Petar Kehayov

Taboo intensifiers as polarity items: Evidence from Estonian[1]

Abstract

This paper discusses the development of expressions such as "who/God knows WH" into negative degree adverbs in Estonian. The study provides an answer to the question of *how* certain negative polarity items emerge in a language. This question is rarely answered in the literature focused on polarity; hence, the case-study at hand is an attempt to contribute to the general understanding of the grammaticalization/lexicalization of polarity items. Two issues are studied in detail: first, the pragmatic prerequisites, and second, the structural idiosyncrasies triggering the lexicalization of these types of expressions into NPIs. The insights obtained allow challenging the view (see von Bergen & von Bergen 1993: 130) according to which the historical change in the meaning of the linguistic elements is not relevant for the synchronic description of polarity items.

Keywords: taboo intensifiers, negative degree adverbs, (non)veridicality, lexicalization, grammaticalization

1. Introduction

Consider the expressions in italics in the following examples:

(1) He thinks he is *I don't know how* smart.
(2) He thinks he is *God knows how* smart.

In his work on indefinite pronouns, Haspelmath (1997: 130–133) notes that in many European languages the type of expression in (1) has served as a source for indefiniteness markers; witness, for instance, French *je ne sais quel* 'some kind of'. Haspelmath also briefly comments on examples such as (2), and adds that he knows of "no case where such a source construction has been grammaticalized strongly" (ibid.: 131) into a marker of indefiniteness.

This study examines expressions of the second type, henceforth referred to as TABOO INTENSIFIERS. Taboo intensifiers pick up an agent from certain lexical domains denoting entities that are perceived as inviolable by the speech community. The most common domains of taboo besides Gods are demons, "obscene" body parts and diseases. Consider the following examples from Polish:

(3) a. *On myśl-i, że jest diabl-i wiedz-ą jak mądry.*
he think-3SG that be.3SG devil-PL know-3PL how smart
'He thinks he is devils know how smart.' (Robert Bielecki, p.c.)

b. *On myśl-i, że jest chuj wie jak mądry.*
he think-3SG that be.3SG dick know.3SG how smart
'He thinks he is so smart (lit. 'He thinks he is dick knows how smart'). (Robert Bielecki, p.c.)

c. *On myśl-i, że jest cholera wie jak mądry.*
he think-3SG that be.3SG cholera know.3SG how smart
'He thinks he is so smart (lit. 'He thinks he is cholera knows how smart'). (Robert Bielecki, p.c.)

[1] The research reported in this paper was sponsored by the Estonian Science Foundation (grant no. 5202: Typological Grammar of Estonian (Syntax)). I would also like to thank Vahur Aabrams, Rogier Blokland, Mati Erelt, Ērika Krautmane, Johanna Laakso, Liina Lindström, Helle Metslang, Matti Miestamo, Renate Pajusalu, Haldur Õim, Johan van der Auwera and Teresa Fanego for valuable comments and suggestions on previous drafts of the paper.

In Estonian, as well as in the closely related Finnish, taboo intensifiers of this kind have undergone an interesting change. Unlike their equivalents in better known European languages, Estonian and Finnish intensifiers show a strong tendency to drop the taboo agent. This tendency can be ascribed to a lexicalization process whereby the collocation [TABOO AGENT *know*-3SG WH] can no longer be accessed analytically and becomes more context dependent. A variant of this collocation, namely [Ø *know*-3SG WH AdjP/AdvP], is then lexicalized into a degree adverb which in Estonian, interestingly enough, is sensitive to POLARITY. Compare (4a) and (4b):

(4) a. *Lapimaa on jumal tea-b kui kaugel.*
Lapland be.3SG God know-3SG how far
'Lapland is God knows how far.'

b. *Rootsi pole Ø tea-b kui kaugel.*
Sweden NEG.be Ø know-3SG how far
'Sweden is not very far.' (lit. 'Sweden is not knows how far.')

In other words, whereas the construction with an overt taboo agent (4a) is mostly found in positive sentences, the reflexes with a zero agent (4b) are licensed by the presence of a negative elsewhere in the sentence.

2. The scope of the study

The linguistic evidence suggests that the agentless construction in (4b) is a direct descendant of the construction with an agent in (4a). The aim of the present study is to identify the factors accounting for the lexicalization of a complex construction with a taboo agent into negative degree adverb. On the assumption that GRAMMATICALIZATION and LEXICALIZATION are not mirror images, but somewhat complementary processes (see, among others, Moreno Cabrera 1998, Lehmann 2002, Rostila 2006), I will take a rather liberal attitude towards the notorious problem of telling grammaticalization and lexicalization apart (on this issue see in particular Brinton & Traugott 2005: 62 ff). As regards developments involving intensifiers, the decision adopted depends, to a very large extent, on the kind of item standing at the endpoint of the cline. Haspelmath (1997: 130–133), for instance, is concerned with the development of expressions such as that in (1) into indefinite pronouns, and therefore speaks of grammaticalization. In the case under discussion, however, the outcome of the process is adverbs (i.e. arguably lexical words), and for this reason I will assume here that this is a case of lexicalization.

Discovering the factors behind the change in polarity is possibly the most challenging task of this research. There is widespread agreement in the literature on polarity that negative polarity items cluster cross-linguistically in certain semantic domains, but there are very few studies concerned with their grammaticalization and/or lexicalization,[2] apart from those carried out by researchers based at the University of Groningen (e.g. Hoeksema 1994, Hoeksema & Klein 1995, 1996). Two reasons seem to account for this lack of attention to the grammaticalization of polarity items. First, due to the lack of sufficient historical evidence, it is often impossible to trace down shifts in polarity over time. Second, grammaticalization research views language development as an ongoing process, and is thus interested in studying the relationships between the various components of a language both diachronically and synchronically, but research on polarity sensitivity over the past forty years has in fact been dominated by the generative paradigm, which does not generally recognize grammaticalization as a distinct process (see, for instance, Newmeyer 1998) and tends to focus solely on the synchronic axis. It is a matter of theory commitment, therefore, that most research on negative polarity has looked only at the synchronic aspects of polarity items, and has sometimes explicitly pronounced their historical development to be irrelevant for their synchronic description (see von Bergen & von Bergen 1993: 130).

The paper is organized as follows: Section 3 gives an overview of the corpora employed, and offers some preliminary figures on the data retrieved from them. Section 4 provides a brief account of

[2] The grammaticalization of pure negators, by contrast, has attracted much more attention.

the notion of polarity item. Section 5, which is descriptive, examines in detail the distribution of the construction under analysis. Section 6 is explanatory: the first subsection investigates the historical connection between the construction with taboo agent and the construction with zero agent, and examines the degree to which these two constructions are lexicalized. In the second subsection, I will dwell upon the puzzling aspects of the shift in polarity. We will take a closer look at the pragmatics and semantics of taboo intensifiers and at their interplay with the structure of Estonian. In the end of the second subsection, I will discuss a specific syntactic reanalysis, which led to an increase of the polarity sensitivity of a certain form. In Section 7, I will try to offer an areal-typological perspective by briefly discussing the development of similar items in Finnish and Latvian. Section 8 summarizes the results of the study.

3. Survey of the corpus

The data were collected from three electronic corpora, namely the Corpus of Estonian Literary Language (CELL),[3] the Grammatical Corpus of Contemporary Estonian (GCCE),[4] and the Text Corpus of the Institute of the Estonian Language (TCIEL).[5] In addition, the examples from the entry for *teadma* 'to know' in the *Dictionary of Literary Estonian* (EKSS) were also included.

CELL contains about five million words and is divided into a number of subcorpora organized by decades, the first dating back to the 1890s. The subcorpora covering the 1980s and the 1990s are considerably larger than the others: they comprise, respectively, 22.5 and 22.4 per cent of the total corpus. All subcorpora, excluding the 1980s, contain fiction and journalistic texts, which are equally shared out. In addition to fiction and journalistic texts, the subcorpus from the 1980s contains essays, biography, and science. GCCE totals seven million words composed of contemporary fiction (two million words) and journalistic texts (five million words) collected from four different newspapers covering the period 1990-2002. Finally, the size of TCIEL is 10.4 million words; it comprises newspapers, magazines and fiction, as well as some recordings of spoken language (60.000 words). The newspaper texts are all from the 1990s and make up 86.5 per cent of the corpus.

The total size of the corpora examined for this study is thus 17.4 million words. However, some of the journalistic and fiction texts occur in more than one corpus; once they were excluded, the size of the corpus was reduced to 16.4 million words. The three corpora are not parallel in structure, and hence their overall representativeness is doubtful. Unfortunately, these were the only online resources available for Estonian at the time the search was conducted (September 2004).

Regarding the concordancing procedure, consider once again examples (4a)-(4b), repeated here for ease of reference:

(5) a. *Lapimaa on jumal tea-b kui kaugel.*
Lapland be.3SG God know-3SG how far
'Lapland is God knows how far.'

b. *Rootsi pole Ø tea-b kui kaugel.*
Sweden NEG.be Ø know-3SG how far
'Sweden is not very far.'

I used four different search strings, each containing the third person singular present tense form (*teab*) of the verb *teadma* 'to know' and one of the four possible sequences of two phonemes (*mi*, *ku*, *ke* and *ka*) occurring in the beginning of the Estonian question words. In all cases, the slot for the taboo agent was left unspecified, so that the exhaustiveness of the search was ensured. Since I was looking for collocations that had already begun to lose their compositionality, the search was limited to collocations not containing an interclausal comma, as free collocations will often require comma punctuation after *teab* 'knows' to mark the clause boundary. After the search was completed, I added another 14 instances from the EKSS; this reference work contains only attested examples and is thus fully reliable as a resource.

[3] <http://www.cl.ut.ee/ee/corpusb/tykk.html>
[4] <http://test.cl.ut.ee/corpora/grammatika/index.et.html>
[5] <http://www.eki.ee/corpus/>

Overall I collected 405 occurrences of the collocation [TABOO AGENT/ZERO AGENT *teab* WH], out of which 201 have an overt taboo agent and 204 a zero agent. The following agents (in order of decreasing frequency; see Appendix A) were recorded in the corpus: *jumal* 'God', *kes* 'who', *kurat* 'Devil', *tont* 'ghost; bogey', *pagan* 'pagan', *taevas* 'heaven', *issand* '(the) Lord', *põrgu* 'hell', *pergel* 'devil, pagan' and *juudas* 'Judas'. According to the *Phraseological Dictionary of Estonian* (Õim 2000: 493), the lexemes *tühi* 'empty' and *mõni* 'some' are also possible in the agent slot of this construction, but they did not occur in my material.

It can be seen that, unlike Polish (3b), (3c), Estonian does not use the names of body parts or diseases as taboo agents, but only "extreme" individuals (either sacred or demonic) or their substitutes, such as the pronouns *kes* 'who' and *mõni* 'some', and the adjective *tühi* 'empty'. Although I am going to discuss the semantic issues more thoroughly in section 6, it must be said at this point that the use of *kes* 'who', *mõni* 'some' and *tühi* 'empty' in this expression is not surprising, since these items elicit here the same implicature as the taboo agents. This means that sequences like *kes teab kus* 'who knows where' and *mõni teab kus* 'some(body) knows where' are not assigned here their literal meaning, but rather state that 'nobody (= none of the humans) knows where'. The universal reading of the implicature is directly related to the meaning of taboo, and more specifically to an "extreme" individual, either sacred or demonic. I therefore assume that these items function here as euphemistic substitutes of an underlying taboo agent. This is most obvious in the case of *tühi* 'empty', which can be considered a shortened variant of *vanatühi* 'Devil, Old Nick' (lit. 'old empty').

Estonian is a language rich in nominal inflections. Appendix B shows the items occurring after the verb form *teab* 'knows' in the construction under concern. All the items listed in the table except for the last one (*kas*), which is a *yes*/*no*-word function as *wh*-words. The lines in the table indicate lexeme borders; that is, if no line divides two items in the table, they are inflectional forms of a single word. It can be seen that the nominative form *mis* of the word 'what' and the word *kui* 'how' (in bold) have a much higher frequency than the rest of the *wh*-words, making up more than half of the examples. I will therefore discuss the collocations [TABOO AGENT *teab mis*] and [TABOO AGENT *teab kui*] in separate sections, and the remaining collocations together.

Although this research makes use of corpora, it should be stressed that it is not corpus-based, in the sense that the data retrieved from the corpora are too limited to allow us to elucidate the process of lexicalization of taboo intensifiers in all its complexity. This granted, I nevertheless believe that the material collected suffices to demonstrate the correlation between certain *wh*-constructions and negative polarity. To check this, two tests of statistical significance have been applied in the descriptive part of the paper (Section 5), namely, the chi-square test,[6] and, where the number of instances is less than five, the Fisher Exact Probability Test. Occasionally, I have supplemented the attested data with examples invented by myself, and checked these examples for grammaticality with native speakers.

4. The notion of polarity item

As defined by Giannakidou (2001: 99), POLARITY ITEMS "form a more or less homogeneous class of expressions whose distribution is restricted by conditions which must appeal to some kind of polarization (negation or affirmation) for wellformedness". They can be classified into POSITIVE POLARITY ITEMS (PPIs; e.g. Eng. *somewhat*, as in "John liked it somewhat") and NEGATIVE POLARITY ITEMS (NPIs; e.g. some uses of Eng. *any*, as in "John didn't buy any books").[7]

Most of the research devoted to polarity consists of case studies on individual polarity items, such as the English indefinite pronoun *any* (Kadmon & Landman 1993), but there is also a considerable number of analyses concerned with the typology of PIs (e.g. Ladusaw 1979, Krifka 1995, Zwarts 1995, Israel 1996, van der Wouden 1997). Overall, one of the central issues is the question of what kinds of properties trigger (or license) PIs. Thus, a semantic line of research starting with Ladusaw

[6] To improve the accuracy of the chi-square test in the case of 2x2 tables I have also applied Yates' correction for continuity, which reduces the magnitude of the difference between expected and observed frequencies by 0.5.

[7] In addition, the literature on polarity also recognizes FREE CHOICE ITEMS, that is, items of restricted distribution but not subject to a polarity dependency, as is the case of Eng. *any* in interrogatives ("Did you find any interesting books?") or imperatives ("Pick any apple!").

(1979) argues that PI licensing is a semantic, rather than a syntactic, relation. This view contrasts with a syntactic tradition in the study of polarity which goes back to Klima (1964; see also Progovac 1994) and looks for the exact structural (syntactic) relationship between negation and polarity items. Finally, there is also a pragmatic tradition in the description of polarity phenomena, represented by Lakoff (1969), Linebarger (1981) and Israel (1996).

In the present study I adopt a semantic approach based on the notion of (NON)VERIDICALITY (see Zwarts 1995, Giannakidou 1998). (Non)veridicality is linked to the availability of a truth entailment: roughly, a SEMANTIC OPERATOR (*Op*) is NONVERIDICAL if it does not entail the truth of the proposition it embeds. Adverbs like *possibly* or *perhaps* (e.g. "Perhaps Roxanne has arrived") and modals (e.g. "Paul may hit Frank") are typical nonveridical operators. Other nonveridical environments include negation, nonassertive speech acts (questions, imperatives, exclamatives), the protasis of conditionals, the scope of strong INTENSIONAL verbs such as *want* or *hope* (e.g. "I hope there is a piece left"), or the future. By contrast, VERIDICAL operators entail the truth of their complements. The adverbs *yesterday* ("Yesterday, Paul saw a snake") and *actually* ("Paul actually hit Frank") and the assertion operator ("Paul saw a snake") are typical examples of veridical operators. Finally, ANTIVERIDICAL (or AVERIDICAL; see Giannakidou 1998: 106, footnote) operators form a subset of the nonveridical; antiveridical operators entail the falsity of the proposition they embed. Direct (or clausemate) negation ("John didn't meet anyone") is the prototypical antiveridical operator.

As a general rule, NPIs are licensed in nonveridical contexts (cf. *any* in "Did you see any people?" and in "*I saw any people."). There are also a small number of "hyperstrong" NPIs, such as Eng. *lift a finger* ("He didn't lift a finger") or n-words in Serbo-Croatian (cf. Giannakidou 1998: 162) that are licensed solely by negation, and hence by antiveridicality.

In sections 5.1 and 5.2 below we will see that the (non)veridicality hypothesis virtually captures the licensing conditions for taboo intensifiers as NPIs.

5. The distribution of Estonian taboo intensifiers

5.1. The type [TABOO AGENT / Ø *teab mis*]

Table 1 shows that the collocation [TABOO AGENT *teab mis*] is restricted almost exclusively to positive sentences, whereas the collocation with a zero agent is sensitive to direct negation.

Table 1. Polarity distribution of *teab mis*[8]

polarity	with a taboo agent	with a zero agent
positive	60	25
negative[9]	4	76

As illustrated in (6), Estonian *mis* 'what' behaves like English *what* in that it has both pronominal (6a) and modifier (6b) uses. Their distribution with respect to polarity is given in Table 2.

(6) a. *Homne päev võis tuua kurat teab mis.*
tomorrow's day could bring Devil knows what
'Tomorrow could bring the Devil knows what.' (i.e. 'Tomorrow could bring anything.')
(CELL: ILU1980\stkt0112)

b. *Homne päev võis tuua kurat teab mis jama.*
tomorrow's day could bring Devil knows what hitch.PRTV
'Tomorrow could bring the Devil knows what problems.' (i.e. 'Tomorrow could bring all sorts of problems.')

Table 2. Polarity distinctions according to the syntactic function of *mis*

	with a taboo agent	with a zero agent

[8] The distribution is significant (chi-square = 71.9326, p = .000).
[9] Here and in the rest of the tables the label "negative" refers to explicit negation.

polarity	pronoun	modifier	pronoun	modifier
positive	29	31	8	17
negative	–	4	1	75

The figures[10] in Table 2 suggest that zero-agent collocations with pronominal *mis* (henceforth mis_{Pron}) correlate with positive polarity, whereas zero-agent collocations with *mis* as a modifier (henceforth mis_{Mod}) are sensitive to negation: 75 out of a total of 92 examples of this type are found in negative sentences.

Also worthy of mention in connection with the data in Table 2 is that about two thirds of the 17 instances of the collocation [Ø *teab* mis_{Mod}] occurring in positive sentences are used in nonveridical contexts, such as irreal clauses marked by conditional mood (7a), comparative clauses (7b), or concessives (7c):

(7) a. *Kingituse eest tuli tänada ja see pisike must mees oli seda nägu, nagu ole-ks*
gift for came thank and this tiny black man was this face as if be-COND
ta and-nud ***teab mis kalli asja,***
he give-PART knows what precious.PRTV thing.PRTV
'S/he had to thank him for the gift and this tiny black man had an expression on his face as if he had given away something very precious,' (TCIEL: Fiction)

b. *Kui mõni Neitsi jätabki nohiku mulje, on niisugustest rohkem kasu kui*
if some Virgo leaves nerd impression is such more profit than
teab mis jutupauniku-te-st
knows what long-winded_person-PL-ELAT
'Although some Virgos leave an impression of being nerdy, they are more helpful than any kind of long-winded people.' (CELL: AJAE1990\ee0347)

c. *Lennaku õhus* ***teab mis vidina-d****, ei jää elu sinu jaoks koos*
fly-JUSS air knows what thingamajig-PL NEG stay life you for with
maakeraga seisma, pigem pöörleb koos sinu uute mõtetega edasi.
earth stay rather spin with your new thoughts forwards
'Whatever kinds of thingamajigs are flying in the air, life and the earth won't stand still for you, rather they will keep spinning round with your new ideas.' (GCCE: EE1999)

As already noted, the collocations with an overt taboo agent, and also those with a zero agent and mis_{Pron}, occur predominantly in positive sentences. This does not mean, however, that they are restricted to veridical contexts. Thus, three out of the eight positive examples with a zero agent and mis_{Pron} were used in nonveridical contexts, and of the 60 instances with an overt taboo agent in positive sentences, 22 were found in nonveridical contexts too, such as the scope of modal verbs (8a) or conditional sentences (8b):

(8) a. *eile nimetasid mind nõiaks, täna kaardimooriks, homme peab mu*
yesterday called me witch today fortune-teller tomorrow must my
nimi öökull ehk ***kes teab mis muu-d veel*** *ole-ma…*
name owl or who knows what else-PRTV more be-SUP
'yesterday you called me a witch, today a fortune-teller, and tomorrow my name maybe will be owl or who knows what else…' (TCIEL: Fiction)

b. *Kui ma weel kauem selle wastu seisa-ksi-n, tee-ksi-wad nad*
If I still longer this against stay-COND-1SG do-COND-3PL they
minu-ga ***Jumal teab mis****.*

[10] The chi-square test gives a significant distribution (chi-square = 91.7710, p = .000). However, since two of the cells are below five, the test is not completely reliable. The application of the Fisher Exact Probability Test to two 2x2 tables, one for the collocations with an overt taboo agent and another one for the collocations with a zero agent, gives a nonsignificant distribution for the left table (p = .120) and significant for the right table (p = .000).

me-COM God knows what
'If I keep confronting this, God knows what they will do to me.' (CELL: ILU1910\ilu0008)

This confirms, therefore, that all collocational types may occur in nonveridical contexts. However, it is clear that the type [Ø *teab mis*$_{Mod}$] correlates with nonveridical contexts and with negation much more frequently than any of the others.

As can be seen from the English translations of the examples in (7), [Ø *teab mis*$_{Mod}$] has come to function either as an indefinite adjectival pronoun (such as *any* in (7b) and *whatever* in (7c)) or as a degree adverb (cf. *very* in (7a)), depending on the category of the constituent it modifies. The examples in (9) exemplify its use with a NP (9a), an AdjP (9b) and an AdvP (9c) respectively; the raw frequencies in the corpus of each of these categories are shown in Table 3.

(9) a. *Et polegi ta üldse* ***teab mis Paulus****, hoopis äraandja Peetrus!*
but NEG.be he at_all knows what Paul rather traitor Peter
'So he is definitely not a kind of Paul, rather a betraying Peter!' (CELL: ILU1980\stkt0051)

b. *Kahjuks ei ole tunne* ***teab mis mõnus****.*
unfortunately NEG be feeling knows what pleasant
'Unfortunately, the feeling is not very pleasant.' (CELL: ILU1990\ilu0597)

c. *Ning tegelikult pole see staari enesega* ***teab mis teisiti.***
and_also actually NEG.be this star self knows what differently
'And actually the star doesn't feel very differently.' (GCCE: PM1996)

Table 3. Categories of constituents modified by [Ø *teab mis*$_{Mod}$]

polarity	NP	AdjP	AdvP	ambiguous
positive	12	–	–	5
negative	37	16	3	19

The label "ambiguous" refers to examples with alternative scope interpretations, as in (10), where it is not clear whether *teab mis* modifies only the adjective *hea* 'good' or the whole NP *hea nalja* 'good joke':

(10) *Kiikavad Katet ja naeravad nagu* ***teab mis hea nalja*** *peale.*
cast a glance Kate-PRTV and laugh like knows what good.GEN joke.GEN on
'They cast a glance at Kate and laugh as though at a very good joke / at some good joke.'
(CELL: ILU1980\stkt0080)

The distribution in Table 3 is not significant (chi-square = 5.5706, p = .134), but the data indicate that [Ø *teab mis*$_{Mod}$] + NP is not as strongly associated with negation as [Ø *teab mis*$_{Mod}$] + AdjP or AdvP. Importantly enough, the data in the table appear to suggest, first, that [Ø *teab mis*$_{Mod}$] + AdjP or AdvP is blocked in positive sentences; second, that the five cases of scope ambiguity in positive clauses should probably be interpreted as involving the use of [Ø *teab mis*$_{Mod}$] as a modifier of the following NP, rather than of the following adjective.

In order to check whether this hypothesis was correct I had to resort to constructed examples. I therefore tried using the collocation [Ø *teab mis*$_{Mod}$ AdjP/AdvP] in a number of positive sentences with both veridical (11a) and nonveridical environments, the latter including conditionals (11b), disjunctives (11c), concessives (11d), clauses with epistemic verbs (11e) or with adversative verbs meaning 'deny, doubt' (11f), comparative clauses (11g), and clauses with implicit, rather than direct, negation (11h). The resulting sentences turned out to be ungrammatical in all cases, though comparative clauses and clauses with implicit negation sound slightly more acceptable than the others:

(11) a. **Jaan on teab mis tark.*
Jaan is knows what smart
'Jaan is very smart.' (lit. 'Jaan is knows what smart')

b. **Kui Jaan ole-ks teab mis tark, siis ta ole-ks õpetaja-le selle-st rääkinud.*

if Jaan be-COND knows what smart then he be-COND teacher-ALL this-ELA spoken
'If Jaan was very smart, then he would have talked to the teacher about this.'

c. *Jaan on kas teab mis tark või täielik idioot.
Jaan is either knows what smart or total idiot
'Jaan is either very smart or a total idiot.'

d. *Jaan olgu teab mis tark, aga praegu peab ta vanema-id kuulama.
Jaan be-JUSS knows what smart but now must he parents-PL.PRTV listen
'Jaan may be very smart, but now he must listen to his parents.'

e. *Jaan võib teab mis tark olla, aga minu meele-st käitub ta nüüd valesti.
Jaan can knows what smart be, but my mind-ELAT behaves he now wrongly.
'Jaan can be very smart, but in my opinion he behaves wrongly now.'

f. *Ma kahtlen, kas Jaan on teab mis tark.
I doubt whether Jaan is knows what smart
'I doubt if Jaan is very smart.'

g. ?Jaan on pigem väga töökas kui teab mis tark.
Jaan is rather very hard_working than knows what smart
'Jaan is very hard working rather than very smart.'

h. ?Ilma et ta teab mis tark ole-ks, pea-ks ta seda teadma.
without but he knows what smart be-COND must-COND he this know
'He doesn't have to be very smart to know that'. (lit. 'Without him being very smart he should know that.')

In view of the above, it can be argued that the collocation [Ø *teab mis*$_{\text{Mod}}$ AdjP/AdvP] is restricted to antiveridical contexts, such as sentences with direct negation (see Table 3) or with implicit negation as in (11h); in other words, it is licensed by operators which entail the falsity of the proposition. Considering, however, the limited acceptability of (11g), we have to admit that antiveridicality might be too strong a restriction.

5.2. [TABOO AGENT / Ø *teab kui*]

In this collocation the *wh*-slot is occupied by the word *kui* 'how', which can modify only adjectives and adverbs, that is, word classes capable of being inflected for comparison. Not surprisingly, therefore, this collocation has developed into a degree adverb meaning 'very'.

Table 4. Polarity distribution of *teab kui*[11]

polarity	with a taboo agent	with a zero agent
positive	20	7
negative	9	36

With regard to the data shown in Table 4, nine out of the 20 positive instances of the type [TABOO AGENT *teab kui*] occur in nonveridical contexts, which shows that this collocation is not necessarily associated with veridicality. However, the correlation of the type [Ø *teab kui*] with nonveridicality is clearly much stronger: it occurs 36 times in negative clauses and, in addition, six out its seven occurrences in positive sentences are also found in nonveridical environments; witness its use with the conditional mood (12a) and in the scope of a verb of propositional attitude (12b):

(12) a. *Mulle ilmus see pärandus lausa päästjana – muidu oleks piin*
I.ALL turned_up this inheritance simply rescuer otherwise be-COND torture
teab kui kaua kest-nud.

[11] The distribution is significant (chi-square = 18.3262, p = .000).

knows how long last-PART
'This inheritance simply turned up to be a saviour for me – otherwise the torture could have lasted very long.' (CELL: ILU1970\ilu0105)

b. *Peab ennast teab kui targaks.*
considers self-PRTV knows how smart-TRNSL
'S/he thinks s/he is very smart.' (EKSS)

I argued in Section 5.1 that the combination [Ø *teab mis* AdjP/AdvP] is licensed only in antiveridical contexts. [Ø *teab kui* AdjP/AdvP], despite its strong association with direct negation, appears to have a less restricted distribution. First, to judge at least from native speakers' intuitions, its use in positive clauses in concessive constructions (13a), with epistemic (13b) and adversative verbs (13c), comparatives (13d), and implicit negation (12e) seems to be more acceptable than was the case with the [Ø *teab mis* AdjP/AdvP] pattern.

(13) a. ?*Jaan ol-gu teab kui tark, aga praegu peab ta vanema-id kuulama.*
Jaan be-JUSS knows how smart but now must he parents-PL.PRTV listen
'Jaan may be very smart, but now he must listen to his parents.'

b. ?*Jaan võib teab kui tark olla, aga minu meelest käitub ta nüüd valesti.*
Jaan can knows how smart be but my mind-ELAT behaves he now wrongly
'Jaan can be very smart, but in my opinion he behaves wrongly now.'

c. ?*Ma kahtlen, kas Jaan on teab kui tark.*
I doubt whether Jaan is knows how smart
'I doubt if Jaan is very smart.'

d. ?*Jaan on pigem väga töökas kui teab kui tark.*
Jaan is rather very hard_working than knows how smart
'Jaan is very hard working rather than very smart.'

e. ?*Ilma et ta teab kui tark ole-ks, pea-ks ta seda teadma.*
without but he knows how smart be-COND must-COND he this know
'He doesn't have to be very smart to know that.' (lit. 'Without him being very smart he should know that.')

Secondly, the corpora yielded one instance of [Ø *teab kui* AdjP/AdvP] in a veridical context (see Table 4 above). In order to find analogous examples, I performed a search on the internet combining the string [Ø *teab kui*] with frequently used adjectives and adverbs. The results suggest that [Ø *teab kui*] can be used in veridical contexts only with adjectives meaning 'big' or 'high' (14). Such instances are, however, extremely rare and have a strongly colloquial flavour.

(14) *Edasise üle otsustasid kohalikud ülemad – jaoülema-st kompanii- ja teab*
rest over decided local commanders squad_commander-ELAT company and knows
kui kõrge ülema-ni välja.
how high commander-TRM untill
'Only local commanders, – starting from squad commanders and finishing with company and other very high (lit. knows how high) commanders, decided over the rest.'
(http://deepthought.ttu.ee/ajaleht/1996/20veebruar1996/tomp1.html)

Summing up so far, despite a few exceptions, the data presented in this and the preceding section confirm the decisive role played by nonveridicality in the licensing of the two agentless patterns used as polarity items, at least in written Estonian. In Section 6 below I will take a closer look at the development of these constructions and will try to show that the apparent exceptions noted above can be seen as instances of their earlier meaning. Such coexistence of old and new meanings known as "layering" is one of the essential characteristics of both grammaticalization and lexicalization (see Brinton and Traugott 2005: 143).

5.3. The remaining collocations

As already noted, the figures in Appendix B show that the nominative form *mis* of the word 'what' occurs far more frequently in the construction under discussion than any other question word. This is largely the result of a tendency, spreading out from the spoken language, to use *mis* instead of the other case forms of the word 'what' (e.g. *mi-da* 'what-PRTV), and instead of the words *milline*, *missugune* and *mihuke* (all meaning 'what kind of') and their case forms (see the items glossed in Appendix B as 'what_kind_of-'). This tendency naturally leads to the increasing marginalization of forms such as *millisteks* in (15), which is a form of *milline* inflected for the plural and the translative case:

(15) *Pea-te meid teab millis-te-ks metslas-te-ks,*
consider-2PL us knows what_kind_of-PL-TRNSL savage-PL-TRNSL
'You think we are some kind of savages,' (CELL: ILU1950\ilu0022)

The figures for each individual collocation in Appendix B are too low to warrant any conclusions regarding polarity preferences, but, in general, they all have much higher rates in positive than in negative sentences. This applies both to the patterns with an overt agent and to those with a zero agent.

In Estonian, as in many other languages (see Haspelmath 1997: 131), these collocations function as weakly grammaticalized indefiniteness markers. They can very often be replaced by indefinite pronouns.

6. Pathways of development

6.1. From [TABOO AGENT *teab* WH-] to [*Ø teab* WH-]

One may wonder whether there is any evidence to assume that the collocations without an agent are derived from those with a taboo agent. The third person marker *-b* in *teab* 'knows' suggests that there must have been an agent phrase at some earlier stage, but this does not imply, of course, that there ever was a taboo agent.

Although the recorded history of Estonian does not provide such evidence, support comes from Finnish, as exactly the same formal pattern is attested in Finnish. Taboo intensifiers can occur either with an overt agent (e.g. *Herra*/*Luoja*/*piru ties WH-* 'the Lord/Creator/Devil knows WH-') or without it (*ties WH-* 'knows WH-'). The element *–s* on the 'know' predicate in Finnish seems to go back to the ending *-si* of the third person imperfect tense form (NS; Yli-Vakkuri 1986: 176–179). This would mean that, in Finnish, expressions involving taboo intensifiers were originally used only with past-time reference; nowadays, however, *-s* in this construction is opaque, as will be discussed in what follows.

The shortening of the imperfect marker *-si* to *-s* is still productive in colloquial Finnish, as shown in (16), where the shortened form *ties* is necessarily interpreted as having past-time reference:

(16) *Pekka tie-s missä Jukka on.*
Pekka know-PAST.3SG where Jukka is
'Pekka knew where Jukka is.'

In (17), by contrast, though *ties* is spelt identically and indeed originates in the same morphophonemic material, it is no longer analyzable as a tense morpheme, as is clear from the translation into English. Moreover, native speakers of Finnish would find it difficult to explain the semantic import of the element *–s*. The taboo agent itself, unlike *Pekka* in (16), is semantically empty; that is, it does not contribute to the sentence's propositional meaning, and would be interpreted simply as an emphatic expression. Finally, native speakers would also agree that *Herra* 'Lord' could be omitted (i.e. *ties missä Jukka on* 'nobody knows where Jukka is') without altering the acceptability or the meaning of the sentence.[12]

[12] Although it would be slightly less forceful as an exclamation.

(17) *Herra ties missä Jukka on!*
Lord knows where Jukka is
'Nobody knows where Jukka is!' (lit. 'Lord knows where Jukka is!')

The fact that the opaque *–s* is used in Finnish only in two very specific construction types, one with a taboo agent and the other without it, shows that the latter, shorter, construction type must derive from the former, and points, furthermore, to an analogous path of development for the corresponding Estonian patterns. Coming now back to these, it deserves mention that the [Ø *teab* WH-] pattern has clearly lost its syntagmatic character and is therefore treated as an idiomatic unit. This can be seen from Estonian examples in (18) and (19). While the topicalization of the complement clause *mis mees tast saab* "what kind of a man he will become" is fully grammatical in the case of collocations with overt agent (18b), it is ungrammatical with zero agent collocations (19b):

(18) a. *Jumal teab mis mees ta-st saab!*
God knows what man he-ELAT becomes
'God knows what kind of a man he will become!'

b. *Mis mees ta-st saab, jumal teab!*
what man he-ELAT becomes God knows
'What kind of a man he will become, God knows!'

(19) a. Ø *Teab mis mees ta-st saab!*
Ø knows what man he-ELAT becomes
'(Who) knows what kind of a man he will become!' (EKSS)

b. **Mis mees ta-st saab,* Ø *teab!*
what man he-ELAT becomes Ø knows
'*What kind of a man he will become, Ø *knows*!'

This pattern indicates that zero agent collocations are linguistically treated as idioms and that their idiomatization (= weak lexicalization) correlates with the drop of the agent.

6.2. The shift in polarity

The development of negative polarity items from expressions such as "who/God/devil knows WH" can be partly explained in terms of CONVERSATIONAL IMPLICATURE and the ONTOLOGY OF THE LANGUAGE TABOO. Let us examine the common (at least in the European languages) linguistic functions of the expression [TABOO AGENT *knows* WH].

The literature on polarity has drawn attention to the fact that in many languages the class of taboo expressions is rich in polarity items (see, e.g., von Bergen & von Bergen 1993: 137; van der Wouden 1997: 68; Postma 2001). Consider, for instance, the PPI *hell of* in "He is a hell of a man" or the NPI *a damn thing* in "I didn't see a damn thing". None of the studies of polarity of which I am aware, however, considers the type of taboo expressions the present paper investigates.[13] It was noted in Section 3 that an 'extreme' individual (either sacred or demonic) fills the agent slot in these types of expression. The use of a sacred or demonic individual as a possessor of certain knowledge implies, in fact, that this knowledge is inaccessible to humans.[14] This is an example of generalized conversational implicature, cf. (20):

(20) **God knows** what the weather is in Chukotka.

[13] What concerns Estonian, the taboo words *jumal* "God"and *kurat* "Devil" have been previously studied in connection with their function as intensifiers (see Metslang 1997; Pajusalu 2006), but not as polarity sensitive expressions.

[14] And this is explicitly shown by bywords like *God only knows* or *Heaven only knows* (compare also the Estonian saying *seda teab ainult jumal taevas* "Only God in the Heaven knows this.")

⊃ (¬∃: person *x*) (*x* knows what is the weather)
Nobody knows what the weather is in Chukotka.

The universal reading[15] of the implicature derives from the basic meaning of taboo. According to Postma (2001: 325), TABOO, in its anthropological sense, refers to anything that brings about "emotional ambivalence", i.e. something extremely good or extremely bad. Intensifiers such as "God knows WH" thus possess an inherent quantification that Postma calls "extreme-degree quantification" (p. 300). If the knowledge possessed by God (or other taboo-agent) is considered the farthest extreme on the scale of knowledge possessed by a single individual, and the knowledge about the weather in Chukotka is considered extremely unavailable in the speech situation, then *God* 'conversationally implies' that the speaker is referring to 'none (of the humans)'.

What is important here is that the implicature is negative. Taboo intensifiers are associated exclusively with negative implicature. For the sake of clarity, consider another example:

(21) This box was full of **God knows what**.

An interpretation of (21) in which the speaker found useful and/or expected things in the box is not possible. The only acceptable interpretation would be that the box yielded useless material, in which case the English affix *-less* would convey the negative meaning.

Returning to the situation in Estonian, we can hypothesize that the negative implicature conveyed by the expression [TABOO AGENT *teab* WH] and the polarity sensitivity of its subtypes [Ø *teab mis* AdjP/AdvP] and [Ø *teab kui* AdjP/AdvP] are related to each other. As a next step, we need to specify the nature of this relation. Hoeksema (1994) has suggested that meaning may be the crucial factor in the grammaticalization of NPIs. He notes that NPIs are not distributed arbitrarily over the lexicon so inferences may be drawn as to both the meaning domains in the lexicon in which NPIs most probably will be found, as well as to those domains that will not be rich sources of NPIs. Among the frequently mentioned domains where NPIs are found are those containing expressions of intolerance, dislike or indifference[16] (see van der Wal 1996 for a comprehensive list of domains and expressions). It is evident that taboo intensifiers very often express such meanings. Consider the following corpus example expressing dislike on behalf of the speaker:

(22) *Eesti* *riik* *ei* *ole* *huvitatud* *kas* *talumaad* *lähevad* *põllumees-te*
Estonian government NEG be interested whether farmlands go farmer-PL.GEN
järeltulija-te *või* *jumal* *teab* *ku-st* *välja* *ilmu-nud* *äritseja-te*
descendant-PL.GEN or God knows where-ELAT out appear-PART trafficker-PL.PART
kätte ...
into_hand
'The Estonian government is not interested whether farmlands go to the descendants of farmers or to shady businessmen who show up from God knows where.' (GCCE: PM1999)

Hoeksema extends the importance of meaning to grammaticalization further and assumes that certain linguistic units are predestined to become polarity sensitive by reason of their original meaning. Given this basic assumption, I suggest that taboo intensifiers were predestined to become NPIs because of their pragmatic environments: they were steadily associated with negative implicature, which in the course of the lexicalization process was encoded as sensitivity to grammatical negation.

However, the development of the items [Ø *teab mis* AdjP/AdvP] and [Ø *teab kui* AdjP/AdvP] into NPIs cannot be explained in terms of pragmatics or semantics alone. Otherwise, one may ask why the remaining collocations (discussed in section 5.3) did not become NPIs. We saw that they occur mostly in positive sentences. Consider again the *wh*-words presented in Appendix B. In principle, each of these *wh*-words determines the word class of the fixed collocation [TABOO AGENT *teab* WH], and, because of the types of *wh*-words present in the appendix, there are only two major possibilities: the collocation is analysed either as indefinite pronoun (e.g. Eng. *God-knows-where*) or as modifier (e.g. Eng. *God-knows-what-kind-of*, *God-knows-how*). The relevant question here is why Estonian

[15] Note that ¬∃xWx = ∀x¬Wx (where W is 'knows what the weather is in Chukotka').
[16] Note that these domain labels are derived with the negative prefixes *in-* and *dis-*.

lexicalized into NPIs only those collocations that owing to their *wh*-words were analyzed as modifiers, more specifically as AdjP/AdvP-modifiers. In other words, why did Estonian lexicalize negative modifiers and not negative pronouns, despite the fact that taboo intensifiers provided an adequate supply of suitable pronoun candidates?

The answer to this question unfolds as we look at the patterns of polarity marking according to word class. Consider the indefinite pronoun system of Estonian. Unlike Finnish, Hungarian and many other Finno-Ugric languages, Estonian does not have a distinct class of indefinite pronouns used with verbal negation (cf. Fi. *kukaan* 'nobody' and Hu. *senki* 'nobody'). Consequently, Estonian does not even have the relevant class of words into which new items like [Ø *teab* WH_{Pron}] could potentially be admitted. Thus, from two primary options – lexicalization of taboo intensifiers into either negative modifiers or into negative pronouns – the latter was discarded owing this specific structural pattern.

With regard to the first option, we saw that Estonian lexicalized only AdjP/AdvP-modifiers into NPIs. This raises the question of why the nominal modifiers, i.e. expressions like *God-knows-what-kind-of*, did not become negative polarity items. There are five *wh*-words in Appendix B that modify nouns and that in collocations with a taboo (or zero) agent and *teab* 'knows' render complex nominal modifiers. These five words are *milline*, *missugune*, *mihuke* (all meaning 'what kind of')[17], *mitu* 'how many' and *mitmes* 'which (of a definite set)'. Consider (23) where the partitive form of *mihuke* occurs:

(23) *Iga sell varjab, nagu jumalteab mihukes-t ametisaladus-t, aga paistab,*
every guy hides as_if God_knows what_kind_of-PRTV professional_secret-PRTV but seems
nagu ole-ksi-n mina üksi see togu, kes midagi ei tea.
as_if be-COND-1SG I alone that fool who nothing NEG know
'Every guy is secretive as if hiding God knows what kind of a professional secret, but it seems to me that I am the only fool who doesn't know anything.' (CELL: ILU1930/lmg0001)

Additionally, we saw in section 5.1 that the nominative case form of the word *mis* 'what' can also modify nouns, which means that this form generates complex NP-modifiers when fixed in taboo intensifier constructions. Recall the example (6b) where the expression *kurat teab mis* "the Devil knows what" functions as a nominal modifier in *kurat teab mis jama* "the Devil knows what problems" = "all sorts of problems". The answer to the question of why such NP-modifiers did not become NPIs is trivial. These complex items function as pronominal modifiers, and, as such as weakly grammaticalized pronouns, which eventually, in a more advanced stage of their univerbation, could be listed in the grammar as adjectival pronouns (such as Eng. *any* in "I don't have any enemies" or *which* in "I haven't decided which flight to take"). In other words, the grammatical output of the modifier *God-knows-what* in "He eats God-knows-what food." is an adjectival pronoun like *any* in "He eats just any food." Since the only possible output of these complex NP-modifiers are pronouns, and, since Estonian does not have a relevant class of negative pronouns, they could not become NPIs whatsoever.

The situation with degree modifiers is, however, strikingly different. Estonian has a very pervasive system of polarity distinctions in degree modification; most of Estonian degree modifiers are in fact polarity items. Furthermore, the raise of negative degree modifiers can be seen as a compensatory development. Estonian has a significant number of monomorphemic AdjP/AdvP-modifiers (e.g. *üsna* 'rather' in (24), *üpris* 'rather' in (25) and *kaunis* 'pretty' in (26)) that function as positive polarity items.

(24) a. *Jaan on üsna tark.*
'Jaan is rather smart.'
b. **Jaan pole üsna tark.*
*Jaan NEG.be rather smart.
'*Jaan is not rather smart.'

(25) a. *Jaan on üpris tark.*
'Jaan is rather smart.'
b. **Jaan pole üpris tark.*

[17] The nominative case forms of these words did not occur in the corpus. The attested case forms are glossed in Appendix B as 'what_kind_of'.

Jaan NEG.be rather smart
'*Jaan is not rather smart.'

(26) a. *Jaan on kaunis tark.*
'Jaan is pretty smart.'
b. **Jaan pole kaunis tark.*
Jaan NEG.be pretty smart
'*Jaan is not pretty smart.'

Negative modifiers dedicated to filling the functional gap left by the ungrammaticality of the above items in negative sentences are, for example, *kuigi* 'very, especially' in (27) and *nii väga* 'all that' in (28).

(27) a. *Jaan pole kuigi tark.*
Jaan NEG.be very smart
'Jaan is not very smart.'
b. **Jaan on kuigi tark.*
'Jaan is very smart.'

(28) a. *Jaan pole nii väga tark.*
Jaan NEG.be so much smart
'Jaan is not all that smart.'
b. **Jaan on nii väga tark.*
'*Jaan is all that smart.'

These negative degree modifiers exhibit as a rule compositional (multi-morphemic) structure[18] and appear thus to be more recent lexicalizations, much like the negative degree modifiers [*Ø teab mis*] and [*Ø teab kui*]. Based on this evidence, we can stipulate, that zero agent taboo intensifiers that functioned as AdjP/AdvP-modifiers became polarity items because they were subjected to an overall pressure towards an explicit marking of polarity in the degree modification of the language. Similarly to the items in (27) and (28), they seem to have responded to a compensatory need for negative polarity items in the system of degree modification.

Thus, the development of the items [*Ø teab mis* AdjP/AdvP] and [*Ø teab kui* AdjP/AdvP] into NPIs is due to two cooperating conditions, namely, *a)* the negative implicature conveyed by the original meaning of taboo intensifiers; and *b)* the need for explicit marking of the (negative) polarity in the system of degree modification of the language. As the second condition was not satisfied for the taboo intensifiers functioning as pronouns or nominal modifiers, they did not become NPIs.

If these conjectures were true, then one would expect that a categorial shift from one word class to another would lead to a shift in polarity sensitivity of an item. The Estonian data provides a neat example of such a development. Consider the distribution of the NP-modifier [*Ø teab mis*]. We saw in Table 3 (Section 5.1) that this item is attracted by negation, as evidenced by 37 occurrences in negative sentences and only 12 in positive sentences. Furthermore, half of the occurrences in positive sentences were found in nonveridical contexts. Despite this appeal to nonveridical contexts, however, the distribution of this item cannot be encompassed by the notion of nonveridicality, as there were six occurrences found in veridical contexts. They are sound reasons to believe that the AdjP/AdvP-modifier uses of [*Ø teab mis*] developed from the NP-modifier uses of this item, and, as the figures in Table 3 indicate, this development correlates with an increase of this item's sensitivity to overt negation. The bridging contexts for the reanalysis from NP- to AdjP/AdvP-modification are provided by the scope ambiguity exemplified in (10) (Section 5.1). Another example of such ambiguity is presented in (29):

(29) *Mul pole küll praegus-te noor-te-ga* ***teab mis lähedas-t***
I.ADE NEG.be indeed modern-PL.GEN youngster-PL-COM knows what close-PRTV
kontakt-i *...*
contact-PRTV

[18] The word *kuigi* is a composed from the temporal-conditional marker *kui* and the intensifying clitic *-gi*.

‘I really don’t have any close contacts / very close contacts with today’s youth…’ (CELL: ESBI1980\tet0025)

In this example [*Ø teab mis*] has two alternative readings, either as indefinite pronoun (‘any’) or as the degree adverb (‘very’). In the first case it modifies the NP ‘close contacts’, whereas in the second it modifies only the adjacent AdjP ‘close’.[19] Although both readings are synchronically available, it seems that they represent two subsequent stages on the lexicalization cline of [*Ø teab mis*]. In other words, it seems that the NP-modifier [*Ø teab mis*] was reanalysed in contexts like (29) as AdjP-modifier. We have already mentioned in Section 5.3 that the nominative form *mis* of the word ‘what’ tends to be used instead of other *wh*-words. This functional extension of *mis*, however, has certain boundaries. *Mis* can be used as NP-modifier (see 30) as a substitute of *milline*, *missugune* and *mihuke* ‘what kind of’, but not as AdjP/AdvP-modifier (see 31).

(30) *Ma ei tea **mis** **toitu** ta sööb.*
I NEG know what food-PRTV s/he eats
‘I don’t know what kind of food s/he eats.’

(31) **Ma ei tea **mis** **kaugel** Lapimaa on.*
I NEG know what far Lapland is
‘*I don’t know what far Lapland is.’

Taking into account this distribution of *mis*, I assume that the NP-modifier use of *teab mis* in (29) reflects a stage during which this collocation could still be accessed in terms of its components. Although, because of the lack of an agent the collocation *teab mis* is already fixed to a certain extent, the NP-modifier reading of the collocation conforms to the independent use of *mis* as NP-modifier (cf. example (30)). In contrast, the AdjP-modifier reading of *teab mis* in (29) does not conform to the independent uses of *mis* (cf. example (31)), a fact which points to a more advanced stage of the lexicalization process where *mis* is a meaningless element of the lexical item [*Ø teab mis*]. What we observe here is a decrease of the syntactic scope of the expression. In (29) there is ambiguity between the original wide scope interpretation where [*Ø teab mis*] modifies the NP and the narrow scope interpretation where it modifies only the adjacent AdjP. This narrow scope interpretation was triggered by the holistic reading of [*Ø teab mis*], in which the original selection restrictions of the independent question word *mis* no longer applied.

This development provides, in fact, straightforward evidence against the claim of von Bergen and von Bergen (1993: 130) that the historical change in the meaning of the linguistic elements is irrelevant for the synchronic description of polarity items. We saw that the reanalysis of [*Ø teab mis*] from a NP-modifier to a AdjP/AdvP-modifier led to increased polarity sensitivity of this form (recall the figures in Table 3). The circumstantial evidence thus shows how the functional development of a linguistic element had direct consequences for its synchronic description as polarity item. Although the notion of (non)veridicality captures the distributional regularities of polarity items in many languages, and seems thus to reflect a common markedness rule, it appears that the process of “becoming sensitive” to certain contexts is a more gradual one. We saw that the NP-modifier [*Ø teab mis*] was attracted by nonveridical contexts with only few uses in veridical contexts, and that its descendant – the AdjP/AdvP-modifier [*Ø teab mis*] – was strongly attracted by antiveridical contexts with only few colloquial uses in other nonveridical contexts. In both cases, these exceptional uses can be considered as relics of an earlier function of the form [*Ø teab mis*] and their coexistence can be accounted for by the notion of ‘layering’ as formulated within grammaticalization theory (Hopper & Traugott 2003: 106, 124–126).

7. The counterparts of [TABOO AGENT *know*-3SG WH AdjP/AdvP] in Finnish and Latvian

[19] It might seem questionable to some native speakers as to whether there is ambiguity in (29). It probably depends on the semantic properties of both the AdjP and the head of the NP. Nevertheless, the fact that some speakers find such sentences as these to be ambiguous is informative enough.

We have already observed that the ellipse of the taboo agent of the verb ‘to know’ occurs not only in Estonian, but also in Finnish. The Finnish examples in (32) and (33) illustrate this again:

(32) *Ilosaarirock on **luoja** **ties** **kuinka** vanha tapahtuma ja muistoja suure-n*
Ilosaarirock is Creator knows how old event and memories big-GEN
kivisatee-n jälkeis-i-stä ajo-i-sta lähtien löyty-y Alpo Kettuse-n
rockfall-GEN following-PL-ELAT time-PL-ELAT since be_found-3SG Alpo Kettunen-GEN
kirjoitta-ma-sta historiiki-stä.
write-SUP-ELAT chronicle-ELAT
‘Ilosaarirock is very old (lit. the Creator knows how old) event and one can find memories about the time following the big rock fall in Alpo Kettunen’s chronicle. (checked for grammaticality by Juha-Matti Aronen)

(33) *Ilosaarirock on **Ø** **ties** **kuinka** vanha tapahtuma ja muistoja suure-n kivisatee-n*
Ilosaarirock is Ø knows how old event and memories big-GEN rockfall-GEN
jälkeis-i-stä ajo-i-sta lähtien löyty-y Alpo Kettuse-n kirjoitta-ma-sta
following-PL-ELAT time-PL-ELAT since be_found-3SG Alpo Kettunen-GEN write-SUP-ELAT
historiiki-stä.
chronicle-ELAT
‘Ilosaarirock is very old (lit. knows how old) event and one can find memories about the time following the big rock fall in Alpo Kettunen’s chronicle.’ (www.ilossarirock.fi/2000/kohti_rokkia.html)

However, despite the possibility of omitting the agent, the shift in polarity has not taken place in Finnish. Both types of constructions – the ones with an explicit taboo agent like (32) and the ones without like (33) – are, regardless of their *wh*-word, predominantly found in the grammatically positive sentences.

In Latvian on the other hand, we find an opposite development. Latvian is a member of the Baltic branch of Indo-European, and is the immediate neighbour of Estonian in the south. Similarly to the situation in English and in other well-known European languages, the lexicalization process in Latvian did not result in agent drop. The process was conducted at the formal level by phonological attrition of the predicate and its agglutination and contraction with the preceding word for ‘God’ (see ME, LEV):

Dievs zina cik >	*Dievs zin cik* >	*diezin cik* >	*diezcik //*
‘God know-3SG how-(much)’	‘God know-$_{\text{CONTRACTED}}$ how-(much)’	‘God-$_{\text{CONTRACTED}}$-know-$_{\text{CONTRACTED}}$ how-much’	‘God-$_{\text{CONTRACTED}}$-know-$_{\text{CONTRACTED}}$-how

Although Estonian shares the same formal developments with Finnish, namely the ellipse of the agent, in their semantic representation the developments in Estonian seem to be more similar with those that took place in Latvian.[20] Compare the following Latvian sentences:

(34) a. ***Dievs** **zina**, **cik** mums tās ir vajadzīg-i …*
God knows how_much we.DAT those are necessary-MASC.PL
‘God knows how much we need those …’ (<vip.latnet.lv/LPRA/kalme/par_varonu_godinasanu.htm>)

b. *Gan **dievs** **zin**, **cik** tā smaga!*
well God know-$_{\text{CONTRACTED}}$ how_much that.FEM heavy.FEM
‘Well, God knows how heavy that is.’(<www.ailab.lv/Teksti/Senie/ApsJ/ApsJ0403.htm>)

c. *Un **diezin** **cik** ilgi tas ilgs.*
and God$_{\text{CONTRACTED}}$-know-$_{\text{CONTRACTED}}$ how long_time that.MASC lasting
‘And God-knows how long that will last. (<www.dziesmas.lv/database/teksts.asp?ID=1802>)

d. *Baltkrievijā saites starp universitāti un akadēmiju nav bijušas*
Belarus.LOC bonds between universities and academy NEG been
***diezcik** labas …*

[20] It should be noted, however, that there is an adverb in Finnish whose history does not display agent drop but formal changes similar to those of Latvian; consider Fi. *kenties* ‘maybe’ < *ken tie-si* ‘who know-PAST.3SG’.

God-CONTRACTED-know-CONTRACTED-how good
'The relationship between the university and the academy in Belarus has not been very good …'
(www.lza.lv/ZV/zv021100.htm <http://www.lza.lv/ZV/zv021100.htm>)

The exact distribution of each of these items must be studied further, but the similarity with the corresponding Estonian items is striking. In Latvian, the bold items in (34a-b) are found mostly in positive sentences with an interjectional reading, whereas the bold item in (34d) is used as negative degree adverb. What we observe here is exactly the same kind of shift in polarity as the one discussed for Estonian.

8. Conclusion

This paper examined a certain class of expressions called taboo intensifiers. It provided a solution to two problems concerning the properties of this class of expressions in Estonian. The first one was the reanalysis of taboo intensifiers from complex syntactic constructions into lexical items, and the second was the change of polarity associated with this reanalysis.

In the sections dealing with polarity, we arrived at the conclusion that the distributional regularities of the type [TABOO AGENT *teab* WH AdjP/AdvP] are virtually captured by the notion of (non)veridicality. The sensitivity to nonveridical contexts was already present in the initial stage of lexicalization by means of the negative implicature and the semantics of taboo. This inherent sensitivity was afterwards encoded in syntax with a new distributional constraint – sensitivity to overt negation. The raise of this constraint was conditioned by general structural patterns, such as word class specific differences in polarity marking.

At a formal level, lexicalization of the expression [TABOO AGENT *teab* WH] was carried out by a formal reduction of arguments (drop of the taboo agent) and coalescence (increased bondedness of previously autonomous elements). These processes brought about ambiguity, which triggered the raise of a novel lexical item restricted in its occurrence to negative sentences. By recognizing the role of this ambiguity in the process of lexicalization of the negative adverb [*Ø teab mis*], we challenged the view according to which the historical change in the meaning of the linguistic elements is irrelevant for the synchronic description of polarity items.

Finally, it we saw that while the formal developments of Estonian taboo intensifiers follow a common Finnic model, the patterns of their usage seem to be shared with the Indo-European neighbour to the south.

References

Bergen, Anke von and Karl von Bergen, 1993. Negative Polarität im Englischen. Tübingen: Narr.

Brinton, Laurel J. and Elizabeth Closs Traugott, 2005. Lexicalization and Language Change. Cambridge: Cambridge University Press.

EKSS = Eesti kirjakeele seletussõnaraamat. VI köide, 1. vihik. Eesti Keele Instituut. Tallinn 2001.

Giannakidou, Anastasia, 1998. Polarity Sensitivity as (Non)veridical Dependency. John Benjamins: Amsterdam.

Giannakidou, Anastasia, 2001. Varieties of polarity items and the (non)veridicality hypothesis. In: J. Hoeksema, H. Rullmann & V. Sánchez-Valencia (eds.), Perspectives on Negation and Polarity Items. Linguistics Today 40. John Benjamins: Amsterdam–Philadelphia, 99–127.

Haspelmath, Martin, 1997. Indefinite Pronouns. Oxford: Clarendon Press.

Hoeksema, Jack, 1994. On the grammaticalization of negative polarity items. In: S. Gahl, A. Dolbey & C. Johnson (eds.), Proceedings of the 20th annual meeting of the Berkeley Linguistics Society. Berkeley, 273–282.

Hoeksema, Jack and Henny Klein, 1995. The development of a new adverb of degree: the case of Dutch *even*. MS Groningen.

Hoeksema, Jack and Henny Klein, 1996. From comparative to adverb of degree: Dutch *even*. In: R. Jonkers, E. Kaan & A. Wiegel (eds.), Language and Cognition 5. Yearbook 1996 of the Research Group for Theoretical and Experimental Linguistics of the University of Groningen. Groningen, 59–70.

Hopper, Paul J. and Elisabeth Closs Traugott, 2003 [1993]. Grammaticalization. Cambridge: Cambridge University Press.
Israel, Michael, 1996. Polarity Sensitivity as Lexical Semantics. Linguistics and Philosophy 19, 619–666.
Kadmon, Nirit and Fred Landman, 1993. Any. Linguistics and Philosophy 16, 353–422.
Klima, Edward, S. 1964: Negation in English. In: J. A. Fodor & J. J. Katz (eds.), The Structure of Language. Englewood Cliffs: Prentice Hall, 246–323.
Krifka, Manfred, 1995. The Semantics and Pragmatics of Polarity Items. Linguistic Analysis 25, 209–257.
Ladusaw, William, 1979. Polarity Sensitivity as Inherent Scope Relation. Ph. D. dissertation, University of Texas. (Published New York: Garland, 1980.)
Lakoff, Robert, 1969. Some reasons why there can't be *some-any* rule. Language 45, 608–615.
Lehmann, Christian, 2002. New Reflections on Grammaticalization and Lexicalization. In: I. Wischer & G. Diewald (eds.), New Reflexion on Grammaticalization. Typological Studies in Language 49. Amsterdam – Philadelphia, 1–18.
LEV = Karulis, Konstantīns, 2002. Latviešu etimoloģijas vārdnīca. Rīga.
Linebarger, Marcia, 1981. The Grammar of Negative Polarity. Indiana University Linguistics Club. Masachusetts Institute of Technology.
ME = Mīlenbahs, Kārlis, 1923–1946. Latviešu valodas vārdnīca. I–IV. Rediģejis, papildinājis, turpinājis J. Endzelīns. Rīga.
Metslang, Helle, 1997. Maksimaalsuse ja minimaalsuse väljendamisest eesti püsiühendites. In: M. Erelt, M. Sedrik, E. Uuspõld (eds.), Pühendusteos Huno Rätsepale. Tartu Ülikooli eesti keele õppetooli toimetised 7. Tartu, 139−154.
Moreno Cabrera, Juan, C., 1998. On the Relationship between Grammaticalization and Lexicalization. In: A. Giacalone Ramat & P. J. Hopper (eds.), The Limits of Grammaticalization. Typological Studies in Language 37. Amsterdam–Philadelphia, 211–227.
Newmeyer, Frederick, J. 1998. Language Form and Language Function. Cambridge, Massachusetts: MIT Press.
NS = Nykysuomen sanakirja. (1951−1961). Tenth edition (1988 I–V). WSO: Porvoo – Helsinki – Juva.
Õim, Asta, 2000. Fraseoloogiasõnaraamat.Tallinn.
Pajusalu, Renate, 2006. *jumal* ja *kurat* argikeeles. In: E. Kasak & A. Kull (eds.), Kuradi tark jumala loll. 2.−3. veebruaril 2006 Tallinna Ülikoolis toimunud konverentsi "Jumal teaduses" ettekannete kogumuk. Tallinn, 17−32.
Postma, Gertjan, 2001. Negative polarity and the syntax of taboo. In: J. Hoeksema, H. Rullmann & V. Sánchez-Valencia (eds.), Perspectives on Negation and Polarity Items. Linguistics Today 40. John Benjamins: Amsterdam–Philadelphia, 283–330.
Progovac, Liljana, 1994: Negative and Positive Polarity: A Binding Approach. Cambridge: Cambridge University Press.
Rostila, Jouni, 2006. Storage as a way to grammaticalization. Constructions 2006, No. 1. http://www.constructions-online.de/articles/
van der Wal, Sjoukje, 1996. Negative Polarity Items and Negation: Tandem Acquisition. Doctoral dissertation defended August 29, 1996 at the University of Groningen.
van der Wouden, Ton, 1997: Negative Contexts. Collocation, polarity and multiple negation. Routledge Studies in Germanic Linguistics. London – New York.
Yli-Vakkuri, Valma, 1986. Suomen kieliopillisten muotojen toissijainen käyttö. Publications of the Department of Finnish and General Linguistics of the University of Turku 28. Turku.
Zwarts, Frans, 1995. Nonveridical contexts. Linguistic Analysis 25, 286–312.

Abbreviations

1 – first person, 2 – second person, 3 – third person, ABL – ablative case, ADE – adessive case, ALL – allative case, COM – comitative case, COND – conditional mood, DAT – dative case, ELAT – elative case, FEM – feminine, GEN – genitive case, INE – inessive case, JUSS – jussive mood, LOC – locative case, MASC – masculine, Mod – modifier, NEG – negation marker, NPI – negative polarity item, PRTV – partitive case, PAST – past tense, PL – plural, PART – participle, PI – polarity item, PPI – positive polarity item,

Pron – pronoun, SG – singular, SUP – supine, TRM – terminative case, TRNSL – translative case

Appendix A

The taboo agents and their frequencies in the corpus

taboo-agent	occurrences
jumal 'God'	76
kes 'who'	73
kurat 'Devil'	22
tont 'ghost; bogey'	16
pagan 'pagan'	7
taevas 'heaven'	3
issand '(the) Lord'	1
põrgu 'hell'	1
pergel 'devil, pagan'	1
juudas 'Judas'	1
tühi 'empty'	0
mõni 'some'	0

Appendix B

Taboo intensifiers according to the postverbal question word (number on the left – occurrences in positive sentences, number on the right – occurrences in negative sentences)

question word	with a taboo-agent	with a zero-agent
***mis* 'what.NOM'**	**60/4**	**25/76**
mi-da 'what-PRTV'	17/4	14/2
mille 'what.GEN'	2/–	1/–
mille-s 'what-INE'	1/–	–
mille-st 'what-ELAT'	2/1	2/–
mille-le 'what-ALL'	1/–	–
mille-ga 'what-COM'	5/–	1/–
millis-t 'what_kind_of-PRTV'	1/–	–
millise 'what_kind_of.GEN'	1/–	–
millise-st 'what_kind_of-ELAT'	1/–	–
millise-lt 'what_kind_of-ABL'	–	1/–
millise-i-d 'what_kind_of-PL-PRTV'	2/–	–
millis-te-s 'what_kind_of-PL-INE'	–	1/–
millis-te-ks 'what_kind_of-PL-TRNSL'	–	1/–
missuguse 'what_kind_of.GEN'	–/1	–
missuguse-i-d 'what_kind_of-PL-PRTV'	1/–	–
missugus-te 'what_kind_of-PL.GEN'	1/–	–
mihukes-t 'what_kind_of-PRTV'	1/–	–
millal 'when'	–	–/1
mitu 'how_many.NOM'	–	3/–
mitut 'how_many-PRTV'	–	1/–
mitme-st 'how_many-ELAT'	–	1/–
mitmes 'which(of a definite set).NOM'	2/–	1/–
mitmenda-t 'which(of a definite set)-PRTV'	3/1	2/–
mistarvis 'what_for'	1/–	–
mispärast 'what_for'	1/–	–
***kui* 'how'**	**20/9**	**7/36**
kus 'where'	7/–	7/–
kust 'where_from'	9/–	8/1

kuhu ‘where_to’	19/3	3/–
kuidas ‘in_what_manner’	4/–	–
kes ‘who’	4/–	4/1
kelle ‘who.GEN’	5/2	3/–
kelle-le ‘who-ALL’	1/–	1/–
kelle-ga ‘who-COM’	2/–	–
kelle-ks ‘who-TRNSL’	1/–	–
kas ‘yes/no’	1/–	–
total	176/25 (=201)	87/117 (=204)