

## THE DEFINITE<sup>1</sup> ARTICLE IN VOTIC: THE PROCESS OF GRAMMATICALISATION

**Tatiana Agranat**

*Institute of Linguistics of Russian Academy of Sciences*

**Abstract.** The paper is devoted to the grammaticalisation of the deictic pronoun *se* into a definite article in Votic. The author traces this process diachronically demonstrating the results of her studies with the examples of a folktale written down by August Ahlqvist in the middle of the 19th century, and of a text written down by herself at the beginning of the 21st century. The reason for article-like usage of *se* in Votic narrative in the 19th century seems to be the same as in Finnish narratives in the 19th century, as we can judge from Laury (1991): NPs that are prefaced with *se* have references that are important in some way for the narrative. In modern Votic *se* lost its function of marking NPs as prominent and became the marker of identifiability. The author draws the conclusion that the grammaticalisation of *se* in Votic is not yet completed, because *se* is not compulsory.

**Keywords:** Votic, definite article, grammaticalisation, deictic pronouns

**DOI:** <http://dx.doi.org/10.12697/jeful.2015.6.1.03>

### 1. Introduction

It is well known that among the Uralic languages only Hungarian has a true article and in Mordvinian there is a morpheme which is sometimes called the postpositive article. Nevertheless it was noticed rather long ago that in spoken<sup>2</sup> Finnish and Estonian the demonstrative pronoun *se* (Fin.) = *see* (Est.), while it has not already become the article, is at least in the process of grammaticalisation. On the Estonian

---

1 The development of the category of definiteness/indefiniteness sorts with the general trend of the diachronic typology: the movement goes from concrete meaning to generalised. That is why means of expression of definiteness, including definite article, appear in languages earlier than means of expression of indefiniteness. There is still nothing in Votic that could look like an indefinite article.

2 In old written Finnish and Estonian the process of definite article grammaticalisation was overcome otherwise; it has been influenced by Germanic languages contacts (see Laury and Pajusalu 2010; Metslang 2011).

demonstrative pronoun *see*, Renate Pajusalu, who has studied its article-like usage in the spoken Estonian, concludes that the use of the pronoun is not obligatory, “which shows that the article has not become fully grammaticalized” (Pajusalu 1997: 173).

As regards Finnish, some authors consider *se* as an article (for example, Laury 1991<sup>3</sup>, 1996), some others admit that it is well-nigh compulsory, but is not yet fully grammaticalised (Chesterman 1991).

In the Votic language there is also the demonstrative pronoun *se* (plural *ne* or *ned*), which means distal deixis, anaphora and exophora<sup>4</sup>. In Votic, as well as in Finnish and Estonian, the pro-adverbs of distal deixis *siäl/sinne/siält*, instead of the demonstrative pronoun *se*, are used with nouns of some local meaning. These adverbs also have anaphoric and exophoric uses. The system of Votic deictic pronouns used to be binominal in the 19th century and today remains the same (*se* is used as the distal deixis marker and *kase* is used as the proximal one), and the system of deictic adverbs is more complicated and it has changed from the 19th century (for details see Agranat 2008).

## 2. The development of the definite article *se* in spoken Finnish

Ritva Laury (1991) compares the 19th century spoken data to the present-day discourse in Finnish. Her data are two narratives from the 1890s (traditional folktales), which she compares with a spoken narrative recorded in 1984.

When studying the folktales, Laury finds that the two features which best predict the use of *se* pre-nominally are accessibility and noteworthiness. Speakers most likely mark by *se* those NPs whose references are accessible from previous discourse context; occasionally, given NPs are also marked with *se*, but not as often as accessible NPs (Laury 1991: 100).

Laury determines given and accessible NPs according to Chafe: a particular concept at a particular point in time “may be in any one of three different activation states” (Chafe 1987: 25). Information which the speaker assumes to be in the active or focal consciousness of the addressee at the time of speaking, given information is expressed

3 The demonstrative pronoun *se* is being grammaticalised in Finnish into a definite article so that speakers now regularly mark identifiable noun phrases with *se* (Laury 1991: 94).

4 About the distinction between deictic, exophoric and anaphoric uses of demonstratives see for example Kibrik 2011: 37.

differently from the new information, which the speaker assumes the addressee not to be conscious of, and accessible information, which the addressee is peripherally conscious of (Laury 1991: 96).

Additionally, according to her data, in 1890s narratives NPs which are prefaced with *se* have references which are important in some way for the narrative. Thus an NP which is preceded with *se* is typically an NP whose referent is a central character in the story or otherwise a crucial prop. She also notes the clustering of *se*-marked NPs at a crucial point in the narrative. Thus *se*-marked NPs don't only occur with noteworthy participants in a narrative, but they also seem to cluster in foregrounded clauses.

Then, in present-day spoken Finnish *se* has been grammaticalised as an article so that speakers now regularly mark identifiable nouns with *se*. Laury describes this process in that way. From presumably having earlier functioned as a pure demonstrative which picks out new referents by pointing out their concrete spatial location, by the 19th century *se* had come to have a highlighting function in discourse; it was used with prominent accessible referents. Accessible and given concepts share the feature that they are identifiable; and languages which have articles, such as English, mark both accessible and given NPs as definite. By becoming associated with the feature of identifiability, *se* has undergone reanalysis and has been grammaticalised as a definite article in spoken Finnish. After *se* became the marker of identifiability, it lost its function of marking NPs as prominent. (Laury 1991: 115)

### 3. Votic data analysis

Votic is an unwritten language although some texts have been written down by researchers since the 19th century. So Votic only has a spoken variant. Although there were quite a lot of texts collected between the middle of the 19th century and today, however in this paper I will follow Laury (1991) just to collate the data of two closely related languages and I will compare the usage of *se* in folktales with its usage in modern spoken narratives.

I will demonstrate the results of my studies with an example of the folktale “Kuptsaa’ emännää’ poika” (‘The merchant’s wife’s son’)

written down by August Ahlqvist (Ahlqvist 1856)<sup>5</sup>, and a text I wrote down recently during my fieldwork with speakers of Votic.

### 3.1. The 19th century data analysis

Let us first analyse the folktale.

A merchant spent seven years buying goods in another kingdom; he went across many seas and then was heading home. But when he was just one hundred versts from home, his cart got stuck in the ground and his horse could not pull it any more.

1. Õl-i-ø      liika    jumalilliinee'    se    kuptsa    nõis-i-ø  
 be-PST-3sg    very    religious      SE    merchant    become-PST-3SG
- Jumala-a    rukoo-le-maa-se    eb-ko              Jumala    awitta-isi  
 god-PART    pray-SUP-ILL      NEG-3SG-INT    god      help-COND
- Wäll'ää      siitä              paikka-a.  
 away        from.there      place-ILL

‘The merchant was very religious, so he started to pray to God to help him to leave and to get to the place.’

The noun *kuptsa* ‘merchant’ is preceded with *se*, since it is accessible from the previous context. It occurs several times in the text below, but only once is it prefaced with *se*. The merchant is not a prominent participant, but this is one of the crucial points in the narrative. This is why the NP is marked with *se*. The word order in this sentence is emphatic (VS), which can also be a sign of a crucial point.

God had not helped him, so he called for help from the Devil. The Devil promised to help him but asked for his help something that the merchant did not know existing at his home.

At home the merchant met a child, which he had promised to the Devil. Then this child, a young guy left home and his adventures began.

5 Alqvist’s orthography is saved here with the exception of the marking of long vowels: in the original text they are marked with the symbol ‘^’, but in the article they are indicated with digraphs for the sake of simplicity of glossing because in some cases morpheme borders go along long vowels, for example *Jumala-a* ‘god-PART’. The symbol <’> is used by Alqvist not only to mark some sounds missing in spontaneous speech (for example ‘-b’ NEG-3SG = *e-b* ‘NEG-3SG), but also to mark some elements which have disappeared from the language (for example genitive marker -\*n: *kuptsaa-’emännää-’poika* ‘merchant-GEN wife-GEN son’.

The guy met a very strange old lady, who told him that he had been promised to her brother and advised him how to behave with her brother. The following day he met another old lady, who looked like the first one and who told him the same things.

2. Kõlmattõma-la päiwä-lä siis tämä jõutu-si-ø ööse-hsi  
 third-AD day-AD then he get-PST-3SG night-TRNSL  
 sinne atamani-le ene-le-sä.  
 there devil-AL himself-AL-PX3

‘The third day by the night he got to the Devil himself.’

This is the next crucial point of the narrative, and that is why *sinne* is used. It is interesting that although the word *ataman* ‘Devil’ appears in the tale for the first time, it is preceded with the adverb *sinne*. The referent was already mentioned before as a brother of two old ladies, so it is accessible from the previous context. After that, *ataman* appears several times in the narrative, but it is no longer prefaced with *se*.

The Devil told the guy that he would like to eat him, but he would not yet eat him if he cut down seven versts of forest, ploughed them, sowed wheat, cropped it, thrashed it and baked a cake before lunch. The Devil has 12 daughters, the youngest of whom promised to help the guy if he promised to marry her. He promised.

3. Maga-te-za pojo se kõikkiaa-’ noore-pi tütär  
 sleep-INF-IN guy SE all-GEN young-CMPR daughter  
 salamii’ izä-ssä lähätt-i-ø sõzare-t oma-sa töö-le  
 secretly father-EL send-IPF-3SG sister-PL own-PX work-AL

‘While the guy was sleeping, the youngest Devil’s daughter without father’s knowledge sent her sisters to work.’

As we can see, *kõikkiaa’ noorepi tütär* ‘youngest daughter’ is preceded with *se*. The referent appeared for the second time in the narrative and is accessible from the previous context. In all *kõikkiaa’ noorepi tütär* appears 16 times and is prefaced with *se* every time with the exception of when first mentioned. As I will show below this referent appears in the narrative under other names, but it is always marked with *se*. The Devil’s youngest daughter is a central character in the story, and she, and not the guy, is the protagonist. As for the guy, he is the supernumerary,

although it would seem, that the story is about him. This referent is expressed with *lahs* ‘child’, *kuptsaa’emännää’poika* ‘merchant’s wife’s son’ and *pojo* ‘guy’ and never occurs with *se* in the article-like usage.

The story continues. The sisters have done the work. After that the Devil assigned some more supernatural tasks to the guy, and in all the cases the youngest daughter sent her sisters to accomplish them. Then the guy and the youngest daughter of the Devil got married and left the Devil’s house.

4. Siis juttel-i-wat ned tüttäre-t što se pojo ’-b  
 then say-PST-3PL SE.PL daughter-pl that SE guy NEG-3sg  
 mitäi-tä teh-nü mö kōikki te-i-mmä mi-tä čäss-i-t.  
 nothing-PART do-PTCP we all do-PST-1PL what-PART order-PST-2SG  
 ‘Then the daughters said that the guy had done nothing, “We have done all you had ordered.”’

Though *tüttäret* ‘daughters’ were already mentioned several times (this referent also occurred under the name of *sõzaret* ‘sisters’), however it is the first time that it is preceded with *ne* (plural of *se*). This is of course one of the most crucial points of the narrative, because the guy would be near collapse.

As we see, *pojo* ‘guy’ is also prefaced with *se*, but *se* is not in its article-like usage. This phrase has a typical Votic syntax: after the conjunction *što* ‘that’ (borrowed from Russian) instead of the indirect speech which would be expected, the direct speech is used. So *se* which prefaces *pojo* is used as exophora here.

So the Devil ordered his daughters to go and bring back their sister with her husband. The Devil’s youngest daughter guessed it.

5. No se noorikko teč-i-ø ihse ene-sä  
 so SE bride make-PST-3SG herself herself.GEN-PX3  
 owos-tabuni-hsi i oma-sa mehee’ tabušnika-hsi.  
 horse-herd-TRNSL and own-PX3 husband-GEN horse.herder-TRNSL  
 ‘So the bride turned herself in a herd of horses, and turned her husband in a horse herder.’

The central character of the story appears under the name *noorikko* ‘bride’ and it is also preceded by *se*.

6. Tüttääre-t ned takaa-ajaja-t tul-i-wat tabus'nikaa-'  
 daughter-PL SE.PL back-pursuer-PL come-PST-3PL horse.herder-GEN  
 tüwe çüsü-si-wät e-t-ko näh-nü matkalais-i-a mees-sä  
 near ask-PST-3PL NEG-2SG-INT see-PTCP traveler-PL-PART man-PART  
 da nais-sa mene-wää tee-tä möö?  
 and woman-PART go-PTCP way-PART along

‘The daughters, the pursuers, came to the horse herder and asked: “Have you seen travellers, a man and a woman going along the way?”’

NP ‘daughters’ is given and it is referred to twice within the same sentence, first with *tüttääret* ‘daughters’ and then with the right dislocated *ned taka-ajajat* ‘the pursuers’. The daughters are not highly prominent in this point of the narrative, so why is this NP marked with *se*? Let us investigate further.

The guy (= the horse herder) answered that he was tending a herd for a long time but has never seen any travellers on the way.

7. No nämä çääntü-si-wät tagaas kotoo-se i juttel-i-wat ädä-lee  
 so they turn-PST-3PL back home-ILL and say-PST-3PL father-AL  
 si-le alamani-le e-mmä näh-nü muuta mitäi-tä kui' waitas  
 SE-AL devil-AL NEG-1PL see-PTCP just nothing-PART except only  
 tabuni opos-i-a da tabusnikka men-i-ø tee-tä möö.  
 herd horse-PL-PART and hors.herder GO-PST-3SG way-PART along

‘So they returned home and said to their father, the Devil: “We haven’t seen anybody at all, except a horse herd and a horse herder going along the way.”’

The Devil is given and is also referred to twice within the same sentence as the daughters were in the previous example. The Devil is not highly noteworthy at this point of the narrative either.

Then the Devil explained to his daughters that they just had met their sister and her husband, and sent them back again. The youngest Devil’s daughter saw that her sisters had returned and she turned herself in a herd of cows and her husband into a cow herder.

8. Tullas ned sōzare-t čūsü-si-wät nä-i-t-ko kassenna  
 come.IPS SE.PL sister-PL ask-PST-3PL see-PST-2SG-INT here  
 owos-tabuni-a mokoma-a?  
 horse-herd-PART such-PART

‘The sisters came and asked: “Have you seen such a herd of horses here?”’

The sisters are not highly noteworthy at this point of the story; however the NP is prefaced with *ned* in the article-like usage. In addition there is an emphatic word order: VS, just as in example 1.

Not only are the NPs whose referents are central characters in the story preceded with *se*, but also the clustering of *se*-marked NPs occurs at a crucial point of narrative, as we have seen in example 1. The last part of the story after the Devil’s youngest daughter and her husband left house is very dramatic and crucial, so this is why the narrator uses *se* in the article-like usage so often here.

The reason for the article-like usage of *se* in Votic narrative in the 19th century seems to be the same as in Finnish narratives in the 19th century, as we can judge by Laury (1991).

### 3.2. The modern Votic data analysis

Let us now examine modern Votic narratives<sup>6</sup>.

9. enne niku se perä sōa mei-l kassin kõig čülve-tt-i  
 formerly like SE after war.GEN we-AD here all sow-IPS-PST  
 ville-d õzra-a kagra-a rüis-sä.  
 cereal-NOM.PL barley-PART oats-PART rye-PART

‘Formerly, after (the Second World) War, all the crops were sowed here: barley, oats, rye.’

6 This text is published in (Agranat 2012: 24–27); I wrote it down in 2001 in Ust-Luga, the speaker was T. F. Prokopenko (1935–2006). I present my examples here in the phonological transcription, using IPA symbols with the exception of ‘õ’ to mark the middle vowel and of ‘л’ to mark the hard lateral consonant (for palatal ‘l’ is used).



10. siiz ku kõig vaļmissu-b niite-tt-i pan-t-i vihko-a  
 then when all ripen-PRS.3SG reap-IPS-PST put-IPS-PST sheaf-PART  
 vihgo-d teh-t-i.  
 sheaf-NOM.PL make-IPS-PST

‘Then, when they are completely ripened, a sheaf was put, sheaves were made.’

11. siiz õpõzi-j-õ ka ne kõig vihgo-d veite-tt-i  
 then horse-PL-GEN with SE.PL all sheaf-NOM.PL carry-IPS-PST  
 riiga-a.  
 threshing.barn-ILL

‘Then the sheaves were carried to the threshing barn with horses.’

In this sentence *vihgod* ‘sheaves’ is accessible from the previous context, but it is not prominent; however, it is prefaced with *ne*.

12. riiga-a üvässi lämite-tt-i štoby ne kõig vihgo-d  
 threshin.barn-GEN well heat-IPS-PST for SE.PL all sheaf-NOM.PL  
 üvässi kuiva-st-i.  
 well dry-IPS-PST

‘The threshing barn was heated well for all the sheaves to dry well.’

In this example *vihgod* ‘sheaves’ is also preceded by *ne* for the same reason as in the previous example. But although *riiga* ‘threshing barn’ was mentioned above, it is not prefaced with *se*.

13. siiz pan-t-i ne kõig vihgo-d maa-ļļ.  
 then put-IPS-PST SE.PL all sheaf-NOM.PL floor-AD

‘Then all the sheaves were put on the floor.’

As we can see, *vihgod* ‘sheaves’ is again preceded with *ne*; it is accessible but not noteworthy either.

14. siiz nõ-st-i mokomaizõ-d primuzļõ-d.  
 then become-IPS-PST such-NOM.PL flail-NOM.PL

‘Then there were such flails.’

15. siiz neje primuzlõ-j-õ ka kõik raimi-tt-i  
 then SE.PL.GEN flail-PL-GEN with all thresh-IPS-PST  
 ned vihgo-d.  
 SE.PL sheaf-NOM.PL

‘Then the sheaves were threshed with the flails.’

Since *primuzlõd* ‘flails’ and *vihgod* ‘sheaves’ were already mentioned, *ned* is used.

16. siiz kõik ville-d tokku-st-i poiz.  
 then all cereal-NOM.PL fall-IPS-PST away

‘Then all the crops fell.’

17. siiz õl-t-i mokomaizõ-d veejalka-d niku sell  
 then be-IPS-PST such-NOM.PL fan-NOM.PL like there  
 tuulõõtõ-tt-i si-tä villa-a.  
 winnow-IPS-PST SE-PART cereal-PART

‘Then there were such fans to winnow the crops.’

As *villa* ‘crop’ is accessible from the previous sentence, it is used with *se*.

So in present-day Votic the accessibility of a referent may be the only reason to use *se* and noteworthiness is not obligatory. It is just as in spoken Finnish, according to Laury (1991), after *se* became the identifiability marker, it lost its function of marking the prominence of NPs. Votic speakers might mark the identifiability of a referent with *se*, but they are not obliged to do so. That is why, for example, the *riiga* ‘threshing barn’ (example 12) is used without *se* although it had been mentioned in the previous sentence.

Greenberg (1978: 61) has described the diachronic stages of development that definite articles typically go through in a wide variety of languages. The early stages (stage I by Greenberg) may be characterised as follows: “The point at which a discourse deictic becomes a definite article is where it becomes compulsory and has spread to the point at which it means ‘identified’ in general, thus including typically things known from context, general knowledge, or <...> identified because it is the only member of its class”.

As for modern Votic, *se* in the article-like usage possesses almost all the features presented above, except, possibly, the main one: it is not compulsory. So it is on the pathway to grammaticalisation, but it has not yet reached Greenberg's Stage I of the development of definite articles from demonstrative pronouns.

#### 4. Conclusion

The reasons for the article-like usage of *se* in Votic narrative in the 19th century are accessibility from previous discourse context and noteworthiness of NPs which are prefaced with *se*, given NPs are also marked with *se*, but not as often as accessible NPs. In addition NPs which are preceded with *se* have references which are important in some way for the narrative, so an NP which is prefaced with *se* is typically an NP whose referent is a central character in the story or otherwise a crucial prop. Thus *se*-marked NPs also cluster in foregrounded clauses.

In modern Votic *se* lost its function of marking NPs as prominent and became the marker of identifiability. It possesses almost all the features of the definite article (according to Greenberg 1978) with the exception of the main one: it is not compulsory. So it is on the pathway to grammaticalisation, but has not yet reached even Greenberg's Stage I of the development of definite articles from demonstrative pronouns.

I think it will be useful to study data of other closely-related languages.

#### Address:

Tatiana Agranat  
 Institute of Linguistics of Russian Academy of Sciences  
 1 bld. 1 Bolshoi Kislovsky lane  
 125009 Moscow, Russian Federation  
 E-mail: tagranat@yandex.ru

#### Abbreviations

1 – 1st person, 2 – 2nd person, 3 – 3rd person, AD – adessive, AL – allative, CMPR – comparative, COND – conditional, EL – elative, GEN – genitive, ILL – illative, IN – inessive, INF – infinitive, INT – interrogative, IPF – imperfect, IPS – impersonal, NEG – negation, NOM –

nominative, NP – nominal phrase, PART – partitive, PRS – the present tense, PL – plural, PST – past tense, PTCP – participle, PX – possessive, SG – singular, SUP – supine, TRNSL – translative.

**Languages and dialects:** Est. – Estonian, Fin. – Finnish.

## References

- Agranat, Tat'jana B. (2008) "Izmenenie sistemy prostranstvennogo dejksisa v vods-kom jazyke". In *Congressus X Internationalis Fenno-Ugristarum. Joškar-Ola 15.–21.08.2005*. Pars IV. Linguistica, 179–184. Joškar-Ola.
- Agranat, Tat'jana B. (2012) *Vodskie teksty s pomorfemnoj notaciej*. Moskva: Tezaurus.
- Ahlqvist, August (1856) *Wotisk grammatik jemte språkprof och ordförteckning*. (Acta Societatis Scientiarum Fennicae, 5.) Helsingfors.
- Chafe, Wallace (1987) "Cognitive constraints on information flow". In Russell S. Tomlin, ed. *Coherence and grounding in discourse: outcome of a symposium, Eugene, Oregon, June 1984*, 21–51. Amsterdam: Benjamins.
- Chesterman, Andrew (1991) *On definiteness: a study with special reference to English and Finnish*. (Cambridge Studies in Linguistics, 56.) Cambridge, New York: Cambridge University Press.
- Greenberg, Joseph H. (1978) "How does a language acquire gender markers". In Joseph H. Greenberg, Charles A. Ferguson, Edith A. Moravcsik, eds. *Universals of human language 3. Word structure*, 47–82. Stanford: Stanford University Press.
- Kibrik, Andrej A. (2011) *Reference in discourse*. Oxford: Oxford University Press.
- Laury, Ritva (1991) "On the development of the definite article *se* in spoken Finnish". In *SKY 1991. Yearbook of the Linguistic Association of Finland*, 93–121. Helsinki: Suomen kielitieteellinen yhdistys.
- Laury, Ritva (1996) "Sen kategoriasta – onko suomessa jo artikkeli?" *Virittäjä* 2, 100, 162–181.
- Laury, Ritva and Renate Pajusalu (2010) "The marking of definiteness in old written Estonian and Finnish: native or borrowed?" In *Congressus XI Internationalis Fenno-Ugristarum, Piliscsaba 9–14. VIII*. Pars III, 25–26. Piliscsaba.
- Metslang, Helle (2011) "Some grammatical innovations in the development of Estonian and Finnish: forced grammaticalization". *Linguistica Uralica* 47, 4, 241–256.
- Pajusalu, Renate (1997) "Is there an article in (spoken) Estonian?" In Mati Ereht, ed. *Estonian typological studies 2*, 147–175. Tartu: Tartu Ülikooli eesti keele õppetool.

**Kokkuvõte. Tatiana Agranat: Määrav artikkel vadja keeles: grammatiseerumise protsess.** Käesoleva artikli fookuses on vadja keele näitava asesõna *se* grammatiseerumine määravaks artiklikliks. Autor jälgib muutumisprotsessi diakrooniliselt, esitades oma uurimistulemused August Ahlqvisti poolt 19. sajandi keskel üles kirjutatud rahvajutu ning enda poolt 21. sajandi alguses

üles kirjutatud teksti näitel. Viidates 19. sajandi soome narratiivide asesõnade kasutust uurinud Laury'le (1991) järeldeb käesoleva artikli autor, et asesõna *se* artikli-sarnases funktsioonis kasutamise põhjus 19. sajandi vadja ja soome keeles on sama: asesõnaga *se* algavad nominaalfraasid rõhutavad narratiivis olulisi viitesuhteid. Kaasaegses vadja keeles on *se* kaotanud nominaalfraasi prominentsust tähistava funktsiooni ning muutunud identifitseeritavuse markeriks. Autor järeldeb, et asesõna *se* grammatiseerumisprotsess vadja keeles ei ole veel lõppenud, sest *se* kasutamine nominaalfraasis ei ole kohustuslik.

**Märksõnad:** vadja keel, määrav artikkel, grammatiseerumine, näitav asesõna