ESUKA – JEFUL 2025, 16–2: 51–88

VOWEL-STEMMED NUT PARTICIPLE VARIATION IN THE FINNISH USED IN THE WORKS OF MIKAEL AGRICOLA

Heidi Salmi[a], Jussi Salmi[b]
[a] *University of Turku, FI*
[b] *Turku University of Applied Sciences, FI*
heidi.salmi@utu.fi, jussi.salmi@turkuamk.fi

Abstract. Finnish had begun to be written in the 16th century, and the language of Mikael Agricola represents the oldest written Finnish. Agricola's language differs in many ways from contemporary Finnish. This article examines the NUT participle variation found in the Finnish used by Agricola, which no longer exists in the contemporary language.

This study answers whether there are statistically significant variations between NUT participle types caused by linguistic variables. The research shows that phonological variables affect NUT participle variation more than morphological ones. NUT participle variants form an interesting network: variables pair up or group variants so that variables favour a certain variant(s) and alternatively reject other(s).

The overall importance of the variables in selecting a participle was also examined by applying a machine learning classifier algorithm. The most important features in this selection were the stem vowel, if the word has back or front vowels, person and tense.

Keywords: NUT participle, old literary Finnish, Mikael Agricola's Finnish, corpus linguistics, machine learning

DOI: https://doi.org/10.12697/jeful.2025.16.2.02

1. Introduction

Even though Finnish culture has old roots, written Finnish is relatively new. Finnish had not begun to be written until the 16th century. The Reformation, which began in Germany in 1517, was the significant factor for the literary development of Finnish. According to Lutheranism, people had to be able to read the word of God in their own native language, and thus, literary Finnish had to be created.

The first to produce books in Finnish was Mikael Agricola, a Finnish Reformer, headmaster of the cathedral school in Turku and Bishop of

Turku, and thus the language Agricola used in his 16th century works (hereinafter Agricola's Finnish) represents the oldest written Finnish. Agricola's Finnish has been studied from many different perspectives (for recent studies see, e.g., Salmi 2010 on adpositions; Nummila 2011 on agent nouns; Elsayed 2017 on the necessitative verb *pitää* 'to keep, to hold'), and its special features are fairly well known by now.

This article examines Agricola's NUT participles, which are verbs or, in other uses, verb-based words. First of all, Finnish verbs can be divided into finite and nonfinite verbs. NUT participles are nonfinite verbs which are used in Finnish compound tenses: the perfect and the pluperfect tense and negative imperfect (see examples 3–5). Although participles are verbs or based on lexical verbs, they can be nominally inflected and the NUT participles can also be used as adjectives and nouns (see examples 6–8).

The NUT participle marker (i.e. ending) in standard contemporary Finnish is *nut : nyt : nee* (examples 1 and 2):

1. *luke-nut* : *luke-nee-t*
   read-NUT PARTIC read-NUT PARTIC-PL

2. *löytä-nyt* : *löytä-nee-t*
   find-NUT PARTIC find-NUT PARTIC-PL

On the basis of previous research (e.g. Häkkinen 1994: 317), it is clear that there is a great deal of variation in the NUT participle used in Agricola's Finnish, which does not exist in contemporary Finnish. The NUT participle in old literary Finnish is formed with a consonant stem more often than what is found in the contemporary language (e.g. *löyn-nee-t* instead of *löytä-nee-t* 'to find'). In addition to this, there is a great deal of variation in the vowel-stemmed NUT participle (presented in Section 2). This article examines the variation of the vowel-stemmed NUT participle in Agricola's Finnish, and the data includes all the vowel-stemmed NUT participles in plural and singular form and all the cases which appear in the data.

In general, when variations in language are studied, researchers want to describe and explain what kinds of variations exist and in what circumstances these variations occur. That is, why does X vary alongside Y in context Z. The variation may be caused by intra- or extra-linguistic

factors (Bayley 2002: 117, 123; Anttila 2002: 206; Labov 2004: 6–7, 14). Intra-linguistic research examines how phonological, morphological, syntactic and lexical variables affect the use of a certain variant (see Feagin 2002: 23; see also Nuolijärvi & Lappalainen 2020: 872). When it comes to research on 16th century language, intra-linguistic factors are the main focus, and thus this study is on intra-linguistic variables. Extra-linguistic factors, such as gender or age of the writer, are not relevant to this study since most of this language variety's material is from one particular writer, making any comparative study impossible. It should also be noted that intra- and extra-linguistic correlations do not exclude one another. That is, a variant may be, for example, both phonologically conditioned and correlated to, for instance, a speaker's gender or age. It is also important to note that it is a matter of correlations rather than causations.

Previous research on NUT participle variation has shown that phonological issues are intertwined with morphological and syntactic ones (Mielikäinen 1982: 279, 290; Mielikäinen 1995: 333). Aila Mielikäinen (1995: 331) writes that when explaining variation, one must also pay attention to its phonetic environment, stress (which is a phenomenon of spoken language), syntax and vocabulary. This article only focuses on the phonological and morphological features of variation; syntactic and lexical variation are not included. It would be interesting to study the semantics of verbs in NUT participle forms in a later study, as Pirkko Forsman Svensson (1992: 20–21) observed that there are only around 20 different verbs in data concerning 17th century sermons, and the five most common make up nearly 80% of occurrences with NUT participles, and these are qualifiers of nouns (adjectives). On the other hand, when it comes to non-religious literature, the NUT participle as an adjective is rare.

This article responds to the research question of whether some phonological and morphological attributes have affected the variation of the vowel-stemmed NUT participle. (The variables are introduced in Section 3.2.) Article examines the NUT participle variation of Agricola's old Finnish written texts in a new way from the perspective of statistical methodology (introduced in Section 3.3). Statistical methodology is needed, because there are no grammatical rules for the use of the variants, and without studying the subject statistically, it might seem that the use of the variants is random and no grammatical factor

determines which variant of Agricola's NUT participle is used at any given time (on free variation see e.g. Paunonen 1984: 139–154; Kager 1999: 404; Kopf & Weber (eds.) 2023).

The research material has been collected from the morphosyntactic annotated database of Agricola's Finnish (*The Morpho-Syntactic Database of Mikael Agricola's Works version 1.1, Korp*, https://www.kielipankki.fi/corpora/agricola/).

2. The NUT participle and its variation following previous research

Agricola's NUT participle differs from the contemporary NUT participle by having several different marker variants. The following different variants of the marker can be seen in Agricola's Finnish: *ee* (*luke-nee-t*), *ehe* (*luke-nehe-t*), *ue*/*ye* (*luke-nue-t*), *uve*/*yve* (*luke-nuve-t*) *u*(*u*) / *y*(*y*), (*luke-nu*(*u*)*t*), *ua*/*yä* (*luke-nua-t*) and *uva*/*yvä* (*luke-nuva-t*) (nominative plural forms of NUT participles of *lukea* 'to read') (Lehikoinen & Kiuru 2006: 121). All the example words from here on are in the nominative plural form if the case and the number are irrelevant in a particular example. All the different singular and plural cases found in the data are presented in Section 4.6.

The only possible NUT form found in contemporary Finnish is *luke-nee-t*. It was almost exclusively in all 18th century literary Finnish in general (Rapola 1933: 165–166; Tunkelo 1939: 169).

2.1. The NUT participle in written modern Finnish

Finnish NUT participles are nonfinite verbs. Their main purpose is to function as a part of compound tenses: the perfect (example 3) and the pluperfect tense (example 4) – as well as the negative imperfect (example 5), which is also a compound tense. Examples 3 through 8 are in standard contemporary Finnish, devised to demonstrate the phenomenon as simply as possible.

3. *He* *o-vat* *kuol-lee-t.*
   they be-3PL die-NUT PARTIC-PL
   'They are dead.'

4. *He* *ol-i-vat* *kuol-lee-t.*
they be-PST-3PL die-NUT PARTIC-PL
'They were dead.'

5. *He* *ei-vät* *kuol-lee-t.*
they NEG-3PL die-NUT PARTIC-PL
'They did not die.'

In addition to this, the NUT participle can sometimes be a part of a verb chain (example 6) and can also be used in a way similar to adjectives and nouns (examples 7 and 8). According to Helinä Koivisto (1987) NUT participles become adjectives. Such *-nUt* forms can also be considered deverbal derivatives (Hakulinen 1979: 215; Nummila 2011: 93). Primarily *-nUt* derivatives are adjectives, and as nouns they mainly refer to action, which the agent has already performed in the past. For example *juopunut* 'drunk' can achieve this state only after drinking (Nummila 2011: 93; 152–154). In the dialect corpus, *-nUt* nouns mainly refer to people (Ikola et al. 1989: 504, 506).

6. *Tul-i-n* *sano-nee-ksi* *paljon.*
come-PST-1SG say-NUT PARTIC-TRANSL a lot.
'I came to say a lot.'

7. *Kuol-lee-t* *ihmise-t*
die-NUT PARTIC-PL people-NOM-PL
'Dead people'

8. *Näe-n* *kuol-le-i-ta.*
see-1SG die-NUT PARTIC-PL-PART
'I see (some) dead (people).'

The NUT participle has two different markers in contemporary written Finnish: the consonant-ending marker *-nut/-nyt*, both in the nominative and partitive singular, and a vowel-ending marker *-nee* in all other singular cases and all plural cases (examples 1 and 2).

Even though NUT participles are verb-based, they can be nominally inflected. They can be used in both singular and plural forms and can be inflected in all Finnish cases (see Section 4.6). When the NUT participle is a part of a compound verb, it is always in the nominative case

(examples 3–5), and the vowel-stemmed participle suffix in the nominative case will always be plural, as in the NUT participle nominative plural form *luke-**nee**-t* 'to read'.[1] When NUT participle forms are used as nominals, they can be inflected in any case (examples 7 and 8). NUT participles which are a part of a verb chain are only inflected in certain cases (example 6) (for more on the NUT participle see, for example, Numminen 1954; Koivisto 2013; VISK § 521–537).

If a verb has both a vowel and a consonant stem, the marker will affix to the latter (example 7 and 8, stem *kuol*-).[2] If a verb only has a vowel stem, the NUT participle marker will affix to one of two types: either a stem that does not undergo consonant gradation (example 6, stem *sano*-) or a so-called strong stem in verbs that are subject to consonant gradation (VISK § 122).

2.2. The history of NUT participle marker variants

NUT participle marker consists of two elements, **nU*- and *-*t*. The etymological origin of the marker is still unknown. *-*t* in the marker has been compared to the similar element in *lyhy-t* 'short' and *ohu-t* 'thin' but without adequate argument, and the origin of **nU*- is entirely unknown (a remark from a reviewer).

Both the ***ue*/*ye* and *ee* variants** of the NUT participle are caused by a sound change in Finnish, namely *nee* (*luke-nee-t*) < **nUe* (*luke-nue-t*) < **nUδe* (*luke-nuδe-t*) < **nUte* (*luke-nute-t*) (Häkkinen 2002: 118–119). After the **δ* disappeared, the gapped vowel cluster *u-e*/*y-e* remained. The *ee* variant developed from *ue*/*ye* as a result of an assimilation process (Itkonen 1969: 124, 134).

According to the traditional explanation of the ***ehe* NUT participle variant** (*luke-nehe-t*), *h* appeared to fill the gap between *u-e*/*y-e* and led to forms such as *uhe*/*yhe*.[3] After this, the *e* would have assimilated with the preceding *u*, and the development would have continued as

1 The nominative singular forms always take the consonant-stemmed suffix (e.g. NUT participle nom. sg. *luke-**nut*** 'to read').

2 In Agricola's Finnish – which is also the data for this study –, the essive singular ending is usually (95% of nouns) affixed to the consonant stem, and genitive plural forms are also more often consonant-stemmed than in contemporary Finnish (Lehikoinen & Kiuru 2006: 92).

3 These *uhe*/*yhe* forms are found in the 16th century Finnish manuscript *Codex Westh*.

for the original *ehe* group. This sound change tendency is limited to certain forms (Itkonen 1969: 134). The *h* in the *ehe* variant can either be this kind of glide, caused by an empty space between syllables, or an analogy of other suffixes which begin with *h*.[4] This is an old phenomenon, as a similar *h* is widely found in Finnish dialects (Häkkinen 1994: 317).

Penttilä considers the ***uve* variant** (*luke-nuve-t*) to be a contamination of the *ue* and *uva* types, but Itkonen believes that *uve*/*yve* may have been a transition sound type developed from the *ue*/*ye* variant, in which the *v* strengthens the syllabic boundary (Penttilä 1965: 138–139; Itkonen 1969: 124, 132).

The development of *ue/ye* into ***u*(*u*)/*y*(*y*)** (*luke-nu*(*u*)*-t*) is typical for southwestern Finnish dialects (Rapola 1933; Itkonen 1969: 127), on which Agricola's Finnish was strongly based. The *ee* variant is typical for Häme dialects. It is likely that, at least during the time of the oldest written Finnish, the normal NUT participle in the southwestern dialect was of the *u*(*u*)/*y*(*y*) or *u*(*v*)*a*/*y*(*v*)*ä* type. Under what terms and phases the assimilation development *ue/ye* > *u*(*u*)/*y*(*y*) has taken place is unclear (Itkonen 1969: 131, 149, 152).

Rapola considers ***ua*/*yä* and *uva*/*yvä*** (*luke-nua-t* and *luke-nuva-t*) NUT participles to be parallel forms of the *ue*/*ye* type. According to him, the *ue*/*ye*, *ua*/*yä* and *uva*/*yvä* types, for example *tulluet* and *tulluvat* (NUT participle of the verb *tulla* 'to come'), may have developed from the form **tulluδet*. As a comparison, Rapola presents the southwestern dialect type *lyhyvä* = *lyhyet* (standard contemporary Finnish *lyhyt* 'short') and *ohuva* = *ohuet* (standard contemporary Finnish *ohut* 'thin'), which developed as **lühüδet* > *lühüvät* > *lühüvä* (Rapola 1933: 165; Rapola 1963: 38). Aarni Penttilä, on the other hand, does not believe that this kind of development was likely. According to Penttilä, it is possible that *ohut* and *lyhyt* may have been connected to words that have a *vA* suffix. Penttilä also highlights E. N. Setälä's remark that the northern Estonian dialectal participle *elanuvad* and Finnish forms have a clear connection to one another (Setälä 1887: 13; Penttilä 1965: 141–142). Penttilä (1965: 143) believes that the *tulluvat* NUT participle

4 The intervocalic *-h-* in *ehe* has a different origin than *-h-* in forms such as *lampahat* 'sheep' and the like, where it apparently developed from the earlier intervocalic **-s-* (a remark from a reviewer).

type is not a direct result of any phonetic development, even though it was indirectly caused by southwestern phonetic change: *u-e* > *uu*, *y-e* > *yy*, *uu* > *u* and *yy* > *y*. Itkonen (1969: 147, 150), however, considers Penttilä's explanation to be incorrect and finds the *a*/*ä* in this type to be a result of the sound change *u-e/y-e* > *u*(*v*)*a*/*y*(*v*)*ä*. This change must have started when the pronunciation of *u-e*/*y-e* started to approach becoming a diphthong (*uo*/*yö*), at the latest. The sound *v* developed in this new *ua*/*yä* variant to ensure syllabic boundary.[5]

2.3. Mikael Agricola's NUT participle and its variation

In Agricola's Finnish there are thus NUT participle variants *ee* (*luke-nee-t*), *ehe* (*luke-nehe-t*), *ue/ye* (*luke-nue-t*), *uve/yve* (*luke-nuve-t*) *u(u) / y(y)*, (*luke-nu*(*u*)*-t*), *ua/yä* (*luke-nua-t*) and *uva/yvä* (*luke-nuva-t*) (*lukea* 'to read').

The *ue*/*ye* NUT participle (*luke-nue-t*), is the most common and the *ee* (*luke-nee-t*) the second most common variant in Agricola's Finnish according to previous studies. The *ue*/*ye* and *ee* variants make up most of the NUT participles. Other ones all together make up about one-fifth of all the variants. Of these other variants, *u*(*u*)/*y*(*y*) (*luke-nu*(*u*)*-t*) is the most common (Rapola 1933: 165–166; Koskinen 1965: 17, 174).

Koskinen's study has a table of NUT participles showing the frequency of each variant in different cases. Koskinen observes that there are no oblique cases in NUT participles with the *uve/yve* (*luke-nuve-t*) and *uva/yvä* (*luke-nuva-t*) variants. Furthermore, *uva*/*yvä* is statistically common, but it has such a narrow declination, that Koskinen wonders whether it has some kind of special function in sentences (Koskinen 1965: 173).

According to Koskinen, all *ua*/*yä* and *uva*/*yvä* -stemmed participles are plurals and in the nominative case (Koskinen 1965: 173–174). Most of these are a part of a sentence's predicate verb. On the other hand this variant is common in the noun *kirjanoppinut* ('scribe' or 'scholar') and

5 Lauri Kettunen's dialect atlas has a map that describes the areal variation of the NUT participle in Finnish dialects; however, the variants shown in this map are not relevant from the perspective of this study because the forms are not vowel-stemmed (*nähnyt*, *näht*, *nähty*, *nähnä*, *nähnynnä*, *nähn*, from the verb *nähdä* 'to see') (Kettunen 1940: dialect map 32; the *h* between two vowels is described in maps 107–111).

10% of the words, which are in the nominative plural, are represented by the form *kirjanoppe-nuva-t* (Koskinen 1965: 174). This supports the degree of lexicalization of the *uva/yvä* variant. On the other hand, it is obvious that *kirjanoppinut* is a calque from the Swedish *Scrifftlärd* or the German *Schriftgelehrten* (Nummila 2011: 154). The large number of occurrences in this particular word may be caused by the fact that this variant was already on the decline in Agricola's time, and it was better retained in "term-like" words such as *kirjanoppenut* than in other lexemes (Penttilä 1965: 140). Moreover, the fact that occurrences of *ua/yä* and *uva/yvä* are found in Agricola's earlier translations (in the beginning of the New Testament, in the Gospel of Matthew) suggests that this variant was on the decline. This variant has no longer been in use in written Finnish after Agricola (Koskinen 1965: 174; Penttilä 1965: 140–141, see also Itkonen 1969: 124).

Koskinen (1965) focused on examining the occurrence of NUT participle variants in Agricola's different works. According to Koskinen (1965: 174) the *ee* variant (*luke-nee-t*) is more common in Agricola's New Testament and liturgical texts. The *ue/ye* (*luke-nue-t*) variant is common in Agricola's works that include parts of the Old Testament. *ue/ye* is also more common in Agricola's *Rucouskiria* ('prayer book') than the *ee* variant. Other variants especially common in *Rucouskiria* are *u*(*u*)/*y*(*y*) (*luke-nu*(*u*)*-t*) and *ehe* (*luke-nehe-t*).

3. Research data and methodology

3.1. Research data

The research data has been collected from the morphosyntactically annotated database of Agricola's Finnish (The Morpho-Syntactic Database of Mikael Agricola's Works version 1.1, Korp, https://www.kielipankki.fi/corpora/agricola/). The data was acquired by using an extensive search of the database with the following parameter: morphological analysis includes *act pcp2* (= NUT participle). The data set includes all of the NUT participles with vocal stems from the research corpus (N = 2066). False findings, for example consonant stemmed forms, were manually removed. Tense, polarity, grammatical person and mood only apply to verbs (and not to nouns, adjectives and verb chains), and thus with these variables N = 1670.

3.2. Variables

This study attempts to answer the question of whether the following phonological and morphological variables affect variant choice regarding vowel-stemmed NUT participles in Agricola's texts:

(i) Stem vowel
(ii) Vowel harmony
(iii) Consonant gradation (table presented in Appendix)
(iv) The number of syllables in the verbal stem
(v) Stem type
(vi) Word class
(vii) Case and grammatical number
(viii) Tense
(ix) Polarity
(x) Grammatical person
(xi) Mood (table presented in Appendix)

Finnish has the phonological phenomenon of consonant gradation, which involves the plosives *p*, *t* and *k*. Gradation affects the length of the plosive or its quality. Consonant gradation is rather common in the Finnish language, as it appears in one-third of the most common Finnish words. There are two variants concerning consonant gradation: strong and weak grades (e.g. *nukkua*, strong stem *nu**kk**u-*, weak stem *nu**k**u-* 'to sleep'). (On the various consonants and consonant combinations see VISK § 41.) Consonant gradation is based on the quality of a stem: the strong grade precedes an open syllable (ending with a vowel) (e.g. *nuk-ku-vat* 'they sleep') and the weak grade precedes a closed syllable (ending with a consonant) (e.g. *nu-kun* 'I sleep'). Our study shows that according to the two-sided Fisher exact test (Section 3.3) the presence or absence of consonant gradation will not affect variant preference (See Appendix).

Verbal mood (i.e. is the verb indicative or conditional) has a significantly greater p-value than 0.05, and therefore there is no statistically significant difference.

3.3. Methodology

The data has been organised in 11 tables, one for each variable. Each table contains the distribution of the words according to the variable. The question of whether there are statistically significant variations between the NUT participle types in each table has been examined using the two-sided Fisher exact test (Bewick et al. 2003). The purpose of the test is to assess whether there is a statistically significant difference in the variable value distribution for each NUT participle. Null hypothesis H_0 proposes that there is no statistical difference between the NUT participle types, and hypothesis H_1 proposes that there is a statistically significant difference, that is, the variable values tend to concentrate on some NUT participles. Fisher's exact test is similar to the more common Chi squared (χ^2) test, but the Fisher test is preferred in cases when several of the expected frequencies are below 5, which is the case here. The Fisher exact test gives a two-sided p-value. If the p-value is below 0.05, the null hypothesis H_0 can be dropped, and there is a significant difference. The following sections have all the tables providing what significant differences we have found. The other tables are presented in Appendix.

The tables show residual values that we have calculated. A large residual absolute value means that the observed value is substantially greater or less than expected. A residual absolute value over 2 or under -2 is boldfaced and it contributes more to the difference than a smaller absolute residual. A negative residual means that the observed value is smaller than expected.

The overall importance of the variables in selecting the NUT participle was examined by applying a machine learning classifier algorithm to the entire data set. A classifier is a machine learning model that learns the way a NUT participle was selected by Agricola. When presented with the features of a new word, the classifier can then make a prediction on the NUT participle Agricola would have used. The purpose of using the classifier is to find out which variables explain the selection of the NUT participle best. The classifier used in this article produces a diagram of the most important variables (see Section 5).

The efficacy of the classifier, that is, the success rate of predicting the correct NUT participle in the research material, can be estimated by using an AUC (Area Under Curve) metric. A perfect result is 1 and the lowest is 0. A value of 0.5 means that the same result could have

been reached by tossing a coin. A value of 1 is perfect, meaning that the classifier selects the correct class every time. The AUC for this classifier model was 0.92, which is close to perfect. The data was divided into training and test sets for the analysis. The test set was used to produce the AUC value.

A voting algorithm, which trains several classifiers and combines them to get a final result, was used. The classifiers used by the VotingEnsemble (Witten et al. 2016: 497) algorithm were Random Forest (Breiman 2001) and XGBoost (Chen & Guestrin 2016). 80% of the training data was used for training the model and 20% for validating it. The reported AUC value is for the test set. The classes in the training data were balanced by oversampling. The model was built using the AutoML method of Microsoft Azure Machine Learning environment. The code is available from the following link (https://github.com/jussalmi/nut-participle).

4. Results of NUT participle data

4.1. Stem vowel

All Finnish words have a vowel stem ending in a vowel and word stems can end with any Finnish vowel (*a*, *e*, *i*, *o*, *u*, *y*, *ä*, *ö*). If the NUT participle stem ends with two vowels, it is coded in Table 1 as *VV* which can represent a long double vowel or a diphthong. Finnish has 18 possible diphthongs (see VISK § 21). The stem vowel is the only variable affecting all NUT participle variants.

The main NUT participle variants in Agricola's Finnish are *ee* (*luke-nee-t*) and *ue*/*ye* (*luke-nue-t*). Phonological features of the verb in particular affect the choice of these variants in an extremely interesting way, as they show that they are in opposition to each other according to several variables: if a variable favours one, it will reject the other. The stem vowel affects the main NUT participle variants this way, making them an opposing pair: *e*-stemmed verbs favour the *ue*/*ye* variant and reject *ee* (e.g. more likely *luke-nue-t* than *luke-neet* 'to read'). In addition to this, *u*-stemmed verbs favour the *ee* variant (*nukku-neet* 'to sleep') and *o*- and *y*-stemmed verbs reject *ue*/*ye* (Table 1).

The other NUT participle variants are favoured by the following stem vowels: *y*-stemmed verbs favour the *uve*/*yve* and *ehe* variant

(e.g. *pysy-nyvet* or *pysy-nehet* 'to stay'), *o*-stemmed verbs favour the *ua/yä* and *u(u)/y(y)* variant (e.g. *rikko-nuat* or *rikko-nu(u)t* 'to break'), and *a*-stemmed verbs favour *uva/yvä* (e.g. *rakasta-nuvat* 'to love') (Table 1).

Table 1. Distribution of verbal stem vowels by participle type.

Variant	Stem vowel								
	a	e	i	o	u	vv	y	ä	Total
ee	164	281	20	65	62	137	44	90	863
	19.0%	32.6%	2.3%	7.5%	7.2%	15.9%	5.1%	10.4%	100.0 %
	1.5	**–3.4**	–1.2	1.4	**2.3**	1.0	1.7	0.4	
ue/ye	148	436	34	40	41	125	23	95	942
	15.7%	46.3%	3.6%	4.2%	4.4%	13.3%	2.4%	10.1%	100.0%
	–0.9	**3.2**	1.1	**–2.6**	–1.4	–1.0	**–2.4**	0.0	
uve/yve	2	22	2	0	0	4	7	5	42
	4.8%	52.4%	4.8%	0.0%	0.0%	9.5%	16.7%	11.9%	100.0%
	–1.9	1.3	0.7	–1.6	–1.5	–0.9	**4.1**	0.4	
ehe	4	22	1	3	5	12	6	4	57
	7.0%	38.6%	1.8%	5.3%	8.8%	21.1%	10.5%	7.0%	100.0%
	–1.8	–0.1	–0.5	–0.3	1.1	1.3	**2.5**	–0.7	
uä/yä	1	9	0	8	0	0	0	0	18
	5.6%	50.0%	0.0%	44.4%	0.0%	0.0%	0.0%	0.0%	100.0%
	–1.2	0.7	–0.7	**6.4**	–1.0	–1.6	–0.8	–1.3	
uvä/yvä	15	12	0	2	0	3	0	0	32
	46.9%	37.5%	0.0%	6.3%	0.0%	9.4%	0.0%	0.0%	100.0%
	4.1	–0.2	–1.0	0.0	–1.3	–0.8	–1.1	–1.8	
u(u)/y(y)	15	39	5	13	3	19	2	13	109
	13.8%	35.8%	4.6%	11.9%	2.8%	17.4%	1.8%	11.9%	100.0%
	–0.8	–0.7	1.0	**2.3**	–1.2	0.8	–1.1	0.6	
Total	349	821	62	131	111	300	82	207	2063

Fisher exact test
p value **<0.01**

The absolute number and percentage of word instances as well as the residual value (based on the relation between the observed and expected value) are shown in each table cell. The absolute residual values over 2 are in boldface, and they contribute more to the statistical significance.

4.2. Vowel harmony: front and back vowels

Finnish words and thus also Agricola's Finnish has vowel harmony, which means that front (*y*, *ä*, *ö*) and back (*a*, *o*, *u*) vowels cannot concurrently be in the same word. Front vowels *e* and *i* are neutral

regarding this phenomenon. The vowel type found in a word will affect if the NUT variant will be *uve*/*yve* (*luke-nuve-t*), *ua*/*yä* (*luke-nua-t*) or *uva*/*yvä* (*luke-nuva-t*). All these variants (*uve*/*yve*, *ua*/*yä* and *uva*/*yvä*) have developed from the same *ue*/*ye* (*luke-nue-t*) variant, and the vowel quality makes a difference between these variants.

The variable that all in all favours the *uve*/*yve* variant the most is the vowel quality in the word: front vowels strongly favour the *uve*/*yve* variant (e.g. *pysy-nyvet* 'to stay'), and back vowels reject it (Table 2). Moreover, front vowel *y*-stemmed verbs favour the *uve/yve* variant (Table 1). Back vowels favour the *ua*/*yä* and *uva*/*yvä* variants (e.g. *rakasta-nuat* or *rakasta-nuvat* 'to love'). In fact, these variants only occur in words with back vowels, and thus, the *yä* and *yvä* front vowel variants are theoretical (Table 2), so the front vowel variant is thus *uve*.

Table 2. Distribution of front and back vowels by participle type.

Variant	Front and back vowels		
	Front	Back	Total
ee	242	621	863
	28.0%	72.0%	100.0%
	–0.8	0.5	
ue/ye	290	652	942
	30.8 %	69.2 %	100.0 %
	0.7	–0.5	
uve/yve	33	9	42
	78.6 %	21.4 %	100.0 %
	5.8	**–3.8**	
ehe	16	41	57
	28.1 %	71.9 %	100.0 %
	–0.2	0.1	
uä/yä	0	18	18
	0.0 %	100.0 %	100.0 %
	–2.3	1.5	
uvä/yvä	0	32	32
	0.0 %	100.0 %	100.0 %
	–3.1	**2.0**	
u(u)/y(y)	29	80	109
	26.6 %	73.4 %	100.0 %
	–0.3	0.8	
Total	610	1453	2063

Fisher exact test
p value **<0.01**

4.3. The number of syllables in the verbal stem

The number of syllables is counted from the stem to which the NUT suffix is affixed: the consonant stem in two-stemmed verbs and the vowel stem in one-stemmed verbs.

The *ee* and *u*(*u*)/*y*(*y*) variants have a different kind of syllabic structure than the others variants, because the NUT participle marker in the *ee* and *u*(*u*)/*y*(*y*) variants is monosyllabic. In *luke.neet, luke.nu*(*u*)*t* and in the other variants, the NUT participle marker splits into two syllables: *luke.nu.et*, *luke.nu.vet*, *luke.nu.at*, *luke.nu.vat*, *luke.ne.het* ('to read').

The number of syllables in the verbal stem affects the main variants *ee* (*luke-nee-t*) and the *ue*/*ye* (*luke-nue-t*) and also the *u*(*u*)/*y*(*y*) (*luke-nu*(*u*)*-t*) variant. The main variants are in opposition to each other regarding the number of syllables in the following way: If the verb stem is monosyllabic, it favours *ue*/*ye* and rejects the *ee* variant, so *men-nye-t* is more likely to be than *men-nee-t*. If the verb stem is bisyllabic, it rejects the *ue*/*ye* variant (*lu.ke-nue-t*) (Table 3).

It seems that the *ue/ye* (*luke-nue-t*) variant could have been preserved with these monosyllabic verbs in order to conform to the more canonical trisyllabic structure of the NUT participles, and thus, for example, with the monosyllabic stem *kuol-* 'to die', the trisyllabic *kuol.lu.et* is more likely used than bisyllabic *kuol.leet*, and vice versa with the bisyllabic stem *lu.ke-* 'to read', the trisyllabic *lu.ke.neet* is more likely used than the quadrisyllabic *lu.ke.nu.et*.

If the stem has four syllables, such as *kun.ni.oit.ta-* 'to respect', it favours the *u(u)/y(y)* variant, which makes the NUT participle five-stemmed: *kun.ni.oit.ta.nu(u)t*. Moreover, other occurrences of the quadrisyllabic stem favour the same kind of pentasyllabic (and not hexasyllabic) NUT participle – namely *ee* variant – even though this is not statistically significant.

Furthermore in Ilja Moshnikov's (2021) study, which is a study on the variation of the NUT participle in certain Karelian dialects, and also uses a statistical methodology, the number of syllables in the verb stem have an effect on the preference of different variants. In general, the results of Moshnikov's study are comparable to the analysis of Agricola's Finnish only in small parts, because the variants in his study are so different from Agricola's language.

Table 3. Distribution of the number of syllables in the verbal stem by participle type.

Variant	The amount of syllables in the NUT-participle's stem				
	1	2	3	4	Total
ee	297	444	118	4	863
	34.4 %	51.4 %	13.7 %	0.5 %	100.0 %
	–3.0	1.9	1.7	0.6	
ue/ye	443	395	103	1	942
	47.0 %	41.9 %	10.9 %	0.1 %	100.0 %
	2.9	**–2.3**	–0.7	–1.2	
uve/yve	20	18	4	0	42
	47.6 %	42.9 %	9.5 %	0.0 %	100.0 %
	0.7	–0.4	–0.4	–0.4	
ehe	28	27	2	0	57
	49.1 %	47.4 %	3.5 %	0.0 %	100.0 %
	1.0	0.0	1.8	–0.4	
uä/yä	4	14	0	0	18
	22.2 %	77.8 %	0.0 %	0.0 %	100.0 %
	–1.2	1.9	–1.5	–0.2	
uvä/yvä	13	17	2	0	32
	40.6 %	53.1 %	6.3 %	0.0 %	100.0 %
	0.0	0.5	–0.9	–0.3	
u(u)/y(y)	39	56	12	2	109
	35.8 %	51.4 %	11.0 %	1.8 %	100.0 %
	–0.8	0.7	–0.2	**2.7**	
Total	844	971	241	7	2063

Fisher exact test
p value **<0.01**

4.4. Stem type

Finnish words can have either one or two stems. If a word is one-stemmed, it only has a vowel stem, which ends with a vowel (e.g. *elää*, stem *elä-* 'to live'). If a word is two-stemmed, it has both a vowel and consonant stem, the latter ending with a consonant (e.g. *mennä*, vowel stem *mene-* and consonant stem *men-* 'to go'). In two-stemmed words, the NUT participle marker affixes to the consonant stem (e.g. *men-nee-t*).

Contraction verbs are a special kind of two-stemmed verb type. In these kinds of verbs, a consonant between two vowels historically disappeared and caused a decrease in the number of syllables, for example *hypätä* (<**hypät-tä* < *hyppä-δä-n* 'to jump'). Moreover, the long vowel

in the vowel stem of these verbs (*hyppää-*) is caused by this loss of consonant (VISK § 330).

In Finnish language history, vowel stems were the originally and exclusively used type during the oldest known stages of development (Häkkinen 2002: 68). The *ue*/*ye* (*luke-nue-t*) NUT participle variant is historically older than *ee* (*luke-nee-t*), which developed from *ue*/*ye*. It seems that the *ee* variant in written Finnish was first used more frequently in prototypical verbs, which have the old one-stemmed verbal structure: if the verb is one-stemmed, it favours the *ee* variant and rejects *eu*/*ye*, and if the verb stem is two-stemmed, it favours the *ue/ye* variant and rejects *ee* (See Table 4).

Several previous studies have shown that words which are in frequent use tend to undergo sound change faster than less frequent words. Frequently used words are more prone to phonetic reduction, because their continual use exposes them to reductive processes. The reduction of words with high frequency of use reflects their predictability in the discourse (Bybee & Hopper 2001: 10, 17). According to Betty S. Phillips "changes which ignore the phonological integrity of segments and the morphological composition of words affect the most frequent words first. – – Changes which require analysis – whether syntactic, morphological, or phonological – during their implementation affect the least frequent words first; others affect the most frequent words first." (Phillips 2001, 124). The frequency of words goes hand in hand with their prototypicality.

Moreover, the stem vowel is connected to the existence of the consonant stem: *e*-stemmed words in particular have a consonant stem (e.g. *mennä*, stems *mene-* and *men-* 'to go') (Häkkinen 2002: 66), and the *e* stem rejects the *ee* variant like those words that have a consonant stem (see Section 4.1).

These factors also intertwine with the number of syllables in the verbal stem. Many very frequent verbs such as *mennä* 'to go' and *tulla* 'to come' have a monosyllabic consonant stem (*men-*, *tul-*, *kuol-*), and these stems also reject the *ee* variant. Contraction verbs, which are always two-stemmed, differ from other two-stemmed verbs. They favour the *ee* variant and this seems to be connected to they usually having a bisyllabic stem (*hy.pän-nee-t*). Bisyllabic stems reject the *ue*/*ye* variant.

All of this thus suggests a collusion between both the preservation of archaic variants in less regular and less paradigmatic types and prosodic factor *e*-stemmed NUT participles often having monosyllabic stem (*men-nye-t*) and contraction verb NUT participles usually having a bisyllabic stem (*hy.pän-nee-t*).

Table 4. Distribution of the number of verbal stems by participle type.

Variant	The number of stems of the verb			
	1 stem	2 stems	contr.	Total
ee	528	260	75	863
	61.2 %	30.1 %	8.7 %	100.0 %
	2.2	**–3.6**	**2.1**	
ue/ye	468	423	51	942
	49.7 %	44.9 %	5.4 %	100.0 %
	–2.4	**3.6**	–1.7	
uve/yve	17	21	4	42
	40.5 %	50.0 %	9.5 %	100.0 %
	–1.3	1.3	0.7	
ehe	28	23	6	57
	49.1 %	40.4 %	10.5 %	100.0 %
	–0.7	0.3	1.0	
uä/yä	12	6	0	18
	66.7 %	33.3 %	0.0 %	100.0 %
	0.6	–0.3	–1.1	
uvä/yvä	20	11	1	32
	62.5 %	34.4 %	3.1 %	100.0 %
	0.5	–0.3	–0.8	
u(u)/y(y)	72	33	4	109
	66.1 %	30.3 %	3.7 %	100.0 %
	1.5	–1.3	–1.3	
Total	1145	777	141	2063

Fisher exact test
p value **<0.01**

4.5. Word class

Finnish NUT participles can be part of finite compound verbs (from here on in, also shortly verbs) (example 9) or a verb chain (example 10) and can also be adjectives (example 11) and nouns (example 12).[6]

6 *kuolleet* 'dead' are analysed as adverbs in the Agricola corpus when they occur in the sentences such as *hen ylesnousi colluista* (Agricola's Works, Mt-27:64) 'he has been raised from the dead.' (NIV, Matthew 27:64). These are in the word class of nouns in this study.

9. *Te* *ole-tta* *minu-a* *etzi-nee-t* – –
you-PL is-PL2 I-PART search-NUT PARTIC-PL
(Agricola's Works, Jh-13:33)
– – you have looked for me – – (NIV, John 13:33)

10. – – *ette* *he* *vsko-isi-t* *Sinu-n* *minu-a*
that they believe-COND-PL3 you-GEN I-PART
lehette-nee-xi. (Agricola's Works, Jh-11:42)
send-NUT PARTIC-TRA
– – that they may believe that thou hast sent me. (NIV, John 11:42)

11. – – *yles-oiet-ca* – – *ne* *wäsy-nye-t* *Polue-t*
up-stretch out-IMP.PL3 those tired-NUT PARTIC-PL knee-PL
(Agricola's Works, Heb-12:12)
– – lift up – – your weak knees (NIV, Hebrews 12:12)

12. *Ja* *mine* *nä-i-n* *ne* *Coo-lue-t* *seke*
and I see-PST-SG1 those die-NUT PARTIC-PL both
swre-t *ette* *piene-t* – – (Agricola's Works, Ilm-20:12)
big-PL and small- PL
And I saw the dead, both great and small – – (NIV, Revelation 20:12)

This variable intertwines with grammatical cases because finite compound verbs are always in the nominative plural, and all other cases are a class other than a verb.

Word class affects the *ee*, *ehe, ue*/*ye* and *u*(*u*)/*y*(*y*) variants (*luke-nee-t*, *luke-nehe-t*, *luke-nue-t and luke-nu*(*u*)*-t*). The tendency for nouns and adjectives to reject the *ee* variant can be seen to be connected to the prototypical use of NUT participles as a part of compound verbs. The tendency that less frequent words are less prone to change is also relevant in this variable and in the less prototypical variants' rejection of the new variant (Bybee & Hopper 2001: 10, 17). However, the newer *ee* variant spread quite quickly in literary Finnish and displaced all the other variants by the 18th century (except the *ehe* variant, which was still used in literature, especially in poetry) (Rapola 1933: 165–166). The *ue*/*ye* variant is partially contrary to the *ee* variant, because nouns favour it (and reject *ee*).

On the other hand, *ehe* (*luke-nehe-t*) is a non-verbal variant, because adjectives, nouns and verbs, which are a part of a nonfinite verb chain, favour it, and verbs in a finite compound verbal structure reject it. With

this variable, the *ehe* variant is contrary to *ee* (*luke-nee-t*). The *u*(*u*)/*y*(*y*) (*luke-nu*(*u*)-*t*) variant is favoured by adjectives.

Table 5. Distribution of word class by participle type.

Variant	Word class				
	Verb	Adjective	Noun	Verb chain	Total
ee	732	11	90	30	863
	84.8 %	1.3 %	10.4 %	3.5 %	100.0 %
	1.3	**–2.0**	**–2.5**	0.6	
ue/ye	742	22	156	22	942
	78.8 %	2.3 %	16.6 %	2.3 %	100.0 %
	–0.7	0.0	**2.4**	–1.3	
uve/yve	40	0	2	0	42
	95.2 %	0.0 %	4.8 %	0.0 %	100.0 %
	1.0	–1.0	–1.6	–1.1	
ehe	25	6	16	10	57
	43.9 %	10.5 %	28.1 %	17.5 %	100.0 %
	–3.1	**4.1**	**3.0**	**6.2**	
uä/yä	17	0	1	0	18
	94.4 %	0.0 %	5.6 %	0.0 %	100.0 %
	0.6	–0.7	–1.0	–0.8	
uvä/yvä	26	0	6	0	32
	81.3 %	0.0 %	18.8 %	0.0 %	100.0 %
	0.0	–0.9	0.8	–1.0	
u(u)/y(y)	88	9	10	2	109
	80.7 %	8.3 %	9.2 %	1.8 %	100.0 %
	0.0	**4.1**	–1.3	–0.8	
Total	1670	48	281	64	2063

Fisher exact test
p value **<0.01**

4.6. Case and number

Finnish has two grammatical numbers, singular and plural, and 15 different cases. The following cases appear in the research material for this article: the nominative (e.g. *luke-nee-t* 'to read'), genitive (e.g. *luke-ne-i-den*), partitive (e.g. *luke-ne-i-ta*) translative (e.g. *luke-ne-i-ksi*) and essive (e.g. *luke-ne-i-na*) grammatical cases and the inessive (e.g. *luke-ne-i-ssa*), elative (e.g. *luke-ne-i-sta*), adessive (e.g. *luke-ne-i-lla*), ablative (e.g. *luke-ne-i-lta*) and allative (e.g. *luke-ne-i-lle*) local cases. In other plural forms, except for the nominative, *ee* is shortened to *e,* and

ue/*ye* often to *u*/*y* when preceding the plural marker *i* (*luke-nee-t : luke-ne-i-ta* and *luke-nue-t : luke-nu-i-ta*).

Since the NUT participle is always in the nominative plural when part of a compound verb, it mostly appears in this form throughout the data. Nouns and adjectives can be inflected in any case, either singular or plural, and verbs which are a part of a verb chain are inflected in certain cases only. Case and number therefore affect the *ee*, *ue*/*ye*, *ehe* and *u*(*u*)/*y*(*y*) variants (*luke-nee-t*, *luke-nue-t*, *luke-nehe-t* and *luke-nu*(*u*)*-t*) (Table 6).

The only morphological features which affect both the *ee* and *ue*/*ye* variants are case and number. Table 6 shows that this variable has a minor effect on the two most common variants: allative singular rejects the *ee* variant (*luke-nee-lla*) and ablative plural favours *ue*/*ye* (*luke-nue-i-lta*) and genitive plural rejects it (*luke-ne-i-den*).

The connection between case and number and the word class is seen in the *ehe* variant: Genitive singular (*luke-nehe-n*), partitive plural (*luke-neh-i-ta*), inessive singular (*luke-nehe-ssa*) and allative plural (*luke-nehe-i-lla*) all favour the *ehe* variant (these are noun variants), whereas the nominative plural (*luke-nehe-t* e.g. the verb variant) rejects it.

In addition the genitive plural (*luke-nu-i-den*) and the essive singular (*luke-nu(u)-na*) strongly favour the *u*(*u*)/*y*(*y*) variant.

Both this and Koskinen's study show that all *ua/yä* (*luke-nua-t*), *uva/yvä* (*luke-nuva-t*), *uve/yve* (*luke-nuve-t*) participles are in the nominative plural (Koskinen 1965: 173). The fact that variants used by Agricola are only in the nominative plural may be due to the fact that the use of these types was already on the decline during his time. These disappearing variants may have only been connected to specific morphological forms right before abandonment (Penttilä 1965: 143; Itkonen 1969: 152). On the other hand, this is not statistically meaningful as seen in Table 6.

Table 6. Distribution of case and grammatical number by participle type.

Variant	Case and grammatical number													
	nom pl	gen pl	ela pl	abl pl	part sg	tra sg	abl sg	gen sg	ade sg	ela sg	all sg	ine sg	ess sg	Total
ee	782	5	32	1	5	23	1	9	2	3	0	0	0	863
	90.6 %	0.6 %	3.7 %	0.1 %	0.6 %	2.7 %	0.1 %	1.0 %	0.2 %	0.3 %	0.0 %	0.0 %	0.0 %	100.0 %
	0.6	–0.7	–0.9	–1.6	–0.2	0.7	0.9	–1.7	1.3	0.6	**–2.0**	–0.9	–1.4	
ue/ye	833	2	53	8	2	19	0	14	0	2	8	1	0	942
	88.4 %	0.2 %	5.6 %	0.8 %	0.2 %	2.0 %	0.0 %	1.5 %	0.0 %	0.2 %	0.8 %	0.1 %	0.0 %	100.0 %
	–0.1	**–2.0**	1.8	**2.1**	–1.6	–0.5	–0.7	–0.8	–1.0	–0.2	1.6	0.1	–1.5	
uve/yve	42	0	0	0	0	0	0	0	0	0	0	0	0	42
	100.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	100.0 %
	0.8	–0.6	–1.4	–0.5	–0.5	–1.0	–0.1	–0.9	–0.2	–0.3	–0.5	–0.2	–0.3	
ehe	29	0	3	0	6	3	0	13	0	0	2	1	0	57
	50.9 %	0.0 %	5.3 %	0.0 %	10.5 %	5.3 %	0.0 %	22.8 %	0.0 %	0.0 %	3.5 %	1.8 %	0.0 %	100.0 %
	–3.0	–0.7	0.3	–0.5	**9.4**	1.5	–0.2	**11.6**	–0.2	–0.4	**3.3**	**4.0**	–0.4	
uä/yä	18	0	0	0	0	0	0	0	0	0	0	0	0	18
	100.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	100.0 %
	0.5	–0.4	–0.9	–0.3	–0.3	–0.6	–0.1	–0.6	–0.1	–0.2	–0.3	–0.1	–0.2	
uvä/yvä	32	0	0	0	0	0	0	0	0	0	0	0	0	32
	100.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	100.0 %
	0.7	–0.5	–1.2	–0.4	–0.4	–0.9	–0.1	–0.8	–0.2	–0.3	–0.4	–0.2	–0.3	
u(u)/y(y)	89	9	2	0	0	2	0	2	0	0	0	0	5	104
	85.6 %	8.7 %	1.9 %	0.0 %	0.0 %	1.9 %	0.0 %	1.9 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	100.0 %
	–0.3	**9.1**	–1.2	–0.7	–0.8	–0.2	–0.2	0.1	–0.3	–0.5	–0.7	–0.3	**9.5**	
Total	1,825	16	90	10	13	47	1	38	2	5	10	2	5	2064

Fisher exact test; p value **<0.01**

In Table 6, pl stands for plural and sg for singular.

4.7. Tense

Most of the NUT participles are a part of compound verbs, and the following variables apply only to the verbs: tense, polarity, and grammatical person (and mood). The following begins with tense.

The NUT participle can be a part of an imperfect (example 13), a perfect (example 14) or a pluperfect (example 15) compound.

13. *Ja* *e-uet* *he* *he-i-te* *wsko-nee-t*.
and neg-PL3 they they-PL-PART believe-NUT PARTIC-PL
(Agricola's Works, Lk-24:11)
– – and they believed them not. (NIV, Luke 24:11)

14. – – *se-n-woo-xi* *ette* *caiki* *o-uat* *syndi*
it-GEN-flux-TRA that everybody be-PL3 sin-PART
teh-nee-t.
do-NUT PARTIC-PL
(Agricola's Works, Rom-5:12)
– – for that all have sinned. (NIV, Romans 5:12)

15. – – *ia* *lwl-i* *ette* *Fangi-t* *ol-i-t*
and think-PST.SG3 that prisoner-PL be-PST-PL3
pois-pae-nee-t. (Agricola's Works, Ap-16:27)
away-run-NUT PARTIC-PL
– – he thought the prisoners had escaped. (NIV, Acts 16:27)

As Table 7 shows, perfect is the most common tense that includes the NUT participle. (For more on the use of different tenses and moods in 17th century biblical Finnish compared to older and contemporary Finnish, see Ikola 1949; Ikola 1950.) Tense affects the *ua*/*yä* (*luke-nua-t*) and the *uva*/*yvä* (*luke-nuva-t*) variants of the NUT participle: the imperfect strongly favours *ua*/*yä* and the perfect tense rejects both the *ua*/*yä* and *uva*/*yvä* variants (Table 7).

Table 7. Distribution of tense by participle type.

Variant	Tempus			
	Perfect	Pluperfect	Imperfect	Total
ee	415	119	198	732
	56.7 %	16.3 %	27.0 %	100.0 %
	–0.8	1.1	0.1	
ue/ye	467	99	176	742
	62.9 %	13.3 %	23.7 %	100.0 %
	1.5	–0.9	–1.6	
uve/yve	26	3	11	40
	65.0 %	7.5 %	27.5 %	100.0 %
	0.5	–1.2	0.1	
ehe	17	2	6	25
	68.0 %	8.0 %	24.0 %	100.0 %
	0.6	–0.9	–0.3	
uä/yä	**2**	1	14	17
	11.8 %	5.9 %	82.4 %	100.0 %
	–2.5	–0.9	**4.4**	
uvä/yvä	**7**	7	12	26
	26.9 %	26.9 %	46.2 %	100.0 %
	–2.2	1.5	2.2	
u(u)/y(y)	50	10	28	88
	56.8 %	11.4 %	31.8 %	100.0 %
	–0.2	–0.8	0.9	
Total	987	246	451	1,670

Fisher exact test; p value **<0.01**

When allomorphs develop, their variation tends to take on a semantic function (Paunonen 1976: 83). The *ua*/*yä* (*luke-nua-t*) variant is favoured by the imperfect tense and negative verbs (see Section 4.8). These variables intertwine, because all the imperfect forms that include the NUT participle are negative. In contrast, the perfect tense rejects both the *ua*/*yä* and *uva*/*yvä* (*luke-nuva-t*) variants. According to the machine learning classifier algorithm, tense is more important than polarity (see Section 5). Polarity and tense are statistically meaningful variables that also affect variation in Ilja Moshnikov's (2021: 142–143, 146) study.

4.8. Polarity

Finnish finite verbs can be either affirmative or negative (examples16–17). Polarity seems to be only a minor factor in the variation of

the NUT participle, because it only affects the use of the *ua*/*yä* (*luke-nua-t*) variant. Verbs used in the negative form favour the *ua*/*yä* variant and affirmative rejects it (Table 8). This is connected to tense, because all the imperfect forms are in the negative and, therefore, these features go hand in hand (the perfect and pluperfect tenses can be in both the affirmative and negative form).

16. *Ja* *nein* *te* *ole-tta* *vsko-nee-t.*
and so you- PL2 BE-PL2 believe-NUT PARTIC-PL
(Agricola's Works, 1Kor-15:11)
– – and so have ye believed. (NIV, 1 Corinthians 15:11)

17. – – *ia* *e-tte* *te* wsko-nua-t – – (Agricola's Works, Jh-3:12)
and neg-PL2 you- PL2 believe-NUT PARTIC-PL
– – and you did not believe – – (NIV, John 3:12)

Table 8. Distribution of polarity by participle type.

Variant	Polarity		
	Affirmative	Negative	Total
ee	494	238	732
	67.5 %	32.5 %	100.0 %
	–0.5	0.7	
ue/ye	537	205	742
	72.4 %	27.6 %	100.0 %
	1.1	–1.6	
uve/yve	29	11	40
	72.5 %	27.5 %	100.0 %
	0.3	–0.4	
ehe	17	8	25
	68.0 %	32.0 %	100.0 %
	–0.1	0.1	
uä/yä	4	13	17
	23.5 %	76.5 %	100.0 %
	–2.3	**3.4**	
uvä/yvä	13	13	26
	50.0 %	50.0 %	100.0 %
	–1.0	1.9	
u(u)/y(y)	58	30	88
	65.9 %	34.1 %	100.0 %
	–0.4	0.5	
Total	1152	517	1670

Fisher exact test
p value **<0.01**

4.9. Person

NUT participles that are a part of a finite compound verb and have a vowel stem can be in the first-person plural (example 18), second-person plural (example 19) or third-person plural (examples 20).

18. *Me ole-ma nin cauuan ele-nee-t.* (Agricola's Works, Vlt-2:16)
we be-PL1 so long live-NUT PARTIC-PL
'We have lived so long.'[7]

19. *Ja nein te ole-tta vsko-nee-t.*
and so you-PL2 be-PL2 believe-NUT PARTIC-PL
(Agricola's Works, 1Kor-15:11)
– – and so have ye believed. (NIV, 1 Corinthians 15:11)

20. *Tächti, ionga he ol-i-t neh-nye-t Idhe-s*
star which they be-PST-PL3 see-NUT PARTIC-PL east-INE
(Agricola's Works, Mt-2:9)
– – 'the star which they had seen in the East.'[8]

As shown in Table 9, there are three exceptions where the verb *olla* ('to be'), the auxiliary part of the verbal structure, is in the singular. These are ungrammatical forms and were most likely mistakes in Agricola's texts.

Person affects the use of the *ee*, *ua*/*yä* and *u*(*u*)/*y*(*y*) variants of the NUT participle in the following way: first-person plural strongly favours the *u*(*u*)/*y*(*y*) variant (*olemme luke-nu*(*u*)*-t* 'we have read') and rejects *ee* (*olemme luke-nee-t*) and makes these two variants opposite to each other in this variable; second-person plural forms favour the *ua*/*yä* variant (*olette luke-nua-t* 'you have read') and third-person plural rejects *u*(*u*)/*y*(*y*) (*ovat luke-nu*(*u*)*-t*) (Table 9).

7 The NIV translation differs from Agricola's translations: *we have seen it* (NIV Lamentations 2:16)

8 The NIV translation differs from Agricola's translations: *When they saw the star – – (NIV, Matthew 2:9)*

Table 9. Distribution of grammatical person by participle type.

Variant	Grammatical person					
	1st person plural	2nd person plural	3rd person plural	3rd person singular	1st person singular	Total
ee	91	123	518	0	0	732
	12.4 %	16.8 %	70.8 %	0.0 %	0.0 %	100.0 %
	–2.1	0.8	0.6	–0.9	–0.7	
ue/ye	108	105	526	2	1	742
	14.6 %	14.2 %	70.9 %	0.3 %	0.1 %	100.0 %
	–0.5	–1.0	0.7	1.2	0.8	
uve/yve	10	8	22	0	0	40
	25.0 %	20.0 %	55.0 %	0.0 %	0.0 %	100.0 %
	1.6	0.8	–1.1	–0.2	–0.2	
ehe	6	1	18	0	0	25
	24.0 %	4.0 %	72.0 %	0.0 %	0.0 %	100.0 %
	1.3	–1.4	0.0	–0.2	–0.1	
uä/yä	1	7	9	0	0	17
	5.9 %	41.2 %	52.9 %	0.0 %	0.0 %	100.0 %
	–1.0	**2.7**	–0.8	–0.1	–0.1	
uvä/yvä	5	6	15	0	0	26
	19.2 %	23.1 %	57.7 %	0.0 %	0.0 %	100.0 %
	0.5	1.0	–0.7	–0.2	–0.1	
u(u)/y(y)	35	11	42	0	0	88
	39.8 %	12.5 %	47.7 %	0.0 %	0.0 %	100.0 %
	5.9	–0.8	**–2.4**	–0.3	–0.2	
Total	255	261	1146	2	1	1670

Fisher exact test
p value **<0.01**

4.10. Summary of variables and their effect on NUT participle variants

To a different extent, different variables affect the NUT participle variants in the following way: The stem vowel is the most important variable, for it is the only variable which affects all variants. Word class and case and number of the NUT participle affect four variants, and if the verb has front or back vowels, stem type and grammatical person of the verb affect the use of three variants. Table 10 summarises how the different variables affect the selection of the different variants.

Table 10. Effect of variables on variant selection.

Variable		Variant						
		ee	ue/ye	uve/yve	ua/yä	uva/yvä	ehe	u(u)/y(y)
stem type	fav	one, contr.	two					
	rej	two	one					
front/back vowel	fav			front		back		
	rej			back	front	front		
stem vowel	fav	u	e	y	o	a	y	o
	rej	e	o, y					
no of stem's syllables	fav		1					4
	rej	1	2					
word class	fav		noun				verb chain, adj, noun	adj.
	rej	noun, adj					verb	
case and number	fav		abl pl				gen sg, part sg, ine sg, all sg	ess sg, gen pl
	rej	all sg	gen pl				nom pl	
tense	fav				imp			
	rej				perf	perf		
polarity	fav				neg			
	rej				aff			
person	fav				pl 2			pl 1
	rej	pl 1						pl 3

The main NUT participle variants in Agricola's Finnish are ***ee*** and ***ue*/*ye*** (*luke-nee-t* and *luku-nue-t*). Phonological features of the verb in particular affect the choice of these variants in an extremely interesting way, as they show that they are in opposition to each other according to several variables: if a variable favours one, it will reject the other. The variants are in opposition to each other regarding the following features, which intertwine as seen in previous sections: 1) If a verb's stem is monosyllabic, it favours the *ue*/*ye* variant and rejects *ee*, and if the stem is bisyllabic, it rejects the *ue*/*ye* variant. 2) If the verb only has one stem, it favours the *ee* variant and rejects *ue/ye*, and if the verb has two stems, it favours the *ue*/*ye* variant and rejects *ee*. (Moreover, contradictive stems favour the *ee* variant.) 3) *e*-stemmed verbs favour the *ue*/*ye* variant and reject *ee*.

Several variables affect the ***ua*/*yä* and *uva*/*yvä* variants** (*luke-nua-t* and *luke-nuva-t*) in the same way, which is not surprising, because *uva*/*yvä* developed from *ua*/*yä*. First, vowel quality affects both the *ua*/*yä* and *uva*/*yvä* variants. In fact, these variants only occur in words with back vowels, and thus, the *yä* and *yvä* front vowel variants are theoretical. In addition to this, verbal tense affects the use of these variants. The perfect tense rejects both the *ua*/*yä* and *uva*/*yvä* variants, and the imperfect strongly favours *ua*/*yä*. The ***uve/yve* variant** is contrary to these two variants, because it is the front vowel variant whereas the *ua*/*yä* and *uva/yvä* are rejected by front vowels and *uva*/*yvä* is favoured by back vowels.

The ***ehe* variant** (*luke-nehe-t*) is a firmly non-verbal variant. Adjectives, nouns and especially verbs which are a part of a nonfinite verb chain favour the *ehe* variant, but verbs in a finite structure rejects it. The analogy of the *ehe* element found in nouns may have affected the use of these non-verbal *ehe* NUT participles. The inflectional model is found in noun *ee*-stems (e.g. *huonee-*, **huonehe-* 'room'), where **-h-* has a clear etymological origin. With this variable, the *ehe* variant is contrary to *ee* (*luke-nee-t*), which is rejected by both nouns and adjectives. Several cases, which occur in nouns, favour the *ehe* variant, whereas the nominative plural (e.g. the verb variant) rejects it. The only phonological feature affecting the use of the *ehe* variant is the stem vowel of the verb.

The ***u*(*u*)/*y*(*y*) variant** (*luke-nu*(*u*)*-t*) is similar to *ehe*, because the same morphological features mainly affect the use of this variant: word class, case and number and the stem vowel. In addition to this, person also affects the use of the variant: while first-person plural strongly favours the *u(u)/y(y)* variant, and quadrisyllabic stems also favour it, third-person plural rejects it. On the other hand *u*(*u*)/*y*(*y*) is opposite to the *ee* variant: first-person plural forms favour *u*(*u*)/*y*(*y*) and reject the *ee* variant. They are also contrary to one another according to the word class of adjectives, which favour the *u*(*u*)/*y*(*y*) variant.

5. The overall importance of the variables by machine learning

A machine learning classification method was used to find out the importance of the individual variables, as described in Section 3.3. The algorithm learns to predict the type of NUT participle in Agricola's texts based on the variables. In doing so, the algorithm calculates the importance of the different variables. The more important the variable, the more it influences the choice of participle type.

The machine learning classification model describes the grammatical relations in Agricola's Finnish, but it is also meant to describe other language forms close to Agricola's. The closer the language is to this data, the better the model is for it. Thus, this model well describes 16th century literary Finnish in general as well. In the 17th century, written Finnish had changed so much that it is very different from the language examined in this study. Agricola's Finnish is based on the colloquial language that had been spoken in southwest Finland, especially the Finnish used in the Turku region. The model might also give a good insight on 16th century colloquial Finnish and the language prior to that in the southwestern dialect area.

The findings show that the most meaningful feature in Agricola was the stem vowel, which is also the only variable affecting the choice of every variant (see Section 4.10), followed by vowel harmony, that is if the verb has front or back vowels, person and tense (Figure 1).

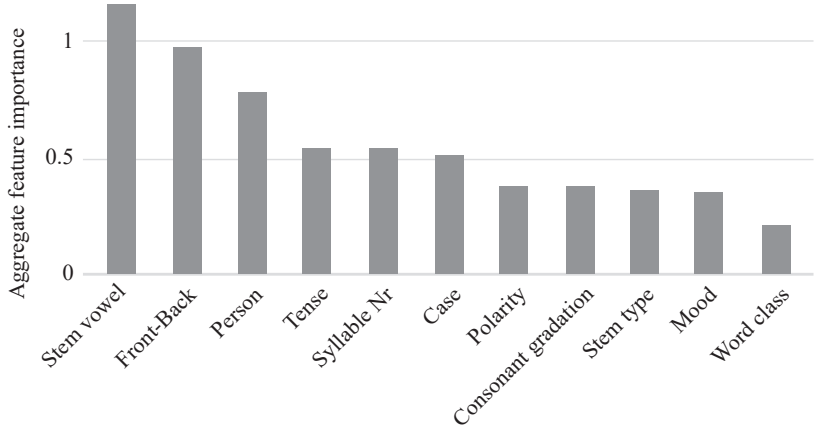

Figure 1. Variable importance in selecting a participle. The X axis shows the weight of the individual variable in the model.

Several phonological variables affecting the use of NUT participle variants – the NUT participle's stem vowel, whether the NUT participle has front or back vowels, the number of syllables in the stem, the stem type, and also consonant gradation – are partially connected to each other, but their importance in the selection of the NUT participle variant is different. These phonological variables, aside from front and back vowels, especially affect the selection of the main variants of the NUT participle *ue*/*ye* (*luke-nue-t*) and *ee* (*luke-nee-t*). Overall, the stem vowel is the most important of these variables, followed by the number of syllables. Consonant gradation was not an important variable according to the Fisher exact test, but it is more important than the stem type according to the machine learning classification, even though their weight in the selection of the variant is rather similar.

The second most important variable, front and back vowel, pairs *uve*/*yve* (*luke-nuve-t*) variant with *uva*/*yvä* (*luke-nuva-t*) and *ua*/*yä* (*luke-nua-t*). *uve*/*yve* is the front vowel variant and the latter ones are back vowel variants. The most important difference between the *uva*/*yvä* and *ua*/*yä* variants is that the newer *ua*/*yä* is strongly favoured by the imperfect tense (the fourth most important variable). Tense and polarity partially intertwine, but tense is more of an important variable than polarity, which is one of the least important variables.

Person is the third most important variable. Both person and particularly tense are more important variables than the Fisher exact test would suggest (they affect only two variants). The *ee* (*luke-nee-t*) and *u*(*u*)/*y*(*y*) (*luke-nu*(*u*)*t*) variants are opposite to each other in regard to person. Variants are also contrary to each other according to the word class, which is the least important variable in the overall scheme, even though according to the two-sided Fisher exact test, it affects the use of four variants. The case and number variable, which is connected to word class, seems to be the key factor in the *ehe* variant. It is more important in the overall scheme than word class, but the weight of case and number in the selection of the NUT participle type also seems less important than it would be considered to be by only looking at the Fisher exact test.

6. Conclusions

In the use of different linguistic variants, there are tendencies and preferences which language users are not aware of. This study shows how different phonological and morphological variables actually influence the choice of NUT participle in Agricola's Finnish.

Both phonological and morphological factors affect the preference of NUT participle variants. All in all, the quality of the vowels seems to be the most important factor in the selection of the NUT participle variant: the stem vowel and if the word has front or back vowels. The third and fourth most important variables, person and tense, are morphological. When allomorphs develop, their variation tends to take on a semantic function (Paunonen 1976: 83), and thus different morphological categories attract different variants: different grammatical persons and tenses especially favour different NUT participle variants.

We discovered that NUT participles variants form an extremely interesting network: variables pair or group variants so that variables favour a certain variant (or variants) and alternatively reject another (or others). For example, the main NUT participle variants *ee* and *ue*/*ye* (*luke-nee-t* and *luke-nue-t*) are contrary to each other in terms of the stem vowel, the number of syllables in the stem, stem type, and also in regard to word class, all in this order of importance. Moreover, also other variants pair or group, and different variables are important in the selection between certain variants.

This study suggests that there is a preservation of archaic variants in less prototypical NUT participle types and that there is first more of a use of the newest *ee* variant in prototypical verbs which have the old one-stemmed verbal structure. The stem type is also connected to the stem vowel. There is also a tendency towards a prototypical syllable structure in the selection of NUT participles. This is in line with previous studies showing that words which are frequent in use usually undergo a sound change faster than less frequent words.

We have examined the effect of phonological and morphological variables on NUT participle variants, and the possible effect of analogy, for instance, is beyond the scope of this study. For example, the analogy of the *ehe* element found in nouns may have affected the use of *ehe* NUT participles, which are often nouns or adjectives. It should also

be noted that other intra- and extra-linguistic factors affect the use of different forms as well, but they too are beyond the scope of this study.

There are archaic words and forms in Agricola's Finnish which have declined from the language. Also most NUT participle variants are such archaic forms. Some of the remnants seem to have been in active use in the 16th century but have soon declined afterwards. Such archaic linguistic features are a result of historical development and show interesting points of language change, and being remnants, they are a window to the earlier stages of the development of the Finnish language (Kolehmainen 2014: 53–54; Mielikäinen 2014: 32; Häkkinen 2015: 127).

Even though the data of this study only includes Agricola's works and gives thus an important insight into the variation of the vowel-ending NUT participle in his Finnish, the machine learning classification model also describes other language forms close to Agricola. The closer the language is to Agricola's Finnish, the better the model works for it. Thus, this model can also describe 16th century literary Finnish in general and furthermore possibly give insight to 16th century colloquial Finnish and how southwestern Finnish was spoken prior to the 16th century, the dialect on which Agricola's Finnish was based.

Acknowledgments

We would like to warmly thank Associate Professor Kirsi-Maria Nummila and both reviewers for their profound work reading the article and their valuable comments, which made the article significantly better. We also thank Leonard Pearl for excellent proofreading of the text. This study was funded by the Academy of Finland (Project 348899).

References

Agricola's Works: *The Morpho-Syntactic Database of Mikael Agricola's Works version 1.1, Korp* [text corpus]. Kielipankki. http://urn.fi/urn:nbn:fi:lb-2019121804. (4 May, 2023).

Anttila, Arto. 2002. Variation and phonological theory. In J. K. Chambers, Peter Trudgill & Natalie Schilling-Estes (eds.), *The Handbook of Language Variation and Change*, 206–243. Malden Mass: Blackwell Publishing. https://doi.org/10.1002/9780470756591.ch8.

Bayley, Robert. 2002. The quantitative paradigm. In J. K. Chambers, Peter Trudgill & Natalie Schilling-Estes (eds.), *The Handbook of Language Variation and Change*, 117–141. Malden Mass: Blackwell Publishing. https://doi.org/10.1002/9780470756591.ch5.

Bewick, Viv, Liz Cheek & Jonathan Ball. 2003. Statistics review 8: Qualitative data – tests of association. *Critical Care* 8. 46–53. https://doi.org/10.1186/cc2428.

Breiman, Leo. 2001. Random Forests. *Machine Learning* 45. 5–32. https://doi.org/10.1023/A:1010933404324.

Bybee, Joan & Paul Hopper. 2001. Introduction to frequency and the emergence of linguistic structure. In Joan Bybee & Paul Hopper (eds.), *Frequency and the Emergence of Linguistic Structure*, 1–24. Typological Studies in Language 45. Amsterdam/Philadelphia: John Benjamins Publishing Company. https://doi.org/10.1075/tsl.45.

Chen, Tianqi & Carlos Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. *KDD'16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785.

Elsayed, Duha. 2017. *Agricolan* pitää – *nesessiiviverbi kielikontaktin ytimessä*. Turku: Turun yliopisto. https://urn.fi/URN:ISBN:978-951-29-7028-5.

Feagin, Crawford. 2002. Entering the Community. In J. K. Chambers, Peter Trudgill & Natalie Schilling-Estes (eds.), *The Handbook of Language Variation and Change*, 20–39. Malden Mass: Blackwell Publishing. https://doi.org/10.1002/9780470756591.ch1.

Forsman Svensson, Pirkko. 1992. *Vanhan kirjasuomen nominaalirakenteista*. Stockholm: Stockholm University.

Häkkinen, Kaisa. 1994. *Agricolasta nykykieleen. Suomen kirjakielen historia*. Porvoo: WSOY.

Häkkinen, Kaisa. 2002. *Suomen kielen historia 1. Suomen kielen äänne- ja muotorakenteen historiallista taustaa.* Turku: Turun yliopisto.

Häkkinen, Kaisa. 2015: *Spreading the Written Word. Mikael Agricola and the Birth of Literary Finnish.* Studia Fennica Linguistica 19. Helsinki: Suomalaisen Kirjallisuuden Seura.

Hakulinen, Lauri. 1979. *Suomen kielen rakenne ja kehitys*. 4., korjattu ja lisätty painos. Helsinki: Otava.

Ikola, Osmo. 1949. *Tempusten ja modusten käyttö ensimmäisessä suomalaisessa raamatussa verrattuna vanhempaan ja nykyiseen kieleen I.* Turku: Turun yliopisto.

Ikola, Osmo. 1950. *Tempusten ja modusten käyttö ensimmäisessä suomalaisessa raamatussa verrattuna vanhempaan ja nykyiseen kieleen II. Indikatiivin imperfektin ja perfektin sekä imperfektiin perustuvien liittomuotojen temporaalinen käyttö*. Turku: Turun yliopisto.

Ikola, Osmo, Ulla Palomäki & Anna-Kaisa Koitto. 1989: *Suomen murteiden lauseoppia ja tekstikielioppia*. Suomalaisen Kirjallisuuden Seuran toimituksia 511. Helsinki: Suomalaisen Kirjallisuuden Seura.

Itkonen, Terho. 1969. Erään partisiippityypin taustaa. *Virittäjä* 73(2). 118–155.

Kager, René. 1999. *Optimality Theory.* Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511812408.

Kettunen, Lauri. 1940. *Suomen murteet III A. Murrekartasto*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Koivisto, Helinä. 1987. *Partisiippien adjektiivistuminen suomen kielessä*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Koivisto, Vesa. 2013. *Suomen sanojen rakenne*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Kolehmainen, Taru. 2014. *Kielenhuollon juurilla. Suomen kielen ohjailun historiaa.* Suomi 204. Helsinki: Suomalaisen Kirjallisuuden Seura.

Kopf, Kristin & Weber, Thilo. 2023. *Free Variation in Grammar: Empirical and theoretical approaches.* Studies in Language Companion Series 234. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/slcs.234.

Koskinen, Juhani. 1965. Tilastollisia havaintoja Mikael Agricolan aktiivin 2. partisiipista. *Virittäjä* 69(2). 172–175.

Labov, William. 2004. Quantitative analysis of linguistic variation. In Ulrich Ammon, Norbert Dittmar, Klaus J. Mattheier & Peter Trudgill (eds.), *Sociolinguistics: an International Handbook of the Science of Language and Society vol. 1*, 6–21. Berlin: De Gruyter. https://doi.org/10.1515/9783110141894.1.1.6.

Lehikoinen, Laila & Silva Kiuru. 2006. *Kirjasuomen kehitys*. Helsinki: Helsingin yliopisto.

Mielikäinen, Aila. 1982. Nykypuhesuomen alueellista taustaa. *Virittäjä* 86(3). 277–294.

Mielikäinen, Aila. 1995. Morfologinen diffuusio. *Virittäjä* 99(3). 321–336.

Mielikäinen, Aila. 2014. Uskonnollisen suomen kielen kehitys Agricolasta nykypäivään. In Riikka Nissi & Aila Mielikäinen (eds.), *Sanaa tutkimassa. Näkökulmia uskonnolliseen kieleen ja sen käyttöön*, 29–68. Helsinki: Suomalaisen Kirjallisuuden Seura.

Moshnikov, Ilja. 2021. *NUT-partisiipin variaatio rajakarjalaismurteissa*. Joensuu: Itä-Suomen yliopisto. https://doi.org/10.23982/vir.109388.

NIV = Bible, New International Version. https://www.biblegateway.com.

Nummila, Kirsi-Maria. 2011. *Tekijännimet Mikael Agricolan teosten kielessä. Henkilötarkoitteisten johdosten merkitykset, funktiot ja rakenteet.* Turku: Turun yliopisto. https://urn.fi/URN:ISBN:978-951-29-4866-6.

Numminen, Paavo. 1954. Aktiivin partisiippien käytöstä ja merkityksestä. *Virittäjä* 58(2). 154–172.

Nuolijärvi, Pirkko & Hanna Lappalainen. 2020. Sosiolingvistiikka. In Milla Luodonpää-Manni, Markus Hamunen, Reetta Konstenius, Matti Miestamo, Urpo Nikanne & Kaius Sinnemäki (eds.), *Kielentutkimuksen menetelmiä I–IV. Esimerkkitapauksia menetelmällisestä kirjosta kielentutkimuksen alueilla, osa IV*, 864–895. Helsinki: Suomalaisen Kirjallisuuden Seura. https://doi.org/10.2307/j.ctv1qp9hgb.29.

Paunonen, Heikki. 1976. *Allomorfien dynamiikkaa*. Helsinki: Helsingin yliopisto.

Paunonen, Heikki. 1984. Vapaasta vaihtelusta. In Heikki Paunonen & Päivi Rintala (eds.), *Nykysuomen rakenne ja kehitys. 2, Näkökulmia kielen vaihteluun ja muuttumiseen,* 139–154. Helsinki: Suomalaisen Kirjallisuuden Seura.

Penttilä, Aarni. 1965. Vanhan kirjasuomen tulluvat ’tulleet’ -tyyppinen partisiippi. *Virittäjä* 69(2). 136–144.

Phillips, Betty S. 2001. Lexical diffusion, lexical frequency, and lexical analysis. In Joan Bybee & Paul Hopper (eds.), *Frequency and the Emergence of Linguistic Structure*, 123–136. Typological Studies in Language 45. Amsterdam/Philadelphia: John Benjamins Publishing Company. https://doi.org/10.1075/tsl.45.

Rapola, Martti. 1933. *Suomen kirjakielen historia pääpiirteittäin. 1, Vanhan kirjasuomen kirjoitus- ja äänneasun kehitys*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Rapola, Martti. 1963. *Henrik Hoffman. Puristinen kielenkorjaaja*. Suomi 110:3. Helsinki.

Salmi, Heidi. 2010. *Mikael Agricolan teosten kielen* ala-*,* ylä- *ja* pää-*vartaloiset adpositiot*. Turku: Turun yliopisto. https://urn.fi/URN:ISBN:978-951-29-4506-1.

Setälä, E. N. 1887. *Zur Gexchichte der Tempus- und Modusstammbildung in dern finnisch-ugrischen Sprachen*. SUSA.

Tunkelo, E. A. 1939. Lisiä suomen ja karjalan II partisiipin historiaan I. *Virittäjä* 43. 165–184.

VISK = Hakulinen, Auli, Maria Vilkuna, Riitta Korhonen, Vesa Koivisto, Tarja Riitta Heinonen & Irja Alho. 2004. *Iso suomen kielioppi*. http://scripta.kotus.fi/visk/etusivu.php (15 April, 2024.).

Witten, I. H., Eibe F., Hall, M. A. & Pal, C. J. 2016. *Data Mining: Practical Machine Learning Tools and Techniques,* 4th Edition. Morgan Kaufmann.

Kokkuvõte. Heidi Salmi, Jussi Salmi: Soome keele vokaaltüvelise NUT -kesksõna varieerumine Mikael Agricola töödes. Soome keelt hakati kirjutama 16. sajandil ja Mikael Agricola tööd esindavad vanimat soome kirjakeelt. Agricola keel erineb mitmel moel tänapäeva soome keelest. Käesolevas artiklis uuritakse soome keele NUT-kesksõna variatsiooni, mida Agricla on kasutanud, kuid mis tänapäeva soome keeles enam ei esine.

Artiklis käsitletakse küsimust kas keelematerjalis on näha statistiliselt olulisi keeleliste muutujate põhjustatud NUT-kesksõna tüüpide variante. Uurimus osutab, et häälikulised muutujad mõjutavad NUT-kesksõna variatsiooni rohkem kui morfoloogilised muutujad. NUT-kesksõnade variandid moodustavad huvitava võrgustiku: muutujad loovad paare või rühmi nii, et teatud variant (variandid) on eelistatumad ja vastavalt teised mitte.

Muutujate laiemat tähendust kesksõna valimisel on vaadeldud ka masinõppimise klassifitseerimise algorütmi kasutades. Siin osutus kõige tähtsamaks omaduseks tüvevokaal, sõnas esinev taga- või eesvokaal või ajatunnus.

Märksõnad: NUT-kesksõna, vana soome kirjakeel, Mikael Agricola soome keel, korpuslingvistika, masinõppimine

APPENDIX. Distribution of the consonant gradation and mood by participle type.

Table 1. Distribution of consonant gradation by participle type.

Variant	Consonant gradation		
	No	Yes	Total
ee	414	449	863
	48.0 %	52.0 %	100.0 %
	–1.5	1.5	
ue/ye	513	429	942
	54.5 %	45.5 %	100.0 %
	1.2	–1.3	
uve/yve	20	22	42
	47.6 %	52.4 %	100.0 %
	–0.4	0.4	
ehe	29	28	57
	50.9 %	49.1 %	100.0 %
	–0.1	0.1	
uä/yä	13	5	18
	72.2 %	27.8 %	100.0 %
	1.2	–1.3	
uvä/yvä	17	15	32
	53.1 %	46.9 %	100.0 %
	0.1	–0.1	
u(u)/y(y)	58	51	109
	53.2 %	46.8 %	100.0 %
	0.2	–0.3	
Total	1064	999	2063

Fisher exact test
p value **0.1**

Table 2. Distribution of mood by participle type.

Variant	Modus		
	Indicative	Conditional	Total
ee	709	23	732
	96.9 %	3.1 %	100.0 %
	0.0	0.5	
ue/ye	724	18	742
	97.6 %	2.4 %	100.0 %
	0.2	–0.6	
uve/yve	39	1	40
	97.5 %	2.5 %	100.0 %
	0.1	–0.1	
ehe	25	0	25
	100.0 %	0.0 %	100.0 %
	0.2	–0.8	
uä/yä	17	0	17
	100.0 %	0.0 %	100.0 %
	0.1	–0.7	
uvä/yvä	24	2	26
	92.3 %	7.7 %	100.0 %
	–0.2	1.5	
u(u)/y(y)	85	3	88
	96.6 %	3.4 %	100.0 %
	0.0	0.3	
Total	1615	47	1670

Fisher exact test
p value **0.61**