitive forms are always marked obsolete in register, as is indicated in Thráinsson et al. (2012). Nouns in all other cases are unmarked. Note that although not all genitive forms of nouns are obsolete in the Faroese vernacular, (e.g. genitive forms of city and village names are commonly used) all genitive forms of nouns are still marked as obsolete in our PAROLE (see more in the section on Obsolete Register at the end of this document).

Possessiveness: **possessive and unmarked (P, U)**

Common nouns are analysed as being either possessive or unmarked in regard to possessiveness. In Faroese, possessive forms are colloqually expressed with a clitic "-sa". The clitic usually follows a nominative or accusative form, men not always, as in *mamma* (U) and *mammusa* (P), 'mother' and 'mother's. The "-sa" clitic mainly applies to kinship terminology, proper nouns and 'live' common nouns or common nouns with agency (Thráinsson et al 2012).

Proper Nouns

CatGram 1	SsCatGram 2	3	4	5	6	7	8	9	10	11	12
Noun	Noun category	Gender	Number	Case	=	=	Definiteness	Register	Possessiveness		
<N>	Proper <P>	<M, F, N, U>	<S, P>	<N, A, D, G>	=	=	<I, D>	<U, O>	<P, U>		

Proper nouns are tagged for the same features as common nouns: for gender (M, F, N, U), number (S, P), case (N, A, D, G), definiteness (D, I) and register (O, U).

*Gender***: masculine, feminine, neuter and unmarked (M, F, N, U)**

Not all proper nouns have grammatical gender (M, F, N); those who do not, get tagged as Unmarked (U). Gender is mainly reflected in the inflectional patterns, which varies across the genders (as well as gender-internally).

*Number***: singular and plural (S, P)**

Proper nouns are analysed as either singular or plural. Most proper nouns are always singular, while some proper nouns are always plural, e.g. *Akrar* <NPMP[AN]==IU> 'the place name of Akrar', and a few proper nouns can be inflected in both singular and plural, e.g. *Sund* <NPNS[AN]==IU> 'the place name of Sund (sg.)' and *Sund* <NPNP[AN]==IU> 'the place name of Sund (pl.)'. If the number of a proper noun is unclear, it is often indeclineable and will be tagged as such. If it is not indeclineable, and the number is unclear, singular number is the default tag.

*Case***: nominative, accusative, dative and genitive (N, A, D, G)**

Proper nouns are inflected in four cases. Some case inflections might have identical forms and these cases will be presented in square brackets.

*Definiteness***: definite and indefinite (D, I)**

Proper nouns are analysed as being either definite or indefinite. Definiteness is typically indicated with a suffix, *Landavegur* (I) and *Landavegurin* (D), 'Country Road' and 'The Country Road'. Most proper nouns are indefinite.

Register: **obsolete and unmarked (O, U)**

Proper nouns are analysed as being either obsolete or unmarked. Genitive forms are always marked obsolete, as is indicated in Thráinsson et al. (2012). Nouns in all other cases are unmarked. Note that although not all genitive forms of nouns are obsolete in Faroese vernacular, (e.g. genitive forms of city and village names are commonly used) all genitive forms of nouns are still marked as obsolete in our PAROLE (see more in the section on Obsolete Register at the end of this document).

Possessiveness: **possessive and unmarked (P, U)**

Common nouns are analysed as being either possessive or unmarked in regard to possessiveness. In Faroese, possessive forms are colloqually expressed with a clitic "-sa". The clitic usually follows a nominative or accusative form, men not always, as in *mamma* (U) and *mammusa* (P), 'mother' and 'mother's. The "-sa" clitic mainly applies to kinship terminology, proper nouns and 'live' common nouns or common nouns with agency (Thráinsson et al 2012).

Indeclineable Common Nouns

CatGram 1 Noun	SsCatGram 2 Noun category	3 Gender	4 Number	5 Case	6 =	7 =	8 Definiteness	9 Register	10 Possessiveness	11	12
<N>	Indeclineable Common <I>	<M, F, N, U>	-	-	=	=	-	-	<P, U>		

The subcategory *indeclineable common nouns* will only include indeclineable common nouns and will be tagged as shown above. Indeclinable nouns are tagged for Gender only.

Gender: **masculine, feminine, neuter and unmarked (M, F, N, U)**

Some indeclineable common nouns still have a gender indication and are tagged as such. Indeclineable common nouns with no overt gender are tagged as Unmarked.

Possessiveness: **possessive and unmarked (P, U)**

Common nouns are analysed as being either possessive or unmarked in regard to possessiveness. In Faroese, possessive forms are colloqually expressed with a clitic "-sa". The clitic usually follows a nominative or accusative form, men not always, as in *mamma* (U) and *mammusa* (P), 'mother' and 'mother's. The "-sa" clitic mainly applies to kinship terminology, proper nouns and 'live' common nouns or common nouns with agency (Thráinsson et al 2012).

Indeclinable Proper Nouns

CatGram 1	SsCatGram 2	3	4	5	6	7	8	9	10	11	12
Noun	Noun category	Gender	Number	Case	=	=	Definiteness	Register	Possessiveness		
<N>	Indeclineable Proper <O>	<M, F, N, U>	-	-	=	=	-	-	<P, U>		

The subcategory *indeclineable proper nouns* will only include indeclineable proper nouns and will be tagged as shown above. Indeclinable nouns are tagged for Gender only.

*Gender***: masculine, feminine, neuter and unmarked (M, F, N, U)**

Some indeclineable proper nouns still have a gender indication and are tagged as such. Indeclineable proper nouns with no overt gender are tagged as Unmarked.

Possessiveness: **possessive and unmarked (P, U)**

Common nouns are analysed as being either possessive or unmarked in regard to possessiveness. In Faroese, possessive forms are colloqually expressed with a clitic "-sa". The clitic usually follows a nominative or accusative form, men not always, as in *mamma* (U) and *mammusa* (P), 'mother' and 'mother's. The "-sa" clitic mainly applies to kinship terminology, proper nouns and 'live' common nouns or common nouns with agency (Thráinsson et al 2012).

Examples, nouns

Examples from the lexicon RAVN_lex:

Common nouns

Aðalskrivari 'sectretary general'
aðalskrivari	<NCMSN==IUU>
aðalskrivara	<NCMS[AD]==IUU>
aðalskrivara	<NCM[PS]G==IOU>
aðalskrivarin	<NCMSN==DUU>
aðalskrivaran	<NCMSA==DUU>
aðalskrivaranum	<NCMSD==DUU>
aðalskrivarans	<NCMSG==DOU>
aðalskrivarar	<NCMP[AN]==IUU>
aðalskrivarum	<NCMPD==IUU>
aðalskrivararnir	<NCMPN==DUU>
aðalskrivararnar	<NCMPA==DUU>
aðalskrivarunum	<NCMPD==DUU>
aðalskrivaranna	<NCMPG==DOU>

Lagkøkusnitta 'a slice of layer cake'
lagkøkusnitta	<NCFSN==IUU>
lagkøkusnittu	<NCFS[AD]==IUU>
lagkøkusnittu	<NCFSG==IOU>
lagkøkusnittan	<NCFSN==DUU>
lagkøkusnittuna	<NCFSA==DUU>

lagkøkusnittuni   `<NCFSD==DUU>`
lagkøkusnittunnar   `<NCFSG==DOU>`
lagkøkusnittur   `<NCFP[AN]==IUU>`
lagkøkusnittum   `<NCFPD==IUU>`
lagkøkusnitta   `<NCFPG==IOU>`
lagkøkusnitturnar   `<NCFP[AN]==DUU>`
lagkøkusnittunum   `<NCFPD==DUU>`
lagkøkusnittanna   `<NCFPG==DOU>`

Listafólk 'artist'

listafólk   `<NCN[PS][AN]==IUU>`
listafólki   `<NCNSD==IUU>`
listafólks   `<NCNSG==IOU>`
listafólkið   `<NCNS[AN]==DUU>`
listafólkinum   `<NCNSD==DUU>`
listafólksins   `<NCNSG==DOU>`
listafólkum   `<NCNPD==IUU>`
listafólka   `<NCNPG==IOU>`
listafólkini   `<NCNP[AN]==DUU>`
listafólkunum   `<NCNPD==DUU>`
listafólkanna   `<NCNPG==DOU>`

Proper nouns

Fríðrikkur 'Frederick'

Fríðrikkur   `<NPMSN==IUU>`
Fríðrikk   `<NPMSA==IUU>`
Fríðrikki   `<NPMSD==IUU>`
Fríðriks   `<NPMSG==IOU>`
Fríðrikkursa   `<NPMSN==IUP>`
Fríðrikksa   `<NPMSA==IUP>`

Vaglið

Vaglið   `<NPNS[AN]==DUU>`
Vaglinum   `<NPNSD==DUU>`
Vaglsins   `<NPNSG==DOU>`

Umhvørvisstovan 'The Environment Agency'

Umhvørvisstovan   `<NPFSN==DUU>`
Umhvørvisstovuna   `<NPFSA==DUU>`
Umhvørvisstovuni   `<NPFSD==DUU>`
Umhvørvisstovunnar `<NPFSG==DOU>`

Indeclineable common nouns

mai   `<NIM--==--U>`

náði	<NIF--==--U>
neðra	<NIU--==--U>

Indeclineable proper nouns

Vestan_Garða	<NOM--==--U>
Vatnaskarðar	<NOU--==--U>
Týr	<NOM--==--U>
Twitter	<NOU--==--U>
Nattestad	<NOU--==--U>
Fuglø	<NOU--==--U>

Examples from text material of RAVNUR:

Í eldri tíð fingu børnini eftirnøvn, sum søgdu, hvør ið faðirin var.
'In the past, children were given surnames that referred to their father'
tíð <NCFS[ADN]==IUU> 'time'
børnini <NCNP[AN]==DUU> 'the children'
eftirnøvn <NCNP[AN]==IUU> 'surnames'
faðirin <NCMS[AN]==DUU> 'the father'

Fyrsta føroyska skaldsøgan, sum varð skrivað á føroyskum, var Bábelstornið hjá Regini í Líð.
'The first Faroese novel written in Faroese was Regin í Líð's The Tower of Babel.'
skaldsøga <NCFSN==IUU> 'novel'
Bábelstornið <NPNS[AN]==DUU> 'The Tower of Babel'
Regini <NPMSD==IUU> 'the name Regin'
í_Líð <NOF--==--U> 'the name í Líð'

Verbs

CatGram 1 Verb	SsCatGram 2 Verb category	3 Mood	4 Tense	5 Person	6 Number	7 Gender	8 Definiteness	9 Transcat	10 Voice	11 Case	12 Register
<V>	Main <A>	Indicative <D>	<R, A>	<1, 2, 3>	<S, P>	-	-	-	<A, M>	-	<U, O, F>
		Infinitive <F>	-	-	-	-	-	-	<A, M>	-	-
		Supine <S>	-	-	-	-	-	-	<A, M>	-	-
		Imperative <M>	-	-	<S, P>	-	-	-	<A, M>	-	-
		Subjunctive <J>	-	-	-	-	-	-	-	-	-
		Participle <P>	Present <R>	-	-	-	-	<A, R, U>	-	-	-
			Past <A>	-	<S, P>	<M, F, N>	<I, D>	<A, R, U>	-	<N, A, D, G>	<U, O>

Table 3: Overview of the Faroese PAROLE tag system for verbs

The verbs have the most complex grammar and tagsystem. Verbs are only subcategorized into **main verbs.**

Verbs are tagged for mood, tense, person, number, gender, definiteness, transcategorization, voice, case and register. However, not all forms of verbs are marked for all of these features**.** The different moods have different tagging strings, which we will go into detail about later.

First, there will be a few sections on theory. After that, we go through each of the mood's tagging schemes.

Main verbs

Regarding the **main verbs**: in the Danish Parole (Keson 1999), the SsCatGram category divides verbs into either main verbs or medial verbs, whilst both main and medial verbs can be further categorized as being either in active or passive voice. We have chosen a different approach, however. Seeing as passive voice in Faroese is most commonly being used with the auxiliary verbs *verða/blíva* + past participle of the main verb (e.g. *Báturin varð málaður (av henni)* "The boat was painted (by her)"), it does not make sense to categorize the verbs according to active/passive voice (as in Danish, e.g. *Båden mal*_es_). However, Faroese does have an -st form similar to the Danish -s form, only this ending most commonly has a mediopassive voice rather than a passive voice. Therefore, we categorize voice according to active/mediopassive instead of according to active/passive. Furthermore, the mediopassive -st ending is quite productive and can be formed from most inflectional forms by adding -st after the relevant inflectional ending (Thráinsson et al. 2012) There are some exceptions, however e.g. unergative and unaccusative intransitive verbs + auxiliaries and modal verbs cannot inflect in the mediopassive. The -st form in Faroese does not always have a mediopassive reading, but can also have a reflexive, reciprocal or modal reading. It is common in Faroese linguistics to categorize all of these readings of -st forms as mediopassive (Thráinsson et al. 2012), and we do the same. Therefore, we have chosen to distinguish between active and mediopassive voice in the 11th tagline considering voice and not distinguish between main and medial verbs in SsCatGram. This is also the manner in which Icelandic -st forms are being tagged (in a different tagset than Parole, but the Icelandic and Faroese grammar are more similar than the Danish and Faroese in this regard (Stofnun Árna Magnússonar í íslenskum fræðum 2020)).

The moods

The Faroese verbs are tagged further according to mood, which takes position three. The Faroese verbs have very different conjugations in the various moods. Note that the tag system is different for each mood (it is not the case for any other word class that the tag system is different for the various values in position 3-11). This is <u>not</u> to be understood as a true subcategorisation of verbs, as would be in the second position (SsCatGram). The conjugational tags start with the moods in third position. The various tag systems of the moods will be listed and described in their respective sections.

Furthermore, Faroese verbs in the participle mood conjugate very differently in the two tenses, and we have therefore made the tag system different for Present tense and Past tense (fourth position) in the Participle Mood. We have done this to be able to tag the present participle and past participle separately. If we were to categorize verbs by mood only, we would have to assign the present participle form with U (unmarked) in all positions where it is indeclinable and the past participle has declension, i.e. *<VAPR-UUU[ARU]-UU>. Instead, we categorize verbs according to tense and can therefore assign all present participle forms with the tag <VAPR----[ARU]--->.

Note that this setup with tag systems being different based on the value in third and fourth position is unique for the verb category. But the complexity of the verbal conjugation in Faroese demands a more detailed system, and this way the grammar is reflected more realistically.

The Danish Parole includes **gerund** (gerundium/kentaurkonstruktioner) as a mood. Although we can see why the gerund forms are included in the Danish Parole, we do not think it necessary in the Faroese. In the Danish Parole, the gerund form is a kind of verbal noun, applying the ending -en to essentially any verbs, e.g. *denne gøen natten lang* 'this barking all night long'. This would be an -an ending in Faroese, e.g. *ein skríggjan og rópan* 'screaming and shouting'. As is evident from this example, the construction *does* occur in Faroese, however, seeing as there is limited literature about this construction in Faroese, we have chosen to treat these constructions as common nouns. Many common nouns in Faroese do have an -an ending and function as regular nouns, e.g. *hugsan* 'thought', *skipan* 'system' etc. Some of these can function as both regular nouns and as gerunds, i.e. as a gerund they would have the meaning of 'an ongoing action'. It would be rather time-consuming to have to tag every -an word as either a noun or a gerund depending on the context, and therefore, together with the fact that there is no tradition in Faroese to categorize these as verbs, we have chosen not to categorize verbs as gerunds, but to categorize these words as nouns.

Other

Auxiliaries and **modal** verbs do <u>not</u> get an explicit tag, but rather they are tagged as any other verb.

We do not bring compound verbs together. Rather, each word in a compound verb gets its own analysis, e.g. *hevur kunnað fingið* 'could have gotten' and *fara upp* 'get up'.

Main verbs

The indicative forms

CatGram 1 Verb <V>	SsCatGram 2 Verb category Main <A>	3 Mood Indicative <D>	4 Tense <R, A>	5 Person <1, 2, 3>	6 Number <S, P>	7 Gender -	8 Definiteness -	9 rcategorizatio -	10 Voice <A, M>	11 Case -	12 Register <U, O, F>

As shown in the table above, the indicative forms are tagged for tense, person, number and voice.

Tense: **present tense (R) and past tense (A)**

Indicative verbs can inflect in either the present tense or the past tense, e.g. Present tense: *eg komi* <VAD<u>R</u>1S---A-U> 'I am coming' and Past tense: *eg kom* <VAD<u>A</u>1S---A-U> 'I came'.

Person: **1st person, 2nd person and 3rd person (1, 2, 3)**

Indicative verbs also inflect for person, e.g. 1st person: *eg komi* <VADR<u>1</u>S---A-U> 'I am coming', 2nd person: *tú kemur* <VADR<u>2</u>S---A-U> 'you are coming' and 3rd person: *hon kemur* <VADR<u>3</u>S---A-U> 'she is coming'.

Number: **singular and plural (S, P)**

Indicative verbs can inflect in either singular or plural, e.g. singular: *eg komi* <VADR1<u>S</u>---A-U> 'I am coming', plural: *vit koma* <VADR1<u>P</u>---A-U> 'we are coming'.

Voice: **Active and mediopassive (A, M)**

Lastly, indicative verbs are also marked for voice in either active or mediopassive, e.g. active: *tað kemur* <VADR3S---<u>A</u>-U> 'it comes', passive: *tað kemst av* <VADR3S---<u>M</u>-U> 'it is due to'.

Register: **Unmarked, obsolete and formal (U, O, F)**

In old texts or very formal speech (e.g. Biblical speech), an old inflection of 1. person plural present tense can be used, e.g. *vær biðjum teg* (vs. *vit biðja teg*) 'we ask of you/we pray'. This -um form will be marked as obsolete. Also, the imperative verbs originally inflected in both the 1st and 2nd person plural. However, in modern Faroese, only the 2nd person plural is being used. The 1st person plural inflection can be found in old texts or very formal speech or writing, e.g. *latum okkum øll biðja* 'let us all pray'. The ending is the same in both the old indicative and the old imperative, namely *-um*. This makes it very hard to figure out when the form is used in the imperative or in the indicative mood. Therefore, all of these forms will be tagged in the indicative mood with obsolete register.

Although the subject generally is in the nominative case in Faroese, some verbs can take a non-nominative subject in the indicative mood, e.g. *mær dámar dunnu* 'I-DAT like-3sg duck' and *meg lystir at dansa* 'I-ACC want-3sg to dance'. A dative subject changes the verb's inflection (e.g. *mær dámar dunnu* 'I-DAT like-3sg duck' vs. *eg dámi dunnu* 'I-NOM like-1sg duck'), while the inflection remains the same in a sentences with a nominative or accusative subject, i.e. e.g. *meg droymdi* 'I-ACC dream-1sg' and *eg droymdi* 'I-NOM dream-1sg'.
According to Thráinsson et al. (2012, p. 257), "some non-nominative subjects in Faroese are on the

way out and others have changed case". The subject case is <u>not</u> indicated in the tagset for verbs. But the verbs that can be uttered with either nominative or dative subject can have two different indicative conjugations, though belonging to the same lemma. To differ these forms we use the formal register tag. A verb form is tagged with formal register (F) only if it takes a dative subject AND the conjugation is different than the conjugation with a nominative subject.

All other forms are marked as unmarked (U).

The infinitive form

Verb	Verb category	Mood	Tense	Person	Number	Gender	Definiteness	rcategorizatio	Voice	Case	Register
<V>	Main <A>	Infinitive <F>	-	-	-	-	-	-	<A, M>	-	-

The infinitive verbs only inflect in voice. The infinitive forms can occur with or without the infinitive marker *at*, e.g. *tað kann ikki <u>lata</u> seg gera* 'it can't be done' and *eg fari <u>at renna í dag</u>* 'I'm going to run today'. The infinitive verb and the infinitive marker are not treated as a compound verb, but are analysed separately.

Voice: **Active and mediopassive (M)**

The infinitive verbs can occur both in the active and mediopassive voice, e.g. active: *siga* <VAF------A--> 'to say' and mediopassive: *sigast* <VAF------M--> 'to be said' .

The participle forms

Verb	Verb category	Mood	Tense	Person	Number	Gender	Definiteness	rcategorizatio	Voice	Case	Register
<V>	Main <A>	Participle <P>	Present <R>	-	-	-	-	<A, R, U>	-	-	-
			Past <A>	-	<S, P>	<M, F, N>	<I, D>	<A, R, U>	-	<N, A, D, G>	<U, O>

The participle forms can be used either in their verbal forms, as attributive or predicate adjectives, or adverbs. The past participles inflect like adjectives, i.e. in tense, number, gender, definiteness, trans-categorisation, case and register. When past participles are used as adjectives, number, gender and case agree with the subject.

The present participle does not inflect in number, gender, definitess and case. Therefore, the present participle will get an analysis as <VAPR----[ARU]--->, meaning that only CatGram, SsCatGram, mood, tense and transcategorisation are relevant for the present participle, while the other categories get an analysis as being irrelevant.

Note that if the participle does not exist as an infinitive form, it will be categorised an adjective instead of VA, e.g. *ótolandi* 'unbearable' (*ótola* doesn't exist as a verb) and *avhølvaður* 'beaten up' (*avhølva* doesn't exist as a verb, only as an open compound verb '*at hølva av*').

Tense: **Present (R) and past (A)**

The participle forms are subdivided into Present (R) and Past (A), and the two are tagged differently.

The present participles do not inflect at all, e.g. *ein sovandi hundur* <VAP<u>R</u>----[ARU]---> 'a <u>sleeping</u> dog'. However, it can have different trans-categorizations, depending on the context (see the section *Trans-categorization* for further details).

The past participle inflects in all of the categories, e.g. *ein tikin bilur* <VAP<u>A</u>-SMIA-NU> 'a car which has been <u>taken</u>'.

Number: **Singular or plural (S, P)**

The past participles inflect in number, e.g. singular: *ein tikin bilur* <VAPA-<u>S</u>[MF]IA-NU> 'a car which has been taken', plural: *nakrir tiknir bilar* <VAPA-<u>P</u>MIA-NU> 'some taken cars'.

Gender: **Masculine, feminine and neuter (M, F, N)**

Past participles also inflect in gender, e.g. masculine: *hann er sagdur at vera vakur* <VAPA-S<u>M</u>IU-NU> 'he is <u>said</u> to be beautiful', feminine: *hon er søgd at vera vøkur* <VAPA-S<u>F</u>IU-NU 'she is <u>said</u> to be beautiful', neuter: *tað er sagt at vera vakurt* <VAPA-S<u>N</u>IU-NU> 'it is <u>said</u> to be beautiful'.

Definiteness: **Definite and indefinite (D, I)**

Past participles inflect for definiteness as well, e.g. indefinite: *ein tikin bilur* <VAPA-SM<u>I</u>A-NU> 'a car which has been <u>taken</u>', definite: *tann tikni bilurin* <VAPA-SM<u>D</u>A-NU> 'the car which has been <u>taken</u>'.

Case: **Nominative, accusative, dative and genitive (N, A, D, G)**

Past participles also inflect in case, e.g. nominative: *har er ein stolin bilur* <VAPA-SMIA-<u>N</u>U> 'there is a <u>stolen</u> car', accusative: *eg sá ein stolnan bil* <VAPA-SMIA-<u>A</u>U> 'I saw a <u>stolen</u> car', dative: *eg koyrdi við einum stolnum bili* <VAPA-SMIA-<u>D</u>U> 'I drove a <u>stolen</u> car'.

*Register***: Unmarked or obsolete (U, O)**

As with the other word classes (nouns, adjectives, pronouns), the genitive inflection in participles is marked as being obsolete, while the other cases are marked as unmarked. Past participles inflect like adjectives and are therefore treated in the same manner as adjectives regarding register. And seeing as adjectives in the genitive case are marked as obsolete, so are the past participles in genitive case (see more in the section on Obsolete Register at the end of this document).

Trans-categorization: **Adjectival, adverbial and unmarked use (A, R, U)**

The participles can have adjectival, adverbial or unmarked (i.e. verbal) use. In our lexicon, all the participles are marked for all three trans-categorization options, i.e. [ARU], due to the lack of linguistic context in a lexicon. However, when tagging a corpus, it is possible to infer the use from the context. Here are some examples:

Present participle

Unmarked use: *eg komi gangandi* <VAPR----<u>U</u>---> 'I come on foot'

Adjectival use: *her er brennandi sól* <VAPR----<u>A</u>---> 'there is burning sun here'

Adverbial use: *vegurin er <u>brennandi</u> heitur* <VAPR----<u>R</u>---> 'the road is piping hot'

The present participle is only trans-categorized as an adjective when used attributively in front of a noun (e.g. *brennandi sól*), and only as an adverbial when used attributively in front of an adjective (e.g. *brennandi heitur*). Otherwise it is tagged as U (*svarandi til* 'corresponding to').

Past participle

Unmarked use:
tey stóðu <u>bundin</u> <VAPA-PNIU-NU> 'they were bound'
tað er <u>bundið</u> <VAPA-SNIU-NU> 'it is bound'

Adjectival use:
ein <u>bundin</u> hundur <VAPA-SMIA-NU> 'a bound dog'
tann <u>bundna</u> kaninin <VAPA-SFDA-NU> 'the bound/knitted bunny'

Adverbial use:
tað er <u>avgjørt</u> neyðugt <VAPA-SNIR-NU> 'it is definitely necessary'

Past participles are tagged as adjectives when they occur attributively in front of a noun (e.g. *ein bundin hundur*) and when inflected for case. Unmarked use only occurs in the nominative and not in front of a noun, while adjectival use can occur in all four cases.

They are tagged as adverbs when they occur attributively modifying an adjective (e.g. *tað er avgjørt neyðugt*). However, the adverbial use seems to be very rare.

Nominal use

Sometimes words from other word classes seem to have a nominal use, like verbs (*tann <u>lesandi</u> svav í hjallinum* 'the student slept in the storehouse') and adjectives (*tey <u>øvundsjúku</u> liva styttri* 'the envious live shorter lives'), but we choose to analyse these as nominal phrases, where the noun itself has left an empty space. So the verbs and adjectives complementing the 'invisible' noun (*tann lesandi 'persónurin' svav í hjallinum* 'the studying person slept in the storehouse', *tey øvundsjúku 'fólkini' liva styttri* 'the envious people live shorter lives') will be analysed as being merely verbs and adjectives and will <u>not</u> get a trans-categorisation for nominal use. Their analysis would therefore be as follows: *Tann <u>lesandi</u>* <VAPR----A---> 'the student'.

Verbs taking the participle

Not all kinds of verbs inflect in the participle. Generally, auxiliaries and modal verbs do not inflect in the participle, while transitive verbs (almost always) inflect in the participle (however, there are some exceptions, e.g. *trúgva*). It is a bit more complex regarding the intransitive verbs; intransitive verbs that <u>only</u> take *hava* as their auxiliary verb do *not* inflect in the participle (e.g. *hann hevur yvirlivað* 'he has survived' vs. *hann er *yvirlivdur* 'he is *survived'). Auxiliary *hava* typically governs the supine form in the following verb. While intransitive verbs meaning change of location (*teir til Íslands farnu (past ptc.) menninir* 'the to-Iceland-gone men' or 'the men, who have gone to

Iceland') or change of state (*ísurin er tiðnaður (past ptc.)* 'the ice is melted') taking *vera* (amongst others) as their auxiliary verb do inflect in the participle.

The supine form

Verb	Verb category	Mood	Tense	Person	Number	Gender	Definiteness	rcategorizatio	Voice	Case	Register
<V>	Main <A>	Supine <S>	-	-	-	-	-	-	<A, M>	-	-

The supine forms only inflect for voice. Supine is constructed with the auxiliary *hava* + the supine and is most commonly used to indicate perfect tense, e.g. *hann hevur <u>arbeitt</u>* <VAS------A--> 'he has been <u>working</u>' (note, that the auxiliary *vera* + main verb inflects as a past participle and not as the supine, e.g. *hann er <u>vorðin</u> ríkur* <VAPA-SMIU-NU> 'he has <u>become</u> rich'). Modal verbs in the past tense can also take a supine complement, e.g. *hann mundi <u>dottið</u>* <VAS------A--> 'he almost fell' (Thráinsson et al. 2012, p. 76)

Voice: active and mediopassive (A, M)

The supine verb forms can be in either active or mediopassive voice, e.g. active: *hann hevur <u>arbeitt</u>* <VAS------A-->'he has been <u>working</u>', and mediopassive: *teir hava <u>barst</u>* <VAS------M--> 'they have been <u>fighting</u>'. Not all mediopassive words have a supine form. Only the -st forms with a reciprocal meaning inflect for mediopassive, e.g. *tey hava <u>mussast</u>* <VAS------M-->'they have been <u>kissing</u>'. But "true" mediopassives (*dyrnar <u>steingjast</u>* 'the doors are closing'), and -st forms with modal and reflexive meaning (resp. *Har kemst ikki upp* 'one can't go up there'; *Jónas reistist frá borðinum* 'Jónas got up from the table') cannot inflect for supine. However, some st-forms have other meanings (Thráinsson et al. 2012 p. 71) than the aforementioned meanings, an in those cases, the verbs may behave differently. Some can inflect in the supine, and others cannot.

The imperative form

Verb	Verb category	Mood	Tense	Person	Number	Gender	Definiteness	rcategorizatio	Voice	Case	Register
<V>	Main <A>	Imperative <M>	-	-	<S, P>	-	-	-	<A, M>	-	-

The imperative verbs forms only inflect in number and voice.

Number: singular and plural (S, P)

In Modern Faroese, the imperative verbs only inflect in 2nd person singular or plural, e.g. singular: *et* <VAM--<u>S</u>---A--> 'eat', plural: *etið* <VAM--<u>P</u>---A--> 'eat'.

However, in old texts or very formal speech (e.g. Biblical speech), an old inflection of 1st person plural present tense can be used, e.g. *latum okkum øll biðja* 'let us all pray'. This -um ending is similar to the obsolete 1st person plural ending in indicative voice, e.g. *vær biðjum teg* (vs. *vit biðja teg*) 'we ask of you/we pray'. This makes it very hard to figure out when the form is used in the imperative or in the indicative mood, and the linguistic context is always somewhat archaic. Therefore, all of these forms will be tagged in the indicative mood with obsolete register.

Voice: active and mediopassive (A, M)

The imperative verbs can be in either active or mediopassive voice, e.g. active: *gevið honum hatta* <VAM--S---<u>A</u>-->'give that to him', mediopassive: *gevist við hasum* <VAM--S---<u>M</u>--> 'stop doing that'. Not all mediopassive words have an imperative form. Only the -st forms with a reciprocal meaning inflect for imperative, e.g. *berjist ikki altíð!* 'do not always fight!'. But "true" mediopassives (*dyrnar steingjast* "the doors are closing'), and -st forms with modal and reflexive meaning (resp. *Har kemst ikki upp* "one can't go up there"; *Jónas reistist frá borðinum* 'Jónas got up from the table') cannot (normally) inflect for imperative. However, some st-forms have other meanings (Thráinsson et al. 2012 p. 71) than the aforementioned meanings, an in those cases, the verbs may behave differently. Some might inflect for imperative, and others cannot. In these cases, we will be relying our own native speaker judgments and/or text corpora.

The subjunctive form

Verb	Verb category	Mood	Tense	Person	Number	Gender	Definiteness	rcategorizatio	Voice	Case	Register
<V>	Main <A>	Subjunctive <J>	-	-	-	-	-	-	-	-	-

The subjunctive form ends in *-i* and does not inflect. It is mostly used in fixed expressions, e.g. *Jesus <u>fylgi</u> tær* <VAJ---------> 'Jesus <u>follow</u> you' (vs. *Jesus <u>fylgir</u> tær* <VADR3S---A-U> 'Jesus <u>follows</u> you' (Thráinsson et al. 2012, p. 68).

Verb or adjective

Some words only exist in the participle forms. It can be debated whether these words are verbs or adjectives.

Typically, these words are verbal participle (past and present) forms that are created from a verb + preposition/adverb, i.e. open compound verbs. Examples: *girða inni --> innigirdur, koma aftur --> afturkomin, støva av --> avstøvaður.* Other examples are not as systematically composed (typically present participle forms): *mannkrevjandi, núverandi, óteljandi.*

In Faroese, the conjugation of words in the participle form is identical to the declension of adjectives. The discussion of whether these words belong to one word class or the other is therefore more ideological than practical. In our work, this type of words are categorised as adjectives. Syntactical and semantic functions of this type of participle is also the same as for the adjectives, so differentiating them in i.e. syntactical analysis is very hard (and without purpose).

Participle forms that belong to a conjugational paradigm with an infinitive and indicative forms are categorised as verbs. Participle forms with no infinitive or indicative forms are categorised as adjectives.

Examples, verbs

From lexicon, RAVN_lex:

arbeiða 'to work'

arbeiða	<VAF------A-->
arbeiðast	<VAF------M-->
arbeiði	<VADR1S---A-U>
arbeiðir	<VADR[23]S---A-U>
arbeiða	<VADR[123]P---A-U>
arbeiðist	<VADR[123]S---M-U>
arbeiðast	<VADR[123]P---M-U>
arbeiddi	<VADA[123]S---A-U>
arbeiddu	<VADA[123]P---A-U>
arbeiddist	<VADA[123]S---M-U>
arbeiddust	<VADA[123]P---M-U>
arbeiðandi	<VAPR----[ARU]--->
arbeiddur	<VAPA-SMI[ARU]-NU>
arbeiddan	<VAPA-SMI[ARU]-AU>
arbeiddum	<VAPA-S[MN]I[ARU]-DU>
arbeids	<VAPA-S[MN]I[ARU]-GO>
arbeiddir	<VAPA-PMI[ARU]-NU>
arbeiddar	<VAPA-PMI[ARU]-AU>
arbeiddum	<VAPA-P[FMN]I[ARU]-DU>
arbeidda	<VAPA-P[FMN]I[ARU]-GO>
arbeidd	<VAPA-SFI[ARU]-NU>
arbeidda	<VAPA-SFI[ARU]-AU>
arbeiddari	<VAPA-SFI[ARU]-DU>
arbeiddar	<VAPA-SFI[ARU]-GO>
arbeiddar	<VAPA-PFI[ARU]-[AN]U>
arbeitt	<VAPA-SNI[ARU]-[AN]U>
arbeidd	<VAPA-PNI[ARU]-[AN]U>
arbeiddi	<VAPA-SMD[ARU]-NU>
arbeidda	<VAPA-S[MN]D[ARU]-[AD]U>
arbeidda	<VAPA-S[MN]D[ARU]-GO>
arbeiddu	<VAPA-P[FMN]D[ARU]-[ADN]U>
arbeiddu	<VAPA-P[FMN]D[ARU]-GO>
arbeidda	<VAPA-S[FN]D[ARU]-NU>
arbeiddu	<VAPA-SFD[ARU]-[AD]U>
arbeiddu	<VAPA-SFD[ARU]-GO>
arbeitt	<VAS------A-->
arbeið	<VAM--S---A-->
arbeiðið	<VAM--P---A-->
arbeiðist	<VAM--P---M-->
arbeiði	<VAJ--------->

Adjectives

CatGram 1 Adjective	SsCatGram 2 Adjective category	3 Degree	4 Gender	5 Nunber	6 Case	7 =	8 Definiteness	9 Register	10 TrsCat	11	12
<A>	Normal <N>	<P, C, S, A>	<M, F, N>	<S, P>	<N, A, D, G>	=	<I, D>	<U, O>	<R, U>		
	Indeclineable <I>	-	-	-	-	=	-	-	<R, U>		

Table 4: Overview of the Faroese PAROLE tag system for adjectives

Adjectives get the word class tag A. Typically, adjectives name an attribute of a noun. They usually occur either as an attributive adjective (e.g. *tann gamli hundurin* "the old dog)", or as a predicative adjective (e.g. *bilurin er grønur* "the car is green"). In the tagging, we do not differentiate between attributive and predicative adjectives.

Adjectives are divided into *normal adjectives* and *indeclineable adjectives*.

Normal Adjectives

SsCatGram 2 Adjective category	3 Degree	4 Gender	5 Nunber	6 Case	7 =	8 Definiteness	9 Register	10 TrsCat	11	12
Normal <N>	<P, C, S, A>	<M, F, N>	<S, P>	<N, A, D, G>	=	<I, D>	<U, O>	<R, U>		

Normal adjectives are marked for inflection for degree, gender, number, case, definiteness and register.

Inflection

Degree:
- Positive (P)
- Comparative (C)
- Superlative (S)
- Absolute superlative (A)

(e.g. *góður, betri, best, allarbest* "good, better, best, very best")

All cases of adjectives that are not comparative, superlative or absolute superlative are marked as being positive, also in cases where the adjective is not subject to comparison, e.g. *(meira/mest) gamaldags* '(more/most) old-fashioned'.

Gender:
- Masculine (M)
- Feminine (F)
- Neuter (N)

Number:
- Singular (S)
- Plural (P)

23

Case:
- Nominative (N)
- Accusative (A)
- Dative (D)
- Genitive (G). Note, however, that the genitive form is very restricted in Modern Faroese and is only found in fixed expressions and compounds. The genitive is therefore not likely to occur very often. See also about register in adjectives below.

Definiteness:
- Definite (D)
- Indefinite (I)

Register:
- Unmarked (U)
- Obsolete (O). As mentioned above, the genitive form is very rarely used and obsolete. Therefore, all genitive forms of the adjective will be marked as O, while all other cases will be marked as U (see more in the section on Obsolete Register at the end of this document).

Transcategorisation:
- Adverbial: R
- Unmarked (adjectival): U

The Transcategorisation tag category is used when certain adjectives are used
Note, that nominal use of adjectives (where the nominal is missing) is still being marked as an adjective, e.g. 'tey gomlu' *the old ones*, 'tann triði' *the third one.*

Nominal use of participles are marked as participles (in the CatGram of verbs) trans-categorised to adjectives (A), e.g. 'tann studerandi' *the student*, 'tann ferðandi' *the traveller* (see also the section "Nominal use" (nouns) and the section "Verbs"). These indeclinable adjectives with the suffix *-andi* (present participle) are formed from verbs (for the most part), and they are therefore marked as present participle verbs trans-categorized to adjectives, when the (see section "Verbs").

Indeclineable Adjectives

CatGram 1 Adjective	SsCatGram 2 Adjective category	3 Degree	4 Gender	5 Nunber	6 Case	7 =	8 Definiteness	9 Register	10 TrsCat	11	12
<A>	Indeclineable <I>	-	-	-	-	=	-	-	<R, U>		

Indeclineable adjectives are not tagged further than the SsCatGram position.
Examples: *felags* <AI----=--[RU]> 'common, together', *óbeinleiðis* <AI----=--[RU]> 'indirect', *super* <AI----=--[RU]> 'super', *eyðsæð* <AI----=--[RU]> 'evident'.

This category also includes present participle verbal forms from open compound verbs with an adjectival function, see more under the sections about verbs. Examples: *ótolandi* <AI----=--[RU]> 'unbearable', *heimagangandi* <AI----=--[RU]> 'a stay-at-home-person', *táverandi* <AI----=--[RU]> 'former'.

Examples, adjectives

Examples from Lexicon, RAVN_lex:

fastur 'solid, attached'

fastur	<ANPMSN=IU[RU]>
fastan	<ANPMSA=IU[RU]>
føstum	<ANP[MN]SD=IU[RU]>
fasts	<ANP[MN]SG=IO[RU]>
fastir	<ANPMPN=IU[RU]>
fastar	<ANPMPA=IU[RU]>
føstum	<ANP[FMN]PD=IU[RU]>
fasta	<ANP[FMN]PG=IO[RU]>
fastra	<ANP[FMN]PG=IO[RU]>
føst	<ANPFSN=IU[RU]>
fasta	<ANPFSA=IU[RU]>
fastari	<ANPFSD=IU[RU]>
fastar	<ANPFSG=IO[RU]>
fastrar	<ANPFSG=IO[RU]>
fastar	<ANPFP[AN]=IU[RU]>
fast	<ANPNS[AN]=IU[RU]>
føst	<ANPNP[AN]=IU[RU]>
fasti	<ANPMSN=DU[RU]>
fasta	<ANP[MN]S[AD]=DU[RU]>
fasta	<ANP[MN]SG=DO[RU]>
føstu	<ANP[FMN]P[ADN]=DU[RU]>
føstu	<ANP[FMN]PG=DO[RU]>
fasta	<ANP[FN]SN=DU[RU]>
føstu	<ANPFS[AD]=DU[RU]>
føstu	<ANPFSG=DO[RU]>
fastari	<ANC[FMN][PS] [ADN]=IU[RU]>
fastari	<ANC[FMN][PS]G=IO[RU]>
fastari	<ANCMSN=DU[RU]>
fastara	<ANC[MN]S[AD]=DU[RU]>
fastara	<ANC[MN]SG=DO[RU]>
fastaru	<ANC[FMN]P[ADN]=DU[RU]>
fastaru	<ANC[FMN]PG=DO[RU]>

fastara	<ANC[FN]SN=DU[RU]>
fastaru	<ANCFS[AD]=DU[RU]>
fastaru	<ANCFSG=DO[RU]>
fastastur	<ANSMSN=IU[RU]>
fastastan	<ANSMSA=IU[RU]>
fastastum	<ANS[MN]SD=IU[RU]>
fastasts	<ANS[MN]SG=IO[RU]>
fastastir	<ANSMPN=IU[RU]>
fastastar	<ANSMPA=IU[RU]>
fastastum	<ANS[FMN]PD=IU[RU]>
fastasta	<ANS[FMN]PG=IO[RU]>
fastastra	<ANS[FMN]PG=IO[RU]>
fastast	<ANSFSN=IU[RU]>
fastasta	<ANSFSA=IU[RU]>
fastastari	<ANSFSD=IU[RU]>
fastastar	<ANSFSG=IO[RU]>
fastastrar	<ANSFSG=IO[RU]>
fastastar	<ANSFP[AN]=IU[RU]>
fastast	<ANSN[PS][AN]=IU[RU]>
fastasti	<ANSMSN=DU[RU]>
fastasta	<ANS[MN]S[AD]=DU[RU]>
fastasta	<ANS[MN]SG=DO[RU]>
	<ANS[FMN]P[ADN]=DU[RU
fastastu	]>
fastastu	<ANS[FMN]PG=DO[RU]>
fastasta	<ANS[FN]SN=DU[RU]>
fastastu	<ANSFS[AD]=DU[RU]>
fastastu	<ANSFSG=DO[RU]>

Pronouns

CatGram 1 Pronoun	SsCatGram 2 Pronoun category	3 Person	4 Gender	5 Number	6 Case	7 Poss. number	8 Reflexive	9 Register	10	11	12
<P>	Demonstrative <D>	-	<M, F, N>	<S, P>	<N, A, D, G>	-	-	<U, O>			
	Indefinite <I>	-	<M, F, N, U>	<S, P>	<N, A, D, G>	-	-	<U, O>			
	Interrogative/Relative <T	-	<M, F, N>	<S, P>	<N, A, D, G>	-	-	<U, O>			
	Personal <P>	<1, 2, 3>	<M, F, N, U>	<S, P>	<N, A, D, G>	-	<Y, N>	<U, F>			
	Possessive <O>	<1, 2, 3>	<M, F, N>	<S, P>	<N, A, D, G>	<S, P>	<Y, N>	<U, O, F>			

Table 5: Overview of the Faroese PAROLE tag system for pronouns

The table above is the PAROLE tagset for pronouns in Faroese. Pronouns are a group of typically short words that can function by themselves as noun phrases and refer to either participants in the discourse (*eg, tú* 'me, you' etc.) or something mentioned elsewhere in the discourse (*hon, tað* 'she, it' etc.). Pronouns are subcategorised into: *Demonstrative, Indefinite, Interrogative/Relative, Personal, Possessive*. We will go through the different subcategories of pronouns in the following sections.

Demonstrative pronouns

CatGram 1 Pronoun	SsCatGram 2 Pronoun category	3 Person	4 Gender	5 Number	6 Case	7 Poss. number	8 Reflexive	9 Register	10	11	12
<P>	Demonstrative <D>	-	<M, F, N>	<S, P>	<N, A, D, G>	-	-	<U, O>			

Demonstrative pronouns are marked for *gender* (M, F, N), *number* (S, P), *case* (N, A, D, G) and *register* (O, U).

The Faroese demonstrative pronouns are: *tann, hesin, hasin, hin* (in all their inflectional forms) 'that/the, this, that, the other'.

The forms *tann* <PD-MSD--O>, *tá* <PD-MPA--O> and *teim* <PD-[FMN]PD--O> are accusative and dative forms of the pronoun *tann* <PD-MS[AN]--U>, but they are rarely used and mostly in older, formal texts. There are other more modern forms that are used colloquially in these cases as well. The older forms are therefore marked as obsolete as well as the genitive forms, while the rest is marked as as unmarked. These decisions are based upon Thráinsson et al. (2012), who put the aforementioned obsolete forms in brackets.

It is important to notice that the Faroese PAROLE tagset does <u>not</u> include the word class of articles. The words traditionally classified as (free-standing) definite articles are identical to the pronouns (*tann, hin*), and the tagger does not differentiate between articles or pronouns. These words will be analysed as either demonstrative pronouns or personal pronouns, but never as articles. To differentiate between the analysis of demonstrative or personal pronouns in these cases, the pronouns will be tagged as demonstrative pronouns when they occur adnominally (attached to or modifying a noun), e.g. *tað fitta barnið* 'the sweet child', and they will be tagged as personal pronouns when they occur pronominally (as a 'place holder' for a noun), e.g. *eg sá tað renna úti* 'I saw it running outside'.

Examples, demonstrative pronouns

Examples from lexicon, RAVN_lex:

hasin 'that'

hasin	<PD-MSN--U>
handa	<PD-MSA--U>
hasum	<PD-[MN]SD--U>
handa	<PD-FSN--U>
hasa	<PD-FSA--U>
hasi	<PD-FSD--U>
hatta	<PD-NS[AN]--U>
hasir	<PD-MPN--U>
hasar	<PD-MPA--U>
hasum	<PD-[FMN]PD--U>
hasa	<PD-[FMN]PG--O>
hasar	<PD-FP[AN]--U>
hasi	<PD-NP[AN]--U>

hin 'the other'

hin	<PD-MS[AN]--U>
hinum	<PD-[MN]SD--U>
hins	<PD-[MN]SG--O>
hin	<PD-FSN--U>
hina	<PD-FSA--U>
hinari	<PD-FSD--U>
hinnar	<PD-FSG--O>
hitt	<PD-NS[AN]--U>
hinir	<PD-MPN--U>
hinar	<PD-MPA--U>
hinum	<PD-[FMN]PD--U>
hinna	<PD-[FMN]PG--O>
hinar	<PD-FP[AN]--U>
hini	<PD-NP[AN]--U>

Indefinite pronouns

CatGram 1 Pronoun	SsCatGram 2 Pronoun category	3 Person	4 Gender	5 Number	6 Case	7 Poss. number	8 Reflexive	9 Register	10	11	12
<P>	Indefinite <I>	-	<M, F, N, U>	<S, P>	<N, A, D, G>	-	-	<U, O>			

Indefinite pronouns are analysed for *gender* (M, F, N, U), *number* (S, P), *case* (N, A, D, G) and *register* (O, U).

Examples of indefinite pronouns are: *onkur* 'some(one)', *nakar* 'anybody, *eingin* 'nobody', *hvørgin* 'neither of two', *annar* 'other', *annarhvør* 'every other/second', *ein* 'one, you', *báðir* 'both', *sjálvur* 'self', *summur* 'some', *hvør* 'each, every', *mann* 'you, one' (in all their inflectional forms, here written in masc-sing-nom). (Note that the indefinite pronoun *hvør* is identical to the

interrogative pronoun *hvør;* except that as an indefinite pronoun, the neuter form in nominative and accusative singular is *hvørt*, while the neuter form of the interrogative pronoun is *hvat*). The genitive forms of the indefinite pronouns are marked as obsolete.

When working with indefinite pronouns, it is important to remember that the tagset does not include articles, and that words traditionally classified as indefinite articles (*ein, eitt*) will be tagged as indefinite pronouns.

Examples, indefinite pronouns

Example from lexicon, RAVN_lex:

nakar 'anyone, anybody'

nakar	<PI-MSN--U>
nakran	<PI-MSA--U>
nøkrum	<PI-[MN]SD--U>
nøkur	<PI-FSN--U>
nakra	<PI-FSA--U>
nakrari	<PI-FSD--U>
nakað	<PI-NS[AN]--U>
nakrir	<PI-MPN--U>
nakrar	<PI-[FM]PA--U>
nøkrum	<PI-[FMN]PD--U>
nakrar	<PI-FPN--U>
nøkur	<PI-NP[AN]--U>

Interrogative pronouns (and relative pronouns)

CatGram 1	SsCatGram 2	3	4	5	6	7	8	9	10	11	12
Pronoun	Pronoun category	Person	Gender	Number	Case	Poss. number	Reflexive	Register			
<P>	Interrogative/Relative <T>	-	<M, F, N>	<S, P>	<N, A, D, G>	-	-	<U, O>			

Interrogative pronouns are analysed for *gender* (M, F, N), *number* (S, P), *case* (N, A, D, G) and *register* (U, O).

The interrogative pronouns are the following: *hvør* <PT-[FM]SN--U> 'who' and *hvat* <PT-NS[AN]--U> 'what'. The obsolete forms are genitive forms. (Open form) compound interrogative pronouns are tagged separately, e.g. in <u>*hvat fyri*</u> *eina gentu sást tú* 'which girl did you see', *hvat* and *fyri* are tagged as separate words.

True relative pronouns do not exist in Faroese, only relative particles (*sum, ið* - and these are analysed as conjunctions). The interrogative pronouns in their genitive form can sometimes be used to introduce a relative clause and therefore this category includes both interrogative and relative pronouns (example where "hvørs" can be analysed as a relative pronoun: *Ísland,* <u>*hvørs fíggjarstøða er ring*</u> <PT-[MN]SG--O>'Iceland, <u>whose</u> financial position is bad'). It can be hard to distinguish

between the use of a pronoun as interrogative or relative, and the difference is not specified in the analysis.

Personal pronouns

CatGram 1 Pronoun <P>	SsCatGram 2 Pronoun category Personal <P>	3 Person <1, 2, 3>	4 Gender <M, F, N, U>	5 Number <S, P>	6 Case <N, A, D, G>	7 Poss. number -	8 Reflexive <Y, N>	9 Register <U, F>	10	11	12

Personal pronouns are analysed for *person* (1, 2, 3), *gender* (M, F, N, U) *number* (S, P), *case* (N, A, D, G), *reflexiveness* (Y, N) and *register* (U).

The personal pronouns are *eg, tú, tygum, hann, hon, tað* (in all their inflectional forms). The unmarked gender is only used for 1. and 2. person personal pronouns.

As mentioned, the definite article *tað (tann)* is either tagged as demonstrative pronouns (when occurring in an adnominal position) or personal pronouns (when occurring in a pronominal position).

The pronouns *mín, tín, tygara, okkara, tykkara, hansara, hennara, tess, teirra* (the genitive forms) can be used in possessive constructions (e.g. *tað er <u>hennara</u> kaka* <PP3FSG-NU> 'it is <u>her</u> cake'). But these are in fact personal pronouns, even when occuring in a possessive construction, and will be tagged as such. See more in the section on Possessive pronouns.

Note also that there is a frequent possessive construction in Faroese consisting of *hjá* + dative, e.g. *hetta er bókin hjá mær* 'this is my book' (cf. *hetta er bók mín/hetta er mín bók*, where *mín* is a possessive pronoun). While the construction is possessive, the pronoun *mær* (or any other dative pronoun used in this construction) is not a possessive pronoun, but a personal pronoun.

All personal pronouns are unmarked for register (U), with the exception of *tygum* (and all inflectional forms thereof), which are analysed with a formal register (F).

The personal pronouns are analysed for reflexiveness. Faroese has a special reflexive pronoun in the third person only (*seg, sær, sín* (it doesn't exist in the nominative)), which will be analysed as a personal pronoun with a reflexive attribute. The forms *seg* (accusative) and *sær* (dative) will always be tagged as a personal pronoun with a 'Y' for reflexiveness, and the form *sín* (genitive) will be analysed as a personal pronoun with a 'Y' for reflexiveness only when it occurs as a nominal complement in a prepositional phrase (*til sín*) and as a possessive pronoun in other constructions. See also the section about possessive pronouns to read more about the reflexive pronouns.

Note that the reflexive pronouns *seg, sær, sín* often occur with the indefinite pronoun *sjálvur* (in its corresponding inflection). *Sjálvur* will be analysed separately as an indefinite (non-reflexive) pronoun and not together with the reflexive pronouns.

Possessive pronouns

CatGram 1 Pronoun <P>	SsCatGram 2 Pronoun category Possessive <O>	3 Person <1, 2, 3>	4 Gender <M, F, N>	5 Number <S, P>	6 Case <N, A, D, G>	7 Poss. number <S, P>	8 Reflexive <Y, N>	9 Register <U, O, F>	10	11	12

Possessive pronouns are analysed for *person* (1, 2, 3)*, gender* (M, F, N)*, number* (S, P)*, case* (N, A, D, G)*, possessor* (S, P)*, reflexiveness* (Y, N) and *register* (F, O, U).

The possessive pronouns in Faroese are as follows: *mín, tín, sín, vár* (and the inflectional forms thereof).

Gender marks the gender of the thing **possessed** and not the gender of the possessor, e.g. *mítt barn* <PO1<u>N</u>S[AN]SNU> 'my child', *mín hundur* <PO1<u>M</u>S[AN]SNU> 'my dog'.
Number in position 5 marks the number of the **possessor**, e.g. *míni børn* <PO1N<u>S</u>[AN]PNU> 'my children'. There is no example of this number in the plural, because the possessive pronouns in the plural Number are analysed as the personal pronouns in the genitive, as they are not 'true' possessive pronouns. See section on this below.
Possessive number in position 7 marks the number of the thing **possessed** and not the number of the possessor, e.g. *mín drongur* <PO1MS[AN]<u>S</u>NU> 'my boy', *míni børn* <PO1NS[AN]<u>P</u>NU> 'my children'.

Most possessive pronouns will be unmarked for register (U), with the exception of *vár* (and the inflectional forms thereof) which will be marked as formal (F).

Only *sín* (and the inflectional forms thereof) is tagged 'Y' for reflexiveness, the other possessive pronouns are not reflexive and will be tagged 'N' for reflexiveness.

It is worth noticing that the genitives of personal pronouns in third person (singular and plural) (*hansara, hennara, tess, teirra)* and first and second person (plural) (*okkara, tykkara, tygara*) can be used in possessive constructions. However, these are not possessive pronouns and will be analysed as the personal pronouns they are (PP). Furthermore, the personal pronouns do not show agreement with 'the thing possessed', while possessive pronouns do (Thráinsson et al., 2012). That is, the personal pronouns do not inflect for possessor, but the possessive pronouns do inflect for possessor. This is also the reasoning behind treating the true possessive pronouns differently from the personal pronouns in genitive used in possessive constructions. Therefore, only the true possessive pronouns will be analysed as possessive pronouns (PO).

Note that possessive constructions in Faroese are frequently made with preposition + personal pronoun, with no possessive pronoun involved (e.g. *báturin hjá mær*).

Reciprocal pronouns

The only reciprocal pronoun in Faroese is 'hvør annar' (in all its inflectional forms). However, these two words are in fact indefinite pronouns and behave and inflect as such. Therefore, the category 'reciprocal pronouns' will not be tagged for in the Faroese PAROLE.

Determiners and Articles

Some PAROLE languages include the word classes Determiners and Articles. Faroese PAROLE does <u>not</u> include the categories of Determiners and Articles.

The words traditionally classified as (free-standing) definite articles are identical to the pronouns (*tann, hin*), and the tagset does not differentiate between articles or pronouns. These words will be analysed as either demonstrative pronouns or personal pronouns, but never as articles.

Adverbs

CatGram 1 Adverb <R>	SsCatGram 2 Adverb category General <G>	3 Degree <U, P, C, S, A>	4	5	6	7	8	9	10	11	12

Table 6: Overview of the Faroese PAROLE tag system for adverbs

Adverbs are typically described as modifying the meaning of an adjective, verb, or other adverb, expressing manner, place, time or degree. Here, all adverbs are tagged as adverbs (RG), including "true" adverbs, derived adverbs with endings only used in adverbs, e.g. *-ligani* in *ordiligani*, and lexical adjectives behaving syntactically as adverbs, e.g. T*að var flott málað* 'It was beautifully painted'. See also Petersen, 2015, pp. 77-78 for other derivative endings in adverbs.

Traditionally, adverbs have different subcategories. However, here they are all categorized in one subcategory, 'general' (G). Furthermore, they are marked for degree:

- Unmarked (U) (e.g. *ikki* 'not', *eisini* 'also', *út* 'out')
- Positive (P) (e.g. *ofta* 'often', *leingi* 'for a long time')
- Comparative (C) (e.g. *oftari* 'more often', *longur* 'for a longer time'*)*
- Superlative (S) (e.g. *oftast* 'most often', *longst* 'for the longest time')
- Absolute superlative (A) (e.g. *allarsíðst* 'the very last', *hann stendur <u>allarfremst</u> í køini* 'he is at the very front of the row')

As opposed to the adjectives, all adverbs not being inflected for either comparative, superlative or absolute superlative are marked as unmarked (U), unless they *can* inflect in degree, then they are marked as positive (P).

Note that some words are both marked as adjectives and adverbs in the dictionary, e.g. *flottur* *<ANPMSN=IU>*'pretty, beautiful' and *flott* *<RGP>*'pretty, beautiful'. In the case of corpus tagging, the context will have to decide, whether it should be marked as and adverb or an adjective.

Examples, adverbs

Examples from lexicon, RAVN_lex:

áðrenn   <RGU>
'before'

eftir   <RGU>
'after'

borgarliga   <RGP>
borgarligari   <RGC>
borgarligast   <RGS>

33

'civil'

Adpositions

CatGram 1 Adposition <S>	SsCatGram 2 Adposition category Preposition <P>	3	4	5	6	7	8 Case <N, A, D, G>	9 Register <U, O>	10	11	12
		=	=	=	=	=					

Table 7: Overview of the Faroese PAROLE tag system for adpositions

The category of Adpositions has only one subcategory: *prepositions*.
In grammar, prepositions are a subcategory of *adpositions*, together with *postpositions* and *circumpositions*. Postpositions do usually not occur in Faroese. Circumpositions do occur in Faroese, but they are as a nature made from two words or more, and they will be analysed on a single word level regardless. Therefore the only adpositions in the Faroese PAROLE are prepositions, and will be referred to as such in the following.

Prepositions are tagged for *Case* and *Register* in position 8 and 9.

Case: Prepositions are marked for *case* (N, A, D, G). The prepositions do not inflect in case, but they can govern the case of their complement.
(N=does <u>n</u>ot govern case; A=governs <u>a</u>ccusative; D=governs <u>d</u>ative; G=governs <u>g</u>enitive).
Many Faroese prepositions can govern more than one case, and they should be written in square brackets, if there is not linguistic context to decide semantics and case.

Register: Prepositions are also marked for *register* (O, U). Obsolete register is mostly used with case governing the genitive, e.g. *teirra millum* <SP=====GO> 'between them' vs. *millum tey* <SP=====AU> 'between them' (see more in the section on Obsolete Register).

When given the linguistic context (e.g. in tagging a corpus), prepositions composed of several words (*compound prepositions* and *circumpositions*) should be analysed as:
1) **adverb + preposition** (examples: *inni í húsinum,* 'inside the house'*, niðan á fjallið,* 'up the mountain'*, vekk frá barninum,* 'away from the child').
2) **preposition + noun + preposition** (examples: *við atliti til,* 'in the interest of'*, til ampa fyri,* 'inconvenient for'*, við síðuna av,* 'beside')
3) **preposition + object of preposition + adverb** (examples: *av hundanum til,* 'unfortunate'*, fyri nógvum árum síðani,* 'many years ago'*, frá barnagarðinum av, 'from the kindergarten'*)
4) **preposition + other + noun** (examples: *fyri hennara skyld,* 'for her sake'*, í tínum stað,* 'in your place')

Composed prepositions like *innan fyri,* 'inside, within' and *oman eftir,* 'down the road/way' are at times written as a single word *innanfyri* and *omaneftir.* According to Faroese grammar they are written in two words when they precede a complement, and a single word when there is no complement and have an adverbial function. In PAROLE they are analysed as follows: when they are written in as a single word, complement or not, they will be analysed as an **adverb**. When they are written in two words, they are analysed as **adverb + preposition**, as mentioned above.

Preposition or adverb?

Some words can function as both prepositions or adverbs, depending on their context. For analysing these words, the main rule is:

1) **preposition**: the words occur together **with** a prepositional object (examples: *eg fór við gentuni* <SP=====DU> 'I went <u>with</u> the girl', *hann tók okkurt fyri høvdið* <SP=====AU> 'he put something <u>for</u> his head')

2) **adverb**: the words occur **without** a prepositional object (examples: *tekur tú okkurt við?* <RGU> 'can you bring something', *eg takki fyri* <RGU> 'thank you' (lit.: 'I thank you').

Conjunctions

CatGram 1 Conjunction <C>	SsCatGram 2 Conjunction category	3	4	5	6	7	8	9	10	11	12
	Coordinating <C>										
	Subordinating <S>										

Table 8: Overview of the Faroese PAROLE tag system for conjunctions

Conjunctions are tagged for *subcategory* only (C, S).

Conjunctions are a closed word class. It consists of small words used to connect clauses or sentences or to coordinate words in the same clause. Conjunctions are traditionally divided into two subcategories. PAROLE marks conjunctions for one of two subcategories:

Coordinating: *og, ella, men, jamen, tí* <CC> 'and, or, but, but, because'

Subordinating: *at, so, tá, har, um, nær, hví, hvussu, sum, ið, meðan, vissi, fyri, hóast, líkasum, síðst, beinanvegin, uttan....* <CS> 'that, so, when, because, there, if, when, why, how, that, that, while, if, for, though, like, the last time, right away, without.....

In Faroese grammar, subordinating conjunctions can be further subcategorized (into *complement conjunctions*, *relative conjunctions* and *adverbial conjunctions*), but the analysis in the tagging process will not go any further than subcategory (CS).

Conjunctions containing several words will be tagged for each word, not as a phrase.

According to Faroese linguistic tradition, *sum/ið* will be tagged as a *conjunction* and not as a *relative pronoun* (in Danish, for example, *som* is grouped as both a relative pronoun and as conjunction, depending on the context. In the Danish Parole, however, it is classified as unique (Keson 1999)). The reason for this is that relative pronouns in Faroese inflect for case, but *sum/ið* does not inflect (Thráinsson et al. 2012, p. 126), and therefore, it is more accurate to categorize it as a conjunction or particle.

Conjunction or preposition?

Some words can be both conjunctions and prepositions, for example *frá, uttan, innan, síðan, um, eftir* etc. To distinguish between the use of the words as conjunctions or prepositions, we base the decision on the context and syntax, e.g. *frá hon snakkaði, til hon tagdi* <CS>*; hon mátti rýma frá húsum* (SP=====DU).

Numerals

CatGram 1 Numeral	SsCatGram 2 Numeral category	3 Gender	4 Number	5 Case	6 =	7 =	8 Register	9	10	11	12
<M>	Cardinal <C>	-	-	-	=	=	-				
	Ordinal <O>	<M, F, N>	<S, P>	<N, A, D, G>	=	=	<U, O>				
	Distributive <D>	<M, F, N>	<S, P, D>	<N, A, D, G>	=	=	<U, O>				

Table 9: Overview of the Faroese PAROLE tag system for numerals

Numerals are divided into three subcategories: **cardinal**, **ordinal** and **distributive**. Only the ordinal and distributive inflect for gender, number, case and are marked for register (due to the genitive case, see also section on the obsolete register and the genitive).

In Faroese, the cardinal numbers 1, 2 and 3 (*eitt, tvey, trý*) are the only cardinal numbers to inflect for gender and case. These three numbers also have a distributive plural inflection (*einir, tvinnir, trinnir*). The distributive subcategory therefore consists only of these three numbers, with all their inflections. Technically, the singular inflections of 1, 2, 3 are not distributive, but the inflectional paradigm cannot be divided, so the three lemmas as a whole are tagged as distributive.

The cardinal 1 (*ein, eitt* etc.) is tagged as singular, the cardinal 2 and 3 (tvey, tveir, trý, tríggjar etc.) are tagged as plural, and all the ditributive inflections of 1, 2, 3 (*einir, tvinnir, trinnir* etc.) are tagged with the Dual number. There is a difference in meaning between the plural and the dual, e.g. *tveir skógvar* 'two shoes' vs. *tvinnir skógvar* 'two pairs of shoes'.

When tagging a dictionary/lexicon: The cardinal numbers, that can be used as pronouns, will get tags for each word class they can represent. 'Ein' can be found as a Numeral, Pronoun and Adjective.
When tagging a corpus: for practical purposes, the words ein/eitt 'one' are always tagged with their pronoun tag and not as numerals. This means that the plural einir is also marked as a pronoun, but the plural tvinnir and trinnir are marked as distributive numbers.

If the word/number *ein/eitt* is represented as the number *1* and not the word *ein/eitt*, it is marked as a numeric digit in the residual category (<XC> and <XO>).
Numerals can occur only as written words. Numeric digits and roman numerics are tagged as <XC> (cardinals) and <XO> (ordinals). Numerals occuring together with slash or colon (e.g. 1/2 ár '1/2 year') are tagged as <XR>.

Note that in the ordinals, *annar* 'second' inflects like a strong adjective, *fyrsti* 'first' inflects like a weak adjective, while the alternative form of the same word, *fyrstur* 'first', inflects like a strong adjective. All other ordinals inflect as weak adjectives.

Examples, numerals

Examples from lexicon, RAVN_lex:

tveir 'two'

tveir	<MDMP[AN]==U>
tveimum	<MD[FMN]PD==U>

tveggja	<MD[FMN]PG==O>
tvær	<MDFP[AN]==U>
tvey	<MDNP[AN]==U>
tvinnir	<MDMDN==U>
tvinnar	<MDMDA==U>
tvinnum	<MD[FMN]DD==U>
tvinna	<MD[FMN]DG==O>
tvinnar	<MDFD[AN]==U>
tvinni	<MDND[AN]==U>

 trettandi '13th'

trettandi	<MOMSN==U>
trettanda	<MO[MN]S[AD]==U>
trettanda	<MO[MN]SG==O>
trettandu	<MO[FMN]P[ADN]==U>
trettandu	<MO[FMN]PG==O>
trettanda	<MO[FN]SN==U>
trettandu	<MOFS[AD]==U>
trettandu	<MOFSG==O>

átjan	<MC---==->

 'eighteen'

trýogtrýss	<MC---==->

 'sixty-three'

Interjections

CatGram 1	SsCatGram 2	3	4	5	6	7	8	9	10	11	12
Interjection	=										
<I>	=										

Table 10: Overview of the Faroese PAROLE tag system for interjections

This word class consists of both interjections and onomatopoeic words. Both are indeclinable and can function on their own as independent phrases. Interjections are exclamations, e.g. *ja, nei, farvæl, avv, áh, ókey, ája* <I=> 'yes, no, goodbye, ouch, oh, okay, ah'. Onomatopoeia are words that mimic a sound, e.g. *mæ, mjavv, bipp, krassj, huff, suff, áh, tuff tuff, atju* <I=> 'baa, miaow, beep, crash, phew, sigh, oh, choo-choo, achoo' etc.

Residual

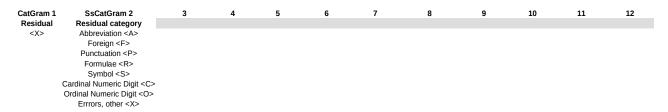| CatGram 1 Residual <X>	SsCatGram 2 Residual category	3	4	5	6	7	8	9	10	11	12
	Abbreviation <A>										
	Foreign <F>										
	Punctuation <P>										
	Formulae <R>										
	Symbol <S>										
	Cardinal Numeric Digit <C>										
	Ordinal Numeric Digit <O>										
	Errrors, other <X>										

Table 11: Overview of the Faroese PAROLE tag system for the residual category

This category is a group of residual words that do not belong to any of the other categories and often have distinct features, e.g. abbreviations, foreign words, punctuation, symbols etc.

Abbreviations

CatGram 1 Residual <X>	SsCatGram 2 Residual category	3	4	5	6	7	8	9	10	11	12
	Abbreviation <A>										

The tag <XA> is given to abbreviations or acronyms, that cannot be categorised as nouns. This is typically abbreviations with symbols or punctuation, *t.d.* <XA> 'e.g.', *km/t* <XA> 'kilometres per hour', *cand.mag.* <XA> 'cand.mag., Master of Arts'.

Foreign words

CatGram 1 Residual <X>	SsCatGram 2 Residual category	3	4	5	6	7	8	9	10	11	12
	Foreign <F>										

Foreign words can be hard to categorize. Loanwords from other languages that have been somewhat implemented in the Faroese language will be treated as regular Faroese words, when appropriate, e.g. *allíkavæl* <RGU> 'still', *faktiskt* <RGU> 'as a matter of fact'. Some foreign words become integrated with Faroese inflectional patterns, like *hotdoggur* <NCMSN==IU> 'hot dog' and are categorised as nouns. This gives us a more informative morphosyntactic analysis. Note, that the version with no inflection would be tagged as *hotdog* <XF> 'hot dog'.
"Real" foreign words that might occur in quotes and/or other forms should be categorized as Foreign Words, e.g. *by the way, awkward, ever* <XF>.

In an automated tagging process, foreign nouns are treated as proper nouns if they start with a capital letter, and if they start with a small letter, they are treated as foreign words (XF). See the section "Foreign nouns" for more on foreign nouns and how to treat them.

41

Punctuation

The following marks are categorized as punctuation marks and are analysed as XP:

period	**.**	exclamation mark	**!**
two dots	**..**	dash/hyphen	**-**
three dots	**...**	left parenthesis	**(**
comma	**,**	right parenthesis	**)**
colon	**:**	quotation mark	**"**
semicolon	**;**	slash	**/**
question mark	**?**	single quotation marks	**' '**
ampersand	**&**		

Other signs that are <u>not</u> analysed as XP:

- - Group compounds are treated as one word, e.g. Gamlarætt-Skopun (a ferry route), although it is often just the last two words of the compound that are hyphenated, e.g. *fast food-matstova*. We will add an underscore to compounds consisting of multiple words and a hyphen, e.g. *fast_food-matstova* 'fast food restaurant', *toll-_og_skattaviðurskifti* 'taxes and customs affairs'.

Formulae

The term 'formulae' here means 'unknown' word forms consisting of letters and numbers and possibly punctuation marks and symbols, e.g. *1+1=2* <XR> '1+1=2', *765/55* <XR> '765/55', *98%* <XR> 'ninety-eight percent', *1/2* <XR> 'half', *5.345* <XR> 'five point three four five'.
The subcategory also includes words with no real formula, e.g. *U19 landsliðið* <XR> 'the youngster national team of age 19', *flogfar RC451* <XR> 'flight RC451' (however, if *U19 landsliðið* was hyphenated, which is the correct spelling, if would be analysed as a noun).

Symbols

CatGram 1 Residual	SsCatGram 2 Residual category	3	4	5	6	7	8	9	10	11	12
<X>	Symbol <S>										

The following symbols have been given an analysis as symbols (XS):

$ (dollar), % (percent), * (asterisk), + (plus), = (equal), § (paragraph) and ° (degree).

Minus (-) collapses with the dash under Punctuation.

Cardinal numeric digits

CatGram 1 Residual	SsCatGram 2 Residual category	3	4	5	6	7	8	9	10	11	12
<X>	Cardinal Numeric Digit <C>										

Cardinal numeric digits, e.g. *1, 6, 58* <XO> 'one, six, fifty-eigh' etc.

Ordinal numeric digits

CatGram 1 Residual	SsCatGram 2 Residual category	3	4	5	6	7	8	9	10	11	12
<X>	Ordinal Numeric Digit <O>										

Ordinal numeric digits, e.g. *1., 12.* <XO> '1st, 12th' etc.

Other

CatGram 1 Residual	SsCatGram 2 Residual category	3	4	5	6	7	8	9	10	11	12
<X>	Errrors, other <X>										

The category *Other* consists of textual errors and linguistic deviations. This could be spelling errors or other. This subcategory is mostly used in tagging a corpus or other continuous text and not so much in the lexicon, as textual errors are a phenomena of continuous text and not entries of single lemmas in a lexicon.

Below some examples are listed, that could be tagged with the <XX> (the list is not exhaustive and at this point is merely suggestions to the <XX> tag):

- **Incorrect use of capitalization**
- **Typing and spelling errors of all sorts**
- **Incorrect separation of compund words**
- **Word units combined with a dash**
- **Two identical words in a row**

43

- Dates and other inconsistent forms (*21. jan-18. feb***)**

- Compound nouns

Compound nouns consisting of nouns and *og* and a hyphen (as in the example *dagstovna-_og_skúlaskipanin* 'day-care centre and school system') are combined with underscores and lemmatised in the Residual word class with the tag XX.

- Nonsense with symbols and punctuation

Many words contain a slash as a part of an abbreviation, in numbers and dates, to imply different usage options etc, e.g. <u>*1990'ini/byrjanin*</u> *av 2000*. In these cases, the words get an XX-analysis (textual error), if it is not known to be an official abbreviation like *km/t* <XA> 'kilometres per hour'.

Unique

CatGram 1	SsCatGram 2	3	4	5	6	7	8	9	10	11	12
Unique	**=**										
<U>	=										

Table 12: Overview of the Faroese PAROLE tag system for the unique category

The words tagged as *Unique* are words that dont belong in any of the other categories and are 'unique' as such.

In our lexicon, RAVN_lex, the only word belonging to this category so far is the infinitive marker, *at*. The infinitive marker, *at,* is unique and not classified as belonging to any of the other word classes. When *at* occurs together with an infinitive verb, it is categorised as unique (U), e.g. *at sita* <U=> 'to sit'. Additionally, when *at* occurs as an infinitive marker in a past participle construction with a subjunctive meaning, it is also categorised as U, e.g. *vit áttu at vent bátinum* <U=> 'we should (have) turn(ed) the boat around'.

Note that when *at* occurs as a marker of a subordinate-clause it is categorised as a conjunction, e.g. *tú segði, at hann var farin* <CS> 'you said that he was gone'. Furthermore, the word *at* also exists as a preposition or an adverb; in those cases, the word will be categorised in these categories, respectively.

Obsolete Register and the Genitive

With the PAROLE tag system of Faroese the feature of Register is active for the word classes Nouns (N), Adjectives (A), Numbers (M), Adpositions (S), Pronouns (P) and Verbs (V). The Faroese language does not have an integrated linguistic system for indicating register. None the less, we have chosen to have the *Register* category in our tagset. There are some word forms and inflections that are are hardly ever used, and only persist because they are a part of the standardised inflectional paradigm (which was created in the process of standardising the Faroese written language in the 1850's, based on its roots of Old Norse and Icelandic), and this is what calls for the *Obsolete* Register tag. The following guidelines are recommandations from the working group behind the Faroese PAROLE. If one wishes to tag differently in another setting and use the Register-tag in another way, one is of course free to do to. After caleful deliberation, these are the guidelines for the linguistic material worked with in the process of generating the Faroese PAROLE:

For the most part the *Obsolete Register* tag is given to word forms in the *genitive* case, that are constructed from a set inflectional paradigm, rather than being word forms that speakers know and use. Thráinsson et al. (2012) use parenthesis around genitive forms of nouns and adjectives to indicate that tey are obsolete, and Petersen (2015) leaves the genitive out altogether in his Faroese Reference Grammar, indicating that the genitive form is not productive in Modern Faroese. This is especially prominent for Nouns and Adjectives and Verbs (the past participle forms).

However, it is worth noting, that not all genitive forms in Faroese are obsolete. Nouns can be part of somewhat fixed phrases with the genitive (*til songar, millum manna, til Dals, til Gjáar*), adjectives can be part of Proper Nouns (*Meginfelag Føroyskra Studenta*), some prepositions can govern the genitive case, but they usually only do so in somewhat fixed phrases (*til havs, teirra millum*).

It would be a very time consuming and somewhat arbitrary task to sort the genitive forms of these six (large) word classes into obsolete forms and non-obsolete (unmarked). All the genitive forms are therefore as a rule tagged with an Obsolete register, knowing that all forms might not be obsolete. So as of now, the Obsolete tag **cannot** be seen as a definite guarantee of a word form being obsolete, but it should be understood as a way of indicating that the genitive case as a whole generates mostly obsolete forms.

The genitive forms of pronouns (especially the personal pronouns) are not as unusual, and they are not tagged as being obsolete, as these are a small group of words in a closed word class, where all

the words are affiliated with the same register, thereby eliminating the task of sorting through obsolete and non-obsolete forms.

Summary

The Faroese PAROLE has been presented in its entirety in this document.

Faroese PAROLE was developed together with the lexicon of the RAVN Project, RAVN_lex, and the manual for the Lexicon is also relevant reading material for anyone working with PAROLE.

To see more on the procedure of corpus tagging, see the Corpus Tag Manual of the RAVN Project.

Happy tagging!