Morphological complexity in Saami: A historical-typological view

Johanna Nichols, University of California, Berkeley

1. Introduction

Saami inflection is very complex, but not evenly so in all parts of the morphology and not equally so in all the languages.[1] Complexity can be measured in various ways, of which the two most commonly used in linguistics will be used here (for kinds of complexity see Dahl 2004; Miestamo 2008; Kusters 2008; Sinnemäki 2011; Anderson 2015; Audring 2017; and many others). One is *enumerative complexity* (also known as *inventory complexity*, *taxonomic complexity*, and other terms): the count of items in a system, such as the number of phonemes in the inventory, the number of cases in noun declension, the number of tenses in verb inflection, and so on. Actually counting all of these things rapidly becomes cumbersome, and leads to comparing unlikes: a language will be considered morphologically complex if it marks local relations with a set of local cases such as inessive, illative, and elative, but not if it marks the same relations with postpositions; the answer is not adding a count of postpositions, since adpositions are usually an open class; and so on. Therefore, the count of enumerative complexity used here limits the survey to just three basic arenas: the phoneme inventory, the inflectional categories of the verb (as laid out in Bickel & Nichols 2013), and a similar inventory for nouns (Sokur & Nichols 2019) (for the whole survey see also Nichols 2019; Nichols & Bentz 2019).

The other measure is what has usually been known as *informational complexity* or *Kolmogorov complexity*, recently *canonical complexity* (Nichols 2020): the amount of information required to describe a system. The amount of information, in the form of prose discussion and listing and

[1] Research and initial drafting done while Helsinki University Humanities Visiting Professor, 2017–2019. Supported in part by a grant from the Russian Academic Excellence Project 5-100 to the Higher School of Economics, Moscow. I thank the SAALS 4 participants for comments and discussion.

annotation of different paradigms, needed to describe the Russian case system, with its four declension classes depending on the gender and arbitrary classifications and with its several instances of case syncretism (e.g., -*u* marks accusative, dative, genitive, or locative singular depending on the declension class; and the markers of accusative are -*u*, -*a*/-Ø, and -*o* in different paradigms), is much greater than what is required to lay out the set of case endings for a Saami language. This kind of complexity amounts to a measure of the non-transparency in a system, and non-transparency is what is invoked in most work seeking correlations between language complexity and sociolinguistic factors (e.g., Dahl 2004; Trudgill 2011). For this reason, though labor-intensive to survey, it is probably more valuable in the long run.

Most measures of informational complexity are computationally based (e.g., Juola 1998; Bentz 2018; Ehret & Szmrecsanyi 2016) and require large corpora. I have proposed a grammar-survey measure (Nichols 2020) making use of the fact that the needed information is minimal in a system that follows biuniqueness, the one-form-one-function ideal of structural analysis and canonicality theory (Corbett 2007, 2015, and several others), and increases as the imbalance between forms and functions increases. Thus, the Russian example just cited, with three forms for the one function of marking the accusative case, is more complex than a comparable segment of Saami case inflection: Aikio and Ylikoski (2010: 39) describe two sets of the case endings for respectively consonant-final and vowel-final stems of North Saami, but these are phonologically predictable variants and not separate declension classes. (For a comparison of Russian to Mongolian, which like Saami has a single noun declension class with phonologically predictable variants, see Nichols 2020.) Since departures from one-form-one-function are non-canonical in the approach of Corbett and others, this complexity measure can be called *non-canonicality-based complexity*, or *canonical complexity* for short.

Keeping a survey of canonical complexity manageable requires limiting the morphology surveyed to a sample of the grammar. I survey noun inflection marking the core relations of subject, direct object, indirect object, and possessor, for singular and plural numbers only; personal pronoun inflection for the same roles, again for singular and plural only; and verb inflection for argument indexation (three persons, singular and plural only) and TAM in the simplest present tense and the simplest past tense (the perfective-like one if the language has more than one). For all I consider default valence or argument marking only: subjects in the strict traditional morphosyntactic sense, which in Uralic languages means nominative case and verb agreement; direct objects in the strict sense, i.e., accusative in many languages, and not the various oblique object cases found, e.g., with verbs of emotion or perception in many languages; and so on. Among these parts of the grammar the survey counts the number of non-biunique patterns such as the number of declension classes (different sets of endings), the number of conjugation classes (likewise), the number of different stems per lexeme, the number of different stem classes in the language, the numbers of all of those that are unpredictable, and numbers of such things as syncretisms, suppletion, allomorphy, portmanteaus, and zero markers other than nominative singular.

2. Complexity levels in Saami languages

2.1. Inventory complexity

The measure of inventory complexity used here (and also used in Nichols 2019; the morphological part is used in Nichols & Bentz 2019) counts these things: in the phoneme inventory, the number of contrastive consonant types and vowel qualities; presence or absence of tones and phonation types; and syllable complexity measured as the maximum number of consonants per syllable; verb inflectional synthesis (number of categories per word) following Bickel & Nichols (2013); in noun inflection, the presence vs. absence of gender , numeral classifiers, number, and possessive classes; in syntax, the number of basic alignments and basic word orders (further measures for syntax, and measures of derivational morphology, are under development). The results for the Saami languages surveyed are shown in Table 1 below.[2]

As these results show, Saami languages vary somewhat in their complexity, and on the whole they are similar to the worldwide average and somewhat lower for morphological complexity. Their morphological inventory complexity includes, for nouns and pronouns, only case and number; also possessor person-number for nouns; and for verbs, one argument, TAM, and voice. Possessive person-number inflection is counted as absent in most Saami languages; where remnants of possessive inflection are preserved they appear to be fossils frozen for use only on certain kin terms, and not a regular inflectional category.[3]

[2] The full database for this paper will be published in an open repository, by approximately the end of 2025.

[3] Grammars give the impression that it is regular on kin terms and some body parts for Kola Saami languages – which would then have inflectional possessive morphology and also an alienability opposition, following the cross-linguistically common pattern of head marking on kin terms and/or body parts and dependent marking (genitive case of possessor) for others. However, I did not count this possessive marking as inflection. Janda and Antonsen 2016 show that loss of possessive inflection is well underway in North Saami but do not claim that it is

Table 1. Enumerative complexity in Saami languages, separately tabulated for all metrics and only morphological ones. The lower portion shows mean enumerative complexity for those Saami languages, for other Uralic languages, and for all languages in my database (N = 245).

Language	*All*	*Morphology*
South	18	7
Pite	20	8
Lule	18	7
North	18	8
Skolt	20.5	9
Kildin	21.5	9
Saami mean	19.3	8.0
Other Uralic mean	17.7	8.7
World mean	19.1	9.8

2.2. Canonical complexity

Canonical complexity in inflectional morphology, for the three parts of speech surveyed here, essentially boils down to two targets. One is the number of different inflectional classes (i.e., sets of endings) for singular inflection of nouns, singular and plural inflection of personal pronouns, and person-number inflection in the two basic verb tenses surveyed here; and the number of these that are unpredictable, i.e., lexically specified and arbitrary. For each of the three parts of speech there is generally just one inflectional class in the Saami languages, so there is no unpredictability. (Note that phonologically predictable allomorphy is not counted here, as it is automatic.) The other survey target is the number of different stem forms and their predictability, and here Saami languages have more complexity, often a great deal. For all but South Saami, there are two or three dimensions of stem complexity in inflection: patterns of consonant gradation, patterns of vowel alternation (or ablaut), and, for the Kola languages, alternation of plain and palatalized consonants. Each set of combined consonant and vowel changes defines a stem paradigm or class; for example, in Kildin Saami, basic strong grade shifts to weak in the genitive and accusative singular, and this occurs both with and

absent. Aikio & Ylikoski (2010): 70ff. describe it as present. Kert (1971) gives full paradigms for Kildin Saami without comment, but Rießler (2022: 226) describes possessives in today's speech, about two generations later, as lost except for remnants. Wilbur (2014: 108) describes them as "nearly fallen out of use" for Pite. Without frequency counts for the other languages I could only treat this as a yes/no variable and, assuming that the situation in these two smaller, more endangered ones is typical, I coded all the varieties as lacking possessive marking.

without vowel alternations, a total of two paradigm types with gradation and two without. (I provisionally lump vowel alternations into a single broad category coded as present vs. absent, which is probably an oversimplification.) (For more discussion and examples see §2.3.) Some grammars lay out the full inventory of paradigm types (e.g., Kert 1971 for Kildin Saami; Morottaja & Olthuis 2020 for Inari; Spiik 1989 for Lule; Wilbur 2015 for Pite). For Skolt Saami, Feist (2015) presents most of the inventory and notes that it is probably not complete. For Akkala, Zajkov (1987) presents the attested alternation patterns but not the possible paradigm types; the impression gained is that there is only one gradation paradigm for nouns and all nouns have gradation, but this is not stated. Tereškin (2002) describes a single gradation pattern and two palatalization patterns (which appear to be in complementary distribution) for nouns, giving the impression that this is the full inventory.

Many grammars list separately the alternations found in vowel-final and consonant-final stems, and/or stems with even vs. odd numbers of syllables. In such cases I took the total number of different patterns found among the stem shapes. (Usually, one stem shape exhibits all types and others have only a subset of those.)

In such counts it is important to distinguish between the *patterns*, on the one hand, and on the other the *factors delimiting their distribution*, or *observations that can be made on their distribution* (henceforth *distributional observations*). The patterns are the alternation types (the gradation and ablaut patterns) as they occur in different cases or person-number forms; these are what is counted here. The distributional observations are the number of syllables and whether the stem ends in a consonant or a vowel. These are not alternation patterns and not predictive or causal factors but grammatical observations about which patterns occur where. Of course, the distributional factors are descended from contextual properties that historically conditioned the alternation in the first place, i.e., they *were* conditioning factors. They may be important for language teaching and pedagogical presentations, and covering them is essential if a grammar is to be useful to historical linguistics, but they are not causal now and are not what is measured in this study.[4]

In most modern Saami languages, gradation and vowel patterns are not predictable (e.g., from the phonological environment or grammatical categories). They can be described as morphologized, or as lexically or morphologically specified. Exceptions are South Saami, which has no gradation, and North Saami, where there may be some phonological

[4] They could be used in a differently designed survey of complexity, e.g., counting the number of different constraints on the distribution of gradation patterns in different languages. By that token Akkala Saami would be maximally simple: at least for verbs, syllable count has no role in the distribution of alternation types (Zajkov 1987: 144), while in the others it does.

predictablity of gradation. (The description of North Saami in Aikio & Ylikoski (2014: §4.2) gives the impression that vowel-final and consonant-final stems are predictable from syllable count and may predict some alternation patterns, but these factors are probably best viewed as distributional observations and not as conditioning. Nickel & Sammallahti (2011: 70–71) use two gradation classes and two vowel alternation classes.)

Consider now the number of stem classes in Saamic vs. Finnish noun inflection. In Saamic they range from one in South Saami (where consonant gradation is absent) to many in Skolt or Pite. In Finnish, noun stems vary due to consonant gradation, but in Finnish the gradation is almost always predictable from the syllable structure: weak grade before a closed syllable, strong before an open syllable.[5] Vowel harmony in Finnish is phonologically predictable, while the vowel alternations of Saamic are not.

Thus, while Finnish and most Saami languages both have consonant gradation and vowel harmony or alternation, producing allomorphy in both languages, most Saami languages have high canonical complexity because the variation is not predictable, while Finnish canonical complexity is low. Table 2 compares Saami to Finnish, Russian, and English, and Table 3 gives the numbers of stem classes for nouns in the Saami languages. Table 4 gives canonical complexity levels for all the Saami languages surveyed, and Table 5 gives mean complexity levels for Saami, the rest of Uralic, and all the languages in my database.

Table 2. Complexity of noun inflection in Saami and selected other languages.

	Saami	*Finnish*	*Russian*	*English*
Ending classes	1	1	4 *	2 ***
Unpredictable	–	–	4	2
Stem classes	1–11	1	1 **	2 ***
Unpredictable	All	–	–	2

* The ending sets now increasingly called “masculine”, “neuter”, and “feminine” plus the traditional “third declension” (now fourth; words like *set′* ‘net’, *put′* ‘path’).
** Stress alternations are not counted.

[5] Predictable, that is, provided the transcription is sufficiently abstract to account for stem shapes such as *tiede,* oblique *tietee*- 'science', where the stem-final syllable is open in both at the surface level but not underlyingly: {tie.teC} *tiede* vs. {tie.te.Ce-} *tietee*-, where C is an abstract consonant that has abstract zero as its weak grade but phonological zero, abstract non-zero unspecified consonant as its strong grade.

*** Nouns like *cat*, *dog*, *house* vs. ones like *tooth*, *mouse*. The second set have a zero plural suffix and a vowel change (*teeth* vs. *cats*).

Table 3. Noun stem classes in Saami languages. For what are sometimes called class suffixes or class markers, see §2.3.

Language	*Classes*	*Source*	*Comment*
Ter	1	Tereškin 2002	
Kildin	11	Kert 1971: 143	
Skolt	10	Feist 2015: 158	6 large classes
Akkala	2 or 3	Zajkov 1987	1 gradation, 2 vowel (+ non-gradation?)
North	4	Aikio & Ylikoski 2010: §4.2	(see text)
Lule	6 or 8	Spiik 1989	
Pite	4 or 5	Wilbur 2014: 98ff.	
South	1	Bergsland 1994	

Table 4. Canonical complexity in Saami languages. All = all measures. Strict = removing any (such as the categories involved) that might be considered inventory complexity.

Language	*All*	*Strict*
South	24	15
Pite	62	45
Lule	57	37
North	45	30
Inari	54	37
Skolt	57	41
Kildin	46	33
Mean	49.3	34.0

Table 5. Mean canonical complexity for Saami and other languages. World = all 116 languages surveyed so far. (New coding principles have been implemented recently, but most of the non-Uralic languages still follow the older principles, so the World total shown here is approximate.)

	All	*Strict*
Saami	49.3	34.0
Other Uralic	25.9	17.2
World	32.4	26.6

2.3. Problems of grammatical analysis

Only seven Saami languages could be fully surveyed for Tables 3–5. Akkala and Ter were surveyed, but available grammars do not give all the information needed. For most of those that are included, uncertainties remain. Many grammars can be described as comparative-historical in their primary orientation: the presentation of inflectional paradigms focuses on syllable count and stem-final vowel vs. consonant, the factors that historically triggered consonant gradation and some vowel alternations and are necessary for reconstructing the historical evolution of Saami. These are essential in any Saami description, and are required for comparative-historical work, but they are not the full picture. Synchronically, these stem shape properties are the distributional observations on stem classes. Not all grammars enumerate and exemplify stem classes in the strictly synchronic general-linguistic sense used here, and none make clear whether they are contrastive or in complementary distribution.

A case in point is what are often called "class suffixes" or "classification" or "class markers" in Saami noun inflection. These are a set of vowels that appear in vowel-final stems. Which stem takes which vowel is unpredictable, but each stem has only one vowel (which undergoes automatic phonological variation in some contexts in some languages). They are generally found throughout the inflectional paradigm. They are found in at least some derived forms (coverage of derivation in most grammars is not full, so this is not certain). Many are consistent in Saami cognate sets, and often appear to be inherited from Proto-Uralic.[6] The grammars I have read show that they are identifiable but not that they are morphologically segmentable.[7] I provisionally analyze them as stem-final or root-final elements, not separate morphemes, so they do not contribute to inflectional complexity.

Another is the number of manners of obstruent articulation. Most grammars list only two, voiceless and voiced, implicitly treating long or geminate consonants as two-consonant sequences. However, long consonants figure in consonant gradation as strong counterparts to singleton weak grades (e.g., Lule 1sg *guláv*, 3sg *gullá* 'hear'; Spiik 1989: 27, 31), and while they can arguably be syllabified as C.C in intervocalic position (e.g., *gul.lá*), a syllable

[6] Proto-Uralic lexical roots are usually CV(C)CV in shape (Aikio 2021: §1.3.4). An example of a *CVCV word with such a vowel is PU **kala* 'fish', Saami **kōlē*, e.g., S Saami *guölie*, N Saami *guolle*, Kildin *kūˊll* (Lehtiranta 2001: set 511).

[7] For instance, in Pite Saami (Wilbur 2014: 98–108) there are three noun classes: one with a final vowel, one with final -Vj, and one consonant-final. The final vowels are visible in the nominative singular, distinguishable from each other there and in their morphophonemic interaction with inflectional suffixes, and distinguishable from -Vj or lack of a final vowel. Inflection and derivation use the stem with its class suffix as base, so the class suffix behaves as though it were an unsegmentable root-final element rather than a separate morpheme.

boundary internal to the geminate is impossible when they occur word-finally (e.g., Kildin Nsg *vuedd* ‘ground’, Rießler 2022: 225, Table 13.3) and in some internal clusters (Akkala loc.sg. *puacckas* ‘part of reindeer harness’, surely {*puacc.kas*}) (Zajkov 1987: 69).

These three points – the full list of gradation and vowel-alternation classes, whether class suffixes are demonstrably morphemes rather than stem-final vowels, and the phonological status of long consonants – are usually not all covered fully in Saami grammars and for the present survey had to be extracted or inferred from partial or scattered description and inspecting what examples are given. To construct a single consistent framework I use not holistic classes but a componential analysis based on which gradation patterns, which vowel alternation patterns, and which (if any) other factors do and do not cooccur. The following gives, as a sample, the decisions behind the numbers of stem classes for noun inflection in three of the languages surveyed.

For Kildin Saami, Kert (1971) lists 9 basic declension classes, plus two more for longer words. The classes are based on consonant gradation, consonant palatalization, and vowel alternation. For the nine basic classes there is one pattern of consonant gradation in the singular paradigm: strong in the nominative, essive, and dative, weak elsewhere.[8] (*Weak* and *strong* are standard terms based on alternations and not on phonetics. The strong grade, found inter alia in the nominative singular of nouns, can be taken as base or underlying; it undergoes lenition-like changes in the weak grade contexts. The output is predictable given the underlying consonant. Any given phoneme may be either weak or strong depending on the underlying consonant: for example, single voiceless stops are strong grade that becomes voiced and/or continuant in the weak grade, and they are the weak grade of long or geminate stops.) Vowel-final stems and stems with three or more syllables lack gradation. Since I am treating syllable count and type of final segment as distributional observations, I lump all stems together and analyze Kildin as having two gradation classes: gradation and no gradation. Which words have gradation is an arbitrary lexical property, and in words that have it the distribution of gradation among the cases is absolutely consistent and morphomic rather than predictable.

Orthogonally to consonant gradation, stems either do or do not have vowel alternation, for which there is again one pattern: a change of vowel quality in the dative singular. Which vowel appears in the dative is predictable from the vowel of the other stems, but whether that vowel change occurs in the dative is not predictable.

[8] Here and below, I use the term dative for the case variously called dative, dative-illative, and illative in different grammars.

The canonical complexity count for noun stem classes has three entries: maximum number of stems per lexeme, number of stem classes per language, and number of those stem classes that are unpredictable. The inflectional paradigms surveyed for nouns are the singular of the cases for A, S, O, G, and T arguments, as well as possessor: nominative, accusative, genitive, dative. For Kildin Saami, in those four cases there are up to three possible stem forms, schematically shown in Table 6:

Table 6. Kildin Saami noun stem gradation and ablaut, and two examples (Kert 1971: 141–142). /ĺ, ŕ/ = palatalized consonant.

	Consonant gradation	Vowel ablaut	'fish'	'tree'
NOM	Strong	basic	*kuĺĺ*	*murr*
ACC	Weak	basic	*kuĺ*	*mur*
GEN	Weak	basic	*kuĺ*	*mur*
DAT	Strong	changed	*kuell-a*	*muŕŕ-e*

There are four stem classes in the language: ± gradation and ± ablaut. Lack of gradation occurs only with consonant-final stems and trisyllabic or longer ones, but these are contextual and not predictive factors. Therefore, all four stem classes are unpredictable. Thus, noun stems contribute 11 complexity points to the Kildin total: 3 for stems per lexeme and 4 for both classes per language and unpredictability of those classes.

Lule Saami has different patterns and interactions:

Gradation, strong > weak	(1 pattern)
Gradation, weak > strong; ± ablaut	(2 patterns)
No gradation	(1 pattern)
Total	4 stem classes, all unpredictable

Skolt Saami has more patterns and factors: two gradation patterns (nominative strong, genitive-accusative weak, as in Kildin; nominative weak, others strong), of which the former has ± stem-final vowel loss and the latter has ± palatalization in the nominative singular. Disyllables contrast in whether it is the medial or the final consonant that undergoes gradation and whether there is gradation or not. Feist sets up 7 stem classes (plus others specific to particular derivational suffixes, not counted here). I count the following:

Strong > weak ± stem-final vowel loss, ± C2 gradation, ± Dat.sg. strong+	4 patterns
Weak > strong ± palatalization in Nom.sg.	2 patterns

No gradation	1 pattern
Total	7 classes

Not all are found in words with all syllable counts, but that is what I call a contextual factor. C2 gradation has to do with the fact that in one of the classes of Feist (2015) it is the final rather than the second consonant or sequence that undergoes gradation. The dative (Feist: illative) has a distinct super-strong grade (strong+) in some nouns.

3. Correlations of complexity with non-linguistic factors

I looked for correlations of complexity with three non-linguistic factors that are easily measured or looked up and which figure in many discussions of linguistic complexity as sociolinguistic factors fostering increased complexity.

(1) Latitude, which also has implications for climate (colder and drier to the north) and ecology (forest in the south, tundra in the north), which determine resource richness and directly impact population size, carrying capacity of the land (range size per capita needed for survival), and prosperity in general. Other things being equal, richer resources and higher prosperity attract others to identifying with the group and adopting its language. The firmly reconstructed prehistory of Saami involves its appearance in Scandinavia beginning perhaps 1500 years ago, followed by a Saami spread to cover all the known historical territory, and absorption of the previous inhabitants of that area into the Saami ethnic and speech communities (Aikio 2012, 2022). More recently, the introduction of reindeer herding made possible large herds and production of surplus on tundra land, the northernmost ecology, so the high latitudes became more prosperous.

(2) Population. Various figures on population size are available, of which I mainly used Krauss 1997. I used the figures not for number of speakers of the language but for overall population size (speakers and non-speakers), since the overall size is a better reflection of the group's size before the speedup of language shift and extinctions in the last several decades. I measured the size not as total or rounded total figures but as the magnitude of the population: hundreds, thousands, tens of thousands. Krauss's figures generally reflect the situation in the mid to late 20th century.

(3) Range. I estimated range sizes by inspecting maps (Rantanen et al. 2022), taking a range the size of Ume or Pite as the basic unit, and estimating by eye the sizes of the other ranges as fractions or multiples of those. This measure is obviously Saami-specific and requires good language and dialect

maps, and it is very approximate, but it is quite adequate as a quick-and-dirty measure.

Table 7. Complexity means and correlations with non-linguistic factors. Canonical complexity for the world sample is based on an earlier coding system and are not exact. Range size estimates are available only for Saami. Population figures for the world sample are not available (see text). Bold = statistically significant (p ≤ .05); italics = fairly close to significant.

	Enumerative:		Canonical:		
	All	Morph. All	Strict		
Mean complexity:					
Saami mean	19.3	8.0	49.3	34.0	
Other Uralic mean	17.7	8.7	25.9	17.2	
World mean	19.1	9.9	32.4	26.6	
Correlation coefficients:					Expected:
Saami (N=6 (enumerative) or 7 (canonical); 0.73 or 0.67 required for $p = 0.05$).					
Latitude	0.216	0.520	**0.702**	**0.664**	+
Population	*−0.622*	−0.267	−0.111	−0.155	-
Range	−0.485	−0.078	*−0.542*	*−0.552*	-
Other Uralic (N=9 (enumerative) or 12 (canonical); 0.58 or 0.50 required)					
Latitude	0.346	**0.627**	0.394	0.353	+
Population	−0.186	**−0.580**	−0.346	−0.270	-
World (N = 245 (enumerative) or 117 (canonical); 0.10 or 0.12 required)					
Latitude	**0.153**	−0.068	**0.197**	**0.171**	+

Correlations of complexity levels with these factors, for Saami and other languages, are shown in Table 7 above. Population figures for the entire database have not been compiled yet.[9] A quick and easy calculation of relative

[9] Nichols (2009) found a highly significant negative correlation worldwide between population size and an earlier and less complete version of enumerative complexity, using a smaller language sample, but showed that it was an artifact of population size differences between very

range sizes like that used here for Saami is problematic for the rest of Uralic and not feasible for the whole world, so correlations with range are not calculated for the other languages. There are two relevant measures of adequacy: statistical significance, hard to reach with a small sample but nonetheless reached for canonical complexity; and whether correlation coefficients are positive or negative as expected. Expected correlations are as shown in the table.

Latitude: A positive correlation is expected (higher latitude = higher complexity), since higher latitude generally means greater distance from centers of spread and economic and cultural centers. (Bentz 2016 finds a worldwide correlation between low latitude and low complexity.) Positive correlations with latitude were found within Saami, significant for canonical complexity, appreciable for morphological enumerative complexity, but very weak for overall enumerative complexity. The largest difference between overall and morphological enumerative complexity is that measures of phonological complexity were included in the overall figure, and this suggests that phonological complexity does not correlate positively with latitude. I did a separate count of phonological complexity, using the more elaborated measure of Hartmann & Nichols in press, and here the correlation was negative (−0.381) but weak, confirming that there is no correlation between latitude and phonological complexity. The correlation for only morphological enumerative complexity is more similar to those for canonical complexity (unsurprisingly, as the measures of canonical complexity include only morphology).

Population: A negative correlation is expected (smaller population = higher complexity). The linguistic literature finds mixed results on whether such a correlation exists worldwide (pro: e.g., Hay & Bauer 2007 for phonological complexity; con: Nichols 2009 for overall enumerative complexity [see again footnote 8]; Donohue & Nichols 2011 for phoneme inventory size), but it continues to be a good hypothesis, since small populations are often sociolinguistically isolated and have not recently absorbed appreciable numbers of adult L2 learners whose presence can decomplexify a language (for this effect see below). Mostly negative correlations were found as expected (close to significant for overall enumerative complexity; others negligible).

---

large areas: mean population sizes are much larger in Eurasia and Africa than in the Pacific or the Americas, and mean enumerative complexity levels are higher in the Pacific and (especially) the Americas. That imbalance accounted for the negative correlation; there was no correlation within any of the three large areas. All of those findings would undoubtedly be true with the current coding and sample as well.

Range size: Negative correlations are expected, for the same reasons as for correlation with population magnitude. Negative correlations were found, most of them appreciable.

These factors are of interest not in themselves but because of the sociolinguistic and economic factors that are likely to have caused them. Those include such things as known language expansions and whether they primarily involved language shift (and consequent absorption of adult L2 learners); known relative economic strength of different groups; number and nature of contact situations; and others. No attempt has been made to measure such factors here, since collecting the data would be a major research project in itself and would require expertise in Saami ethnography and history. A known relevant episode is the relatively recent expansion of North Saami with the introduction of reindeer herding involving language shift (of other Saami populations to North Saami) and territorial expansion to utilize the tundra. This would have meant absorption of adult L2 learners, which can decomplexify a language as noted above, and in fact North Saami is low in both enumerative and canonical complexity compared to the other Saami varieties. South Saami is even less complex in all measures, possibly because of its historical sociolinguistic status but most likely because it diverged first from the rest of Saami (Aikio 2012), before consonant gradation arose.

Principal findings for the three non-linguistic factors are these. With one exception, correlations were positive or negative as expected. The correlations of complexity with latitude and population magnitude found for Uralic are somewhat different for the rest of Uralic, but appreciable, and significant for morphological enumerative complexity. For the whole world, correlations with latitude are highly significant except for morphological enumerative complexity (which is furthermore the only case of sign reversal in these correlations).

4. Discussion

The positive correlation of complexity with latitude found for Saami finds support in the rest of Uralic and strong support worldwide. The negative correlation with population magnitude, for enumerative complexity only, is partly supported in the rest of Uralic. All in all, the other findings give plausibility, and sometimes support, to what was found for Saami and make it worthwhile to continue interpreting them. It is reasonably safe to generalize that grammatical complexity in Saami increases at higher latitudes, with smaller populations, and with smaller range sizes. All three are likely correlates of a sociolinguistic situation similar to what Trudgill (2011) terms *sociolinguistic isolation*: no recent history of absorbing speakers of other languages. Absorbing adult L2 speakers generally decomplexifies a language, but otherwise complexity gradually builds up in intergenerational transmission. Languages which do not absorb adult learners are under no pressure to decomplexify, and can theoretically complexify without limit (though a situation with zero shift to the language and nonzero shift from it means that extinction is inevitable, putting a practical limit to the complexity). Note that sociolinguistic isolation does not mean isolation in the sense of minimal contacts; sociolinguistically isolated languages are often in multiple close contacts but not as dominant or expanding languages. (This is true of the eastern Caucasus, for instance.) They seem to often be in locations that are peripheral to centers of spread, commerce, and population growth. This general picture describes Saami languages, spoken by fairly mobile populations that were in contact with other languages, Saami and non-Saami, but in the ultimate peripheral location at the northern edge of the habitable world.

Ordinarily, findings can be firmed up by expanding the sample, but in the case of Saami the remaining languages – Ume, Akkala, and Ter – have only partial descriptions and could not be surveyed fully enough to include them here. It is easy to see that Akkala and Ter, if included, would upset some of the patterns established above. Though they are far northern languages they have very small range sizes for their latitude, and their gradation systems and phonological inventories appear to be relatively simple compared to those of adjacent Skolt and Kildin (and their phoneme inventories no more complex). Their populations, on the other hand, are very small as predicted for their latitude. It is possible that some of the non-complexity is due to incomplete enumeration of gradation classes in the existing descriptions. Simplification can also be a language death symptom, and that seems more likely in the case of Akkala and Ter: both seem likely to have been losing speakers to their neighbors, and both had their economic institutions and community centers destroyed during collectivization and much of their population scattered (in the case of the hunter-fisher Akkala, scattered to towns or herding collectives, putting them in an unfamiliar social and economic situation); the extent of these losses appears to have been greater than for the Kildin and Skolt. Perhaps comparative documentation of factors like these could be assembled for more Saami groups to improve the qualitative nature of the correlations.

Extensive ethnographic, demographic, historical, and archaeological work has been done on such matters as Saami migration routes, whether they were for transhumant herding, hunting and trapping, or fishing, and their seasonal schedules, geography, and distance covered. Such things could be brought together to give a much richer and more precise description of range size.

5. Conclusions

The unified descriptive framework for assessing Saami canonical complexity includes these points:

• Inflectional paradigms consist of stem, ending, and various distributional observations that delimit distributions by stem shape, syllable count, etc. but not do not predict either stem alternations or form of ending. For the most part, the stem alternations of consonant gradation and vowel ablaut are lexically specified and not predictable.

• Stem classes are defined by consonant gradation and vowel ablaut. The gradation and ablaut classes are defined abstractly: consonant grades are strong vs. weak; for lack of any standard terminology for vowel grades I have called them base (found in the nominative singular) vs. changed. A pattern is a set of consonant or ablaut grades stating which grade appears in which paradigmatic position (which case form, etc.).

• Distributional observations include the syllable count, whether the stem is consonant-final or vowel-final, and whether there is a stem extension in some of the forms.

• Ending classes are defined by same vs. different sets of endings. In Saami languages there is only one ending class for nouns and usually only one for verbs. Allomorphy in endings is usually phonologically conditioned, i.e., predictable.

• Canonical complexity is defined as the number of stem classes, the number of ending classes, and the number of different patterns of syncretism, allomorphy, zero endings (other than nominative), etc.

The main findings on the distribution and correlations of complexity in Saami are these:

• Enumerative complexity in Saami is slightly higher than in other Uralic, but unremarkable overall.

• Canonical complexity is high overall in Saami. It is due to morphologization of what were originally allophonic or subphonemic alternations. It is primarily a matter of stem alternations; endings are simple and regular.

• It is easier to identify the lowest-complexity Saami languages than the highest-complexity ones: South and North Saami have markedly lower complexity. For North Saami this probably reflects expansion to absorb other Saami speakers and, for recent years, standardization and use in media. For

South Saami sociolinguistic factors must be relevant, such as serving as contact frontier with North Germanic from early times on; but the early separation of South Saami from the rest must have meant that it simply did not undergo the post-Proto-Saami sound changes that eventually resulted in gradation and ablaut alternations.

• Complexity correlates positively with more northern latitude and smaller range and/or population. These three factors are interconnected and point to a sociolinguistic situation involving minimal pressure to decomplexify and transmission mostly by childhood L1 acquisition and infrequently by adult L2 acquisition, a situation associated with small and peripheral languages.

A final observation is that the analysis of linguistic complexity requires good grammatical descriptions, good understanding of the terminology and conventions of the descriptive tradition for those grammars, and knowledge of the non-linguistic factors that may explain the distribution of complexity levels across languages. Good descriptions in traditional and structural terms are available for many Saami languages, but there is still a need for comprehensive, typologically oriented reference grammars of as many of the languages as possible. I can vouch for the soundness and tested usefulness of the definitions and coding system used here, but good hypotheses and good explanations will require expertise in Saami linguistics, sociolinguistics, history, ethnography, and probably other fields, and for best results they should be initiated by Saami specialists themselves.

References

Aikio, Ante = Luobbal Sámmol Sámmol Ánte. 2012. An essay on Saami ethnolinguistic prehistory. In Riho Grünthal & Petri Kallio (eds), *A Linguistic Map of Prehistoric Northern Europe* (Suomalais-Ugrilaisen Seuran Toimituksia 266), 63–117. Helsinki: Société Finno-Ougrienne.

Aikio, Ante = Luobbal Sámmol Sámmol Ánte. 2022. Proto-Uralic. In Marianne Bákro-Nagy, Johanna Laakso & Elena Skribnik (eds), *The Oxford guide to the Uralic languages*, 3–27. Oxford: Oxford University Press.

Aikio, Ante & Jussi Ylikoski. 2010. *The structure of North Saami*. Course handout.

Anderson, Stephen R. 2015. Dimensions of morphological complexity. In Matthew Baerman, Dunstan Brown & Greville G. Corbett (eds), *Understanding and measuring morphological complexity*, 11–26. Oxford: Oxford University Press.

Audring, Jenny. 2017. Calibrating complexity: How complex is a gender system? *Language Sciences* 60. 53–68.

Bentz, Christian. 2018. *Adaptive languages: An information-theoretic account of linguistic diversity* (Trends in Linguistics. Studies and Monographs. Volume 316). Berlin: de Gruyter.

Bentz, Christian. 2016. The low-complexity belt: Evidence for large-scale language contact in human prehistory? *The Evolution of Language: Proceedings of the 11th International Conference* (EVOLANG 11). https://evolang.org/neworleans/pdf/EVOLANG_11_paper_93.pdf

Bergsland, Knut. 1994. *Sydsamisk grammatikk*. Karasjok: Davvi Girji.

Bickel, Balthasar & Johanna Nichols. 2013. Inflectional synthesis of the verb. In Matthew S. Dryer& Martin Haspelmath (eds) *WALS Online* (v2020.4) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.13950591

Corbett, Greville. 2007. Canonical typology, suppletion, and possible words. *Language* 83(1). 8–42.

Corbett, Greville. 2015. Morphosyntactic complexity: A typology of lexical splits. *Language* 91(1). 145–193.

Dahl, Östen. 2004. *The growth and maintenance of linguistic complexity* (Studies in Linguistic Companion Series 71). Amsterdam: Benjamins.

Donohue, Mark & Johanna Nichols. 2011. Does phoneme inventory size correlate with population size? *Linguistic Typology* 15(2). 161–170.

Ehret, Katharina & Benedikt Szmrecsanyi. 2016. An information-theoretic approach to assess linguistic complexity. In Raffaela Baechler & Guido Seiler (eds), *Complexity, Isolation, and Variation* (linguae & litterae, volume 57), 71–94. Berlin: de Gruyter Mouton.

Feist, Timothy. 2015. *A grammar of Skolt Saami* (Suomalais-Ugrilaisen Seuran Toimituksia 273). Helsinki: Suomalais-Ugrilainen Seura.

Hartmann, Frederik & Johanna Nichols. (In press.) Geospatial effects on phonological complexity in the world's languages. *Linguistic Typology*.

Hay, Jennifer & Laurie Bauer. 2007. Phoneme inventory size and population size. *Language* 83(2). 388–400.

Janda, Laura A. & Lene Antonsen. 2016. The ongoing eclipse of possessive suffixes in North Saami: A case study in reduction of morphological complexity. *Diachronica* 33(3). 330–366.

Juola, Patrick. 1998. Measuring linguistic complexity: The morphological tier. *Journal of Quantitative Linguistics* 5. 206–213.

Kert, Georgij M. 1971. *Saamskij jazyk (kil'dinskij dialekt). Fonetika, morfologija, sintaksis*. Leningrad: Nauka.

Krauss, Michael E. 1997. The indigenous languages of the North: A report on their present state. In Hiroshi Shoji & J. Janhunen (eds), *Northern minority languages: Problems of survival*, 1–34. Osaka and Fairbanks: National Museum of Ethnology and Alaska Native Language Center, University of Alaska Fairbanks.

Kusters, Wouter. 2008. Complexity in linguistic theory, language learning, and language change. In Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds), *Language complexity: Typology, contact, change* (Studies in Language Companion Series, 94), 3–22. Amsterdam: Benjamins.

Lehtiranta, Juhani. ^{2}2001. *Yhteissaamelainen sanasto* (Suomalais-Ugrilaisen Seuran Toimituksia 200). Helsinki: Suomalais-Ugrilainen Seura.

Miestamo, Matti. 2008. Grammatical complexity in cross-linguistic perspective. In Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds), *Language complexity: Typology, contact, change* (Studies in Language Companion Series, 94), 23–41. Amsterdam: Benjamins.

Morottaja, Matti & Marja-Liisa Olthuis. 2020. *Inarinsaamen koulukielioppi*. Aanaar/Inari: Säämi kulttuurkuávdáš Sajos.
Nichols, Johanna. 2020. Canonical complexity. In Peter Arkadiev & Francesco Gardani (eds), *The complexities of morphology*, 163–192. Oxford: Oxford University Press.
Nichols, Johanna. 2019. Why is gender so complex? New typological considerations. In Francesca Di Garbo, Brunol Olsson & Bernhard Wälchli (eds), *Grammatical gender and linguistic complexity. Volume 1: General issues and specific studies*, 63–92. Berlin: Language Sciences Press.
Nichols, Johanna. 2009. Linguistic complexity: A comprehensive definition and survey. In Geoffrey Sampson, David Gil & Peter Trudgill (eds), *Language Complexity as an Evolving Variable*, 110–125. Oxford: Oxford University Press.
Nichols, Johanna & Christian Bentz. 2019. Morphological complexity of languages reflects the settlement history of the Americas. In Katerina Harvati, Gerhard Jäger & Hugo Reyes-Centano (eds), *New Perspectives on the Peopling of the Americas*, 13–26. Tübingen: Kerns.
Nickel, Klaus Peter & Pekka Sammallahti. 2011. *Nordsamisk grammatikk*. Davvi Girji.
Rantanen, Timo, Harri Tolvanen, Meeli Roose, Jussi Ylikoski & Outi Vesakoski. 2022. Best practices for spatial language data harmonization, sharing and map creation: A case study of Uralic. *PLoS ONE* 17(6). e0269648. https://doi.org/10.1371/journal.pone.0269648
Rießler, Michael. 2022. Kildin Saami. In Marianne Bákro-Nagy, Johanna Laakso & Elena Skribnik (eds), *The Oxford guide to the Uralic languages*, 219–239. Oxford: Oxford University Press.
Ross, Malcolm D. 1996. Contact-induced change and the comparative method: Cases from Papua New Guinea. In Mark Durie & Malcolm D. Ross (eds), *The comparative method reviewed*, 180–217. New York: Oxford University Press.
Sinnemäki, Kaius. 2011. *Language universals and linguistic complexity: three case studies in core argument marking*. PhD dissertation, Helsinki: Department of Modern Languages.
Sokur, Elena & Johanna Nichols. 2019. *Noun vs. verb inflectional synthesis: A complexity tradeoff?* Presented at SLE 2019 meeting, Leipzig.
Spiik, Nils Eric. 1989. *Lulesamisk grammatik*. Luleå: Sameskolsyresen.
Tereškin, Sergej N. 2002. *Jokan'gskij dialekt saamskogo jazyka*. PhD dissertation. Sankt-Peterburg: Rossijskij gos. pedagogičeskij universitet im. A. I. Gercena.
Thurston, William R. 1989. How exoteric languages build a lexicon: Esoterogeny in West New Britain. In Ray Harlow & Robin Hooper (eds), *VICAL 1 – Oceanic languages: Papers from the Fifth International Conference on Austronesian Linguistics*, 555–579. Auckland: Linguistic Society of New Zealand.
Trudgill, Peter. 2011. *Sociolinguistic typology: Social determinants of linguistic structure and complexity*. Oxford: Oxford University Press.
Wilbur, Joshua. 2014. *A grammar of Pite Saami* (Studies in Diversity Linguistics, No 5). Berlin: Language Science Press.
Zajkov, Pekka M. 1987. *Babinskij dialekt saamskogo jazyka (fonologo-morfologičeskoe issledovanie)*. Petrozavodsk: Karelija.

Appendix. All entries for complexity measures and non-linguistic factors. Extracts appear in tables in the text. Canonical complexity values for World (italicized) are based on an older coding system and are not exact. Complexity entries are totals; full breakdowns will be in the database to be published c. 2025.

Languages	*Complexity*				*Non-linguistic factors*			
	Enumerative		Canonical		Lat.	Long.	Population magnitude	Range
	All	Morph.	All	Strict				
South Saami	18	7	24	15	62.4	11.2	4	3
Ume							4	1
Pite	20	8	62	45	66.0	17.5	4	1
Lule	18	7	57	37	68.0	16.0	4.5	1.5
North Saami	18	8	45	30	69.0	24.0	5	5
Inari			54	37	68.5	27.0	3	0.5
Skolt	20.5	9	57	41	69.4	29.0	4	2
Akkala					67.9	32.4	3	0.5
Kildin	21.5	9	46	33	67.0	37.0	4	2
Ter					67.3	38.5	3	0.8
Finnish	16	7	25	15	62.0	25.0	7	
Estonian	20.5	9			59.0	26.0	7	
Erzya	17	8	24	15	53.0	45.5	5	
Mari	15	6	22	16	56.6	48.0	6	
Udmurt	14.5	6			57.5	52.5	6	
Komi	16	7			62.0	50.0	6	

Hungarian	19	8	30	22		47.0	20.0	7
Mansi	18	11	22	16		62.0	62.0	5
E Khanty	17	9	26	14		64.0	78.0	4
N Khanty			20	11		66.3	66.4	5
Selkup	17	9				65.0	82.0	4
T Nenets	20	11	31	21		69.0	72.0	5
Nganasan	22	13	33	25		71.0	93.0	4

Means:

Saami	19.3	8.0	49.3	34.0
Non-Saami Uralic	17.7	8.7	25.9	17.2
All Uralic	18.2	8.4	36.1	24.6
World	19.1	9.9	*32.4*	*26.6*

Notes:

Pite: The very high entries for canonical complexity are contributed mostly by the verb, which is much more complex than the noun (in the rest of Saami they are similar in complexity or nouns are somewhat more complex).

Kildin: Population figure based on Kert 1971 and Rießler 2022.

Other figures for population magnitude based chiefly on population figures in Krauss 1997.

Magn.: Order of magnitude of the population (e.g., 4 = numbering in thousands, 5 = ten thousands, etc.).