

TARTU RIIKLIKU ÜLIKOOLI TOIMETISED
УЧЕННЫЕ ЗАПИСКИ
ТАРТУСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА

ALUSTATUD 1893. a.

VIINIK 172 ВЫПУСК

ОСНОВАНЫ в 1893 г.

Х. А. Хольм

ВЫДЕЛЕНИЕ ПАРАДИГМЫ ПЕРВОГО МОРФОЛОГИЧЕСКОГО
ТИПА НА РАЗЛИЧНЫХ ПОДЪЯЗЫКАХ ПРИ СТАТИСТИКО-
КОМБИНАТОРНОМ МОДЕЛИРОВАНИИ ЭСТОНСКОЙ
МОРФОЛОГИИ



ТАРТУ 1965

TARTU RIİKILKU ÜLIKOOLI TOIMETISED
УЧЕННЫЕ ЗАПИСКИ
ТАРТУСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА

ALUSTATUD 1893. a.

VIINIK 172 ВЫПУСК

Основаны в 1893 г.

Х. А. Хольм

**ВЫДЕЛЕНИЕ ПАРАДИГМЫ ПЕРВОГО МОРФОЛОГИЧЕСКОГО
ТИПА НА РАЗЛИЧНЫХ ПОДЪЯЗЫКАХ ПРИ СТАТИСТИКО-
КОМБИНАТОРНОМ МОДЕЛИРОВАНИИ ЭСТОНСКОЙ
МОРФОЛОГИИ**

Тарту 1965

Редакционная коллегия:

И. Куль, Х. Рятсеп (ответств. редактор), С. Смирнов и Э. Ууспюльд.

Настоящее исследование является попыткой сравнительного изучения различных подязыков эстонского языка.

В основе работы лежат подязыки радиозлектроники и публицистики, обработанные алгоритмом статистико-комбинаторного моделирования языка¹.

Сравнению подвергаются условные вероятности графем по позициям в слове, множество коррелятивных функций букв, ложные шаги алгоритма, истинные шаги алгоритма.

Ход алгоритма статистико-комбинаторного моделирования эстонского языка на материале подязыков радиозлектроники и публицистики уже публикуется автором настоящего исследования², и поэтому в ряде случаев дается не сам ход алгоритма, а его внеалгоритмическая лингвистическая интерпретация.

I. ЗАВИСИМОСТЬ ХОДА АЛГОРИТМА ОТ ПОДЪЯЗЫКА. ХОД АЛГОРИТМА ПРИ РАЗЛИЧНЫХ ПОДЪЯЗЫКАХ

В зависимости от подязыков порядок выделения типов изменяется, но так как объект типизации — язык — остается неизменным, то ясно, что в конечном счете выделяются одни и те же типы. Содержательно выделение типа зависит от лингвистической вероятности отдельных баз в исследуемых текстах³. Так как лингвистическая вероятность базы определяется подязыком, то, например, вероятность того, что словоформа «электрон» встречается в текстах по радиозлектронике, несравнимо больше, чем вероятность этой словоформы в художественной

¹ Н. Д. Андреев, Алгоритмы статистико-комбинаторного моделирования морфологии, синтаксиса, словообразования и семантики. — Материалы по математической лингвистике и машинному переводу II, Л., 1963, стр. 3—44.

² Х. А. Хольм, Выделение первого морфологического типа в эстонском языке на основе статистико-комбинаторного моделирования.

Х. А. Хольм, Выделение первого морфологического типа в эстонском языке на публицистических текстах. — Материалы по математической лингвистике и машинному переводу III (в печати).

³ Н. Д. Андреев, Л. Р. Зиндер, О понятиях речевого акта, речи, речевой вероятности и языка. — Вопросы языкознания, 1963, 3, стр. 15—21.

литературе. Таким образом, подъязык именно своим лексическим составом (т. е. лексической речевой вероятностью) определяет возможную последовательность выделения типов.

I эстонский тип, полученный на материале текстов по радиоэлектронике, характеризуется девятью ложными шагами. Такого числа ложных шагов в языках, исследованных методом статистико-комбинаторного моделирования, до сих пор не встречалось. Причина возникновения большого количества ложных шагов заключается в самой структуре эстонского языка и в соответствующем подъязыке, что ведет нас в область соотношения языка с подъязыками.

В эстонском языке видное место среди возможных путей словообразования занимает словосложение. В языке выработана своеобразная и богатая система словосложения⁴. Словосложение может принимать на себя сравнительно, например, с русским языком, следующие функции:

- 1) аффиксация,
- 2) словосочетание,
- 3) выражение нового понятия посредством новой словоформы.

Например:	<i>meri</i>	— море
	<i>meremees</i>	— моряк
	<i>mereäärne</i>	— приморский
	<i>merelaevandus</i>	— морское судоходство
	<i>merekool</i>	— морское училище
	<i>merevaik</i>	— янтарь
	<i>vihm</i>	— дождь
	<i>vihmamantel</i>	— дождевик
	<i>vihmapiisk</i>	— капля дождя, дождевая капля
	<i>vihmavari</i>	— зонт

В разных целях (И. В. Вески: целесообразность, И. Аавик: красота языка) языковеды стремились к замене сложных слов простыми. Все эти исследования велись в области различных подъязыков. Но методом первоначального создания терминологии является до сих пор словосложение.

В исследуемых текстах по радиоэлектронике терминология подъязыка, за редкими исключениями, создана методом словосложения. Поэтому возникают подобные нижеприведенному ряду слов: *kõrgeheliline*, *kõrgekvaliteediline*, *kõrgemaastmeline*, *kõrgetakistuslik*, *kõrgeväärtuslik*, *kõrglääbivusfiltrid*, *kõrgsagedusaste*, *kõrgeheliline*, *kõrgekvaliteediline*, *kõrgemaastmeline*, *kõrgetakis-*

⁴ J. V. Veski, Ebakohti eesti kirjakeele sõnavaras. — ENSV TA Toimetised, Ühiskonnateaduste seeria, 1956, 2, стр. 106—110.

tuslik, kõrgeväärtuslik, kõrglääbivusfiltrid, kõrgsagedusaste, kõrgsagedusfiltrid, kõrgsagedusliinid, kõrgsageduslik, kõrgsageduspinge, kõrgsagedustakistus, kõrgsagedustrafo, kõrgsagedustransformaator, kõrgsagedusvool, kõrgsagedusvõimendaja, kõrgsagedusvõimendus, kõrgsagedusvõimsus. (Пример взят из I ложного шага на радиоэлектронных текстах. Словоформы приведены в номинативе ед. или мн. числа). Ложный информант \bar{o}_{+2} возник в результате словообразования через словосложение. Общая часть у 18 различных по значению слов сильно повышает условную вероятность графем, попадающих в эту общую часть. Если же одна или несколько графем общей части данного ряда слов входят одновременно и в общую часть другого или нескольких других рядов слов, то ее условная вероятность повышается во столько же раз. Например:

võnkering, *võnkeenergia*, *võnkesagedus*, *võnkumisprotsess*, *võnkumisrežiim* — второй ряд.

põhigrupp, *põhisagedus*, *põhirühm*, *põhilevitüüp*, *põhiliik*, *põhivallem* — третий ряд,

võimsuskadu, *võimsustrafod*, *võimsustegur* — четвертый ряд и некоторые другие ряды.

Такой определенный метод словообразования в сочетании с речевой вероятностью, определяемой подъязыком, создает определенные предпосылки для позиционных вероятностей графем. Условные вероятности (условием является определенная позиция) графем отражают таким образом подязыковую речевую вероятность. С другой стороны, подязыковые условные вероятности отражают относительное расположение графем по позициям в языке. Об этом свидетельствует относительно пропорциональное распределение основной массы графем по позициям при различных подъязыках.

Таким образом, относительно сходные условные вероятности графем — это языковая тенденция распределения графем по позициям, в то время как относительно различающиеся условные вероятности графем — явление, характерное для данного подъязыка. (Таблицы 1, 2, 3.)

Таблица 1

Условные вероятности графем по позициям (радиоэлектронные тексты)

Графема	№ позиции ⁵								
		+1	+2	+3	+4	-4	-3	-2	-1
A		0,068	0,156	0,083	0,050	0,107	0,102	0,091	0,131
B		—	—	0,010	0,010	0,006	0,003	0,001	0,049
D		0,003	—	0,015	0,079	0,045	0,042	0,054	0,103
E		0,087	0,115	0,092	0,091	0,060	0,153	0,111	0,236
F		0,002	—	0,001	—	—	—	—	—
G		0,003	—	0,049	0,032	0,022	0,020	0,061	0,003
H		0,012	0,016	0,086	0,002	0,009	0,008	—	—
I		0,024	0,110	0,102	0,091	0,114	0,164	0,063	0,083
J		0,027	0,003	0,013	0,031	0,007	0,006	0,007	—
K		0,097	0,013	0,068	0,097	0,027	0,046	0,080	0,004
L		0,061	0,081	0,107	0,077	0,092	0,019	0,086	0,055
M		0,069	0,040	0,024	0,051	0,073	0,015	0,015	0,009
N		0,037	0,024	0,100	0,063	0,054	0,057	0,058	0,005
O		0,032	0,110	0,053	0,010	0,038	0,032	0,015	—
P		0,066	0,001	0,020	0,007	0,002	0,001	0,001	—
R		0,036	0,032	0,041	0,046	0,025	0,025	0,011	0,006
S		0,101	0,019	0,063	0,045	0,116	0,078	0,195	0,141
S		—	—	—	—	—	—	—	—
Z		—	—	0,001	—	—	—	—	—
T		0,109	0,004	0,035	0,133	0,087	0,089	0,046	0,138
U		0,001	0,068	0,037	0,069	0,058	0,092	0,081	0,021
V		0,140	0,004	0,009	0,032	0,040	0,043	0,024	0,011
O		—	0,079	0,005	—	0,007	0,001	—	—
A		0,002	0,072	0,009	—	0,007	0,001	—	—
O		—	0,009	0,008	—	—	—	—	—
U		0,027	0,044	0,001	—	—	0,004	—	—
—		0,002	—	—	—	—	—	—	0,005

Таблица 2

Условные вероятности графем по позициям (публицистические тексты)

Графема	№ позиции						
		+1	+2	+3	-3	-2	-1
A		0,064	0,216	0,089	0,137	0,093	0,142
B		0,006	0,001	0,013	0,008	0,002	0,033
C		0,000	0,000	—	0,000	0,001	—
D		0,009	0,005	0,018	0,036	0,067	0,140
E		0,038	0,111	0,085	0,122	0,097	0,211
F		0,004	0,001	0,001	—	0,000	—
G		0,003	0,002	0,030	0,017	0,050	0,001

⁵ +1 позицией считается первая графема с начала слова.

-1 позицией считается первая графема с конца слова.

H	0,024	0,017	0,071	0,016	0,005	0,001
I	0,019	0,091	0,137	0,140	0,092	0,115
J	0,027	0,001	0,023	0,006	0,011	—
K	0,145	0,006	0,025	0,041	0,075	0,006
L	0,061	0,037	0,099	0,045	0,088	0,045
M	0,069	0,014	0,019	0,024	0,038	0,007
N	0,033	0,017	0,102	0,063	0,045	0,009
O	0,029	0,090	0,051	0,028	0,012	0,002
P	0,099	0,004	0,017	0,005	0,001	0,000
R	0,040	0,048	0,057	0,030	0,011	0,006
S	0,089	0,026	0,060	0,107	0,134	0,135
S	—	0,000	0,000	—	—	—
Z	0,000	—	—	—	—	0,000
Z	—	0,000	0,000	—	—	—
T	0,117	0,015	0,040	0,092	0,084	0,100
U	0,007	0,080	0,056	0,069	0,069	0,038
V	0,082	0,006	0,015	0,031	0,013	0,005
W	0,000	—	0,000	—	—	—
O	0,006	0,071	0,001	0,001	0,001	—
А	0,002	0,069	0,008	0,002	0,001	—
О	0,002	0,015	0,024	0,002	0,000	0,000
У	0,023	0,022	0,005	0,001	0,001	—
X	—	—	0,000	—	—	—
Y	0,000	0,000	0,000	—	0,000	0,000
—	0,000	—	—	0,000	0,001	0,003

Таблица 3

Независимые вероятности графем

Номер графемы по убывающей величине	Графема	Вероятность графемы	Номер графемы по убывающей величине	Графема	Вероятность графемы
1.	Е	0,104	24.	.	0,008
2.	А	0,099	25.	Ї	0,006
3.	пробел	0,094	26.	О	0,002
4.	І	0,087	27.	- (дефис)	0,000
5.	Т	0,075	28.	(0,000
6.	S	0,074	29.	"	0,000
7.	L	0,058	30.)	0,000
8.	U	0,049	31.	—	0,000
9.	K	0,039	32.	цитата	0,000
10.	N	0,039	33.	Z	0,000
11.	D	0,038	34.	:	0,000
12.	O	0,038	35.	S	0,000
13.	M	0,031	36.).	0,000
14.	R	0,030	37.),	0,000
15.	V	0,026	38.	"	0,000
16.	G	0,020	39.	"	0,000
17.	J	0,014	40.	:	0,000
18.	H	0,014	41.).	0,000
19.	P	0,011	42.);	0,000
20.	O	0,011	43.	C	0,000
21.	А	0,010	44.	.(0,000
22.	,	0,008	45.	.)	0,000
23.	В	0,008	46.	"	0,000

Множество коррелятивных функций букв по убывающей величине (радиоэлектронные тексты)

1.	$O_{+2}=7,2$	23.	$T_{+1}=1,5$
2.	$H_{+3}=6,1$	24.	$S_{+1}=1,4$
3.	$V_{+1}=5,4$	25.	$L_{+2}=1,4$
4.	$O_{+2}=2,9$	26.	$U_{+4}=1,4$
5.	$N_{+3}=2,6$	27.	$I_{-4}=1,3$
6.	$S_{-2}=2,6$	28.	$A_{-1}=1,3$
7.	$K_{+1}=2,5$	29.	$L_{+4}=1,3$
8.	$K_{+4}=2,5$	30.	$I_{+2}=1,3$
9.	$M_{-4}=2,4$	31.	$T_{-3}=1,2$
10.	$E_{-1}=2,3$	32.	$A_{-4}=1,2$
11.	$D_{+4}=2,1$	33.	$T_{-4}=1,2$
12.	$I_{-3}=1,9$	34.	$I_{+3}=1,2$
13.	$U_{-3}=1,9$	35.	$E_{+2}=1,1$
14.	$S_{-1}=1,9$	36.	$E_{-2}=1,1$
15.	$T_{-1}=1,8$	37.	$A_{-3}=1,0$
16.	$L_{-3}=1,8$	38.	$I_{+4}=1,0$
17.	$T_{+4}=1,8$	39.	$E_{+4}=0,9$
18.	$K_{+3}=1,7$	40.	$S_{+3}=0,9$
19.	$A_{+2}=1,6$	41.	$E_{+3}=0,9$
20.	$S_{-4}=1,6$	42.	$A_{+3}=0,8$
21.	$L_{-4}=1,6$	43.	$E_{+1}=0,8$
22.	$E_{-3}=1,5$	44.	$E_{-4}=0,6$

Множество коррелятивных функций букв по убывающей величине (публицистические тексты)

1.	$P_{+1}=9,0$	17.	$I_{+3}=1,5$
2.	$H_{+3}=5,1$	18.	$S_{-3}=1,5$
3.	$D_{-1}=3,7$	19.	$U_{-2}=1,4$
4.	$K_{+1}=3,7$	20.	$A_{-3}=1,4$
5.	$V_{+1}=3,2$	21.	$A_{-1}=1,4$
6.	$N_{+3}=2,4$	22.	$I_{-1}=1,3$
7.	$A_{+2}=2,2$	23.	$S_{+1}=1,2$
8.	$E_{-1}=2,0$	24.	$T_{-3}=1,2$
9.	$S_{-1}=1,8$	25.	$E_{-3}=1,2$
10.	$S_{-2}=1,8$	26.	$T_{-2}=1,1$
11.	$L_{+3}=1,7$	27.	$E_{+2}=1,1$
12.	$T_{+1}=1,6$	28.	$A_{-2}=0,9$
13.	$I_{-3}=1,6$	29.	$E_{-2}=0,9$
14.	$L_{-2}=1,5$	30.	$A_{3+}=0,9$
15.	$K_{-2}=1,5$	31.	$E_{+3}=0,8$
16.	$L_{-2}=1,5$		

II. ЛОЖНЫЕ ШАГИ АЛГОРИТМА

Так как лингвистическое содержание первого ложного шага на радиоэлектронных текстах приведено выше, то для него дается его алгоритмический исход. Для всех остальных ложных шагов дается внеалгоритмическая лингвистическая интерпретация. Это значит, что лингвистическая интерпретация ложного шага представляет собой выявление лингвистических причин его возникновения.

1. Ложные шаги, полученные на радиоэлектронных текстах

На радиоэлектронных текстах мы получили 9 ложных шагов, что указывает на чрезвычайно развитое словообразование путем словосложения в эстонском языке, а в частности в подъязыке радиоэлектроники.

I ложный шаг

$$K\Phi \bar{O}_{+2} = 7,2$$

Условные вероятности +1 и +3 позиций при \bar{O}_{+2}

+1 позиция	+3 позиция
V=0,557	N=0,318
K=0,093	R=0,214
M=0,064	I=0,213
Ö=0,060	U=0,062
L=0,058	L=0,045
T=0,045	P=0,041
N=0,039	H=0,033
P=0,034	Õ=0,030
S=0,031	S=0,010
I=0,015	E=0,008
H=0,004	T=0,007
	K=0,007
	J=0,007
	V=0,005

Перепадом считается отношение условной вероятности информанта, который последним вошел в цепочку к УВ элемента, следующего за данным информантом в том вероятностном спектре, где этот информант был получен. Если перепад меньше 1,5, то рассматриваемый информант не включается в информантную цепочку, и она тем самым укорачивается на один элемент. Фактически в алгоритме предписывается делать проверки на

меру перепада начиная со второго информанта цепочки (если он разрешен). Предел наращиванию информантной цепочки кладется исходя из предположения, что «средняя длина словоизменительного компонента слова находится в определенной зависимости от средней длины всего слова в рассматриваемом языке».⁶

Мера перепада

$$1,5 \cdot 0,318 = 0,477$$

$$0,477 < 0,557,$$

Медиальное разбиение — это сечение заданного множества языковых элементов на 2 подмножества: верхнее подмножество, элементы которого обладают условной вероятностью более высокой, чем пороговая, и нижнее подмножество, элементы которого обладают условной вероятностью, равной пороговой или меньшей, чем пороговая. За порог принимается полусумма условных вероятностей экстремальных элементов, т. е. $\frac{y_B + y_{\bar{B}}}{2}$.

Всякое медиальное разбиение связано с определенным условием. При другом условии и порог разбиения и само разбиение изменяются.

Медиальное разбиение:

$$\frac{0,557 + 0,004}{2} = \frac{0,561}{2} = 0,281$$

и КФ для верхнего подмножества:

$$V_{+1} = \frac{0,557}{0,026} = 21,5$$

$$N_{+3} = \frac{0,318}{0,039} = 8,1$$

Таким образом, новым информантным аффиксом является

$$V_{+1} \ O_{+2}$$

Предложим формулу $x = D^2 : \Phi$,

где Φ — средняя длина фразы (в словах),

D — средняя длина слова (в буквах),

x — искомая длина словоизменительного компонента (тоже в буквах).

Информантная цепочка должна соответственно состоять не больше чем из антье от x элементов (если x получается меньше еди-

⁶ Н. А. Андреев, Алгоритмы статистико-комбинаторного моделирования морфологии, синтаксиса, словообразования и семантики. — Материалы по математической лингвистике и машинному переводу II, Л., 1963, стр. 7.

ницы, то к антье прибавляется единица).⁷ Информантную цепочку, сформированную с данным ограничением, назовем первым информантным аффиксом и обозначим его самого через I_1 , а его длину через d . Информантными аффиксами, следующими за первым, назовем замену данного информантного аффикса другими сочетаниями букв при той же базе.

Образует информантное подмножество.

«Сочтем членимыми словами $Y_ч$ моделируемого языка те словоформы, длина которых не меньше, чем антье от $\frac{2}{3} D+1$, где D — средняя длина слова. Произведем информантное разбиение множества всех членимых словоформ на два подмножества: подмножество слов $Y_n 1$, содержащих информантный аффикс I_1 , и подмножество слов $Y_n 1$, не содержащих его. Иначе говоря, подставим $Y_ч$ в виде $Y_ч = Y_n 1 + Y_n \bar{1}$. Подмножество $Y_n 1$ назовем информантным подмножеством; слова, попавшие в информантное подмножество, будем рассматривать как состоящие из информантного аффикса I_1 и базы B_1 (т. е. сочтем базой всякий остаток от выделения I_1 , даже если он не является основой в лингвистическом смысле): $Y_n 1 = B_1 + I_1$ ».⁸

Условные вероятности + 3 позиции при $V_{+1} O_{+2}$

$$I = 0,363$$

$$R = 0,387$$

$$N = 0,288$$

$$E = 0,006$$

$$T = 0,006$$

Проверка на меру перепада:

$$1,5 \cdot 0,337 = 0,5055$$

$$0,5055 > 0,363$$

В формуле $\frac{M_1 - M_2}{k \cdot M_1}$ M_1 — количество баз первого вида, т. е. базы, которые сочетаются с нашим предполагаемым информантом после шлейфа; M_2 — количество баз второго вида, т. е. базы, которые могут сочетаться как с предполагаемым информантом, так и с другими сочетаниями букв, которые могут выступать на месте информанта в той же позиции; k — коэффициент недостаточности текста, т. е. число, которое находится между средней длиной слова D и единицей и зависит от длины текста.

⁷ Н. Д. Андреев, Алгоритмы статистико-комбинаторного моделирования морфологии, синтаксиса, словообразования и семантики. — Материалы по математической лингвистике и машинному переводу II, Л., 1963, стр. 8.

⁸ Там же, стр. 8.

$D > k > 1$, где $k = D^{\frac{1}{1+0,02 \log v}}$; v — объем текста в словоформах.

«Если мера редукции не превышает некоторой пороговой величины, то можно считать, что множества B_1 и B_2 связаны грамматически, если же налицо превышение заданного порога, то нужно предполагать отсутствие грамматической связи между ними (т. е. либо словообразовательную связь, либо никакой). Лингвистические соображения, основанные на знании того, что словоизменение, как правило, регулярно, а словообразование гораздо менее регулярно, заставляют считать, что пороговая величина должна быть равной $1 : D$ (но не более 0,5), где D — средняя длина слова. Мету редукции, более низкую, чем порог редукции $1 : D$, будем называть субпороговой; мету редукции, равную $1 : D$ или более высокую, — суперпороговой».⁹

Текстовая неполнота может быть использована для того, чтобы заранее исключать из работы слова, недостаточно частые. Шлейф — это подмножество членимых словоформ, принадлежащих к словам с малой частотностью, куда войдут, в частности, все слова с потенциально полной, но актуально неполной парадигмой. Практически шлейф охватывает словоформы, которые встретились в информантном подмножестве один раз и не встретились при замене данной информантной цепочки другими сочетаниями букв при той же базе.

Баз первого вида после шлейфа 114.

Баз второго вида 2.

Базы второго вида:

võttes—jättes—mõttes
võrele—järele—perele

Мера редукции:

$$\frac{M_1 - M_2}{k \cdot M_1} = \frac{114 - 2}{6,65 \cdot 114} = \frac{112}{758} = 0,147; k = 6,65$$

Мера редукции для эстонского языка равна

$$\frac{1}{D} = \frac{1}{8} = 0,125$$

$$0,147 > 0,125$$

Следовательно, редукция суперпороговая и шаг является ложным.

II ложный шаг:

$$K\Phi H_{+3} = 6,1$$

⁹ Н. Д. Андреев, ук. соч., стр. 10.

Информантное подмножество состоит в основном из слов, общей частью которых являются словоформы *vahe*, *põhi*, а также и *tühi*, *lühi*, *tihe*, *täht*, *vähe*, *maht*, *juht* и основы: *näht-*, *lähe-*, *mähi-*, *laht-* и некоторые другие. Например: *vahe* (*vahemik*, *vaheline*, *vahetu*, *vahendid*), *põhi* (*põhinev*, *põhinema*) и некоторые другие.

Сложные слова от *vahe*: *vahema*, *vahepealne*, *vahekord*, *vahelduvkomponent*, *vahelduvpinge*, *vahelduvvool*, *vahelduvvooluring*, *vahelduvvoolumasin*, *vahelduvvooluvõrk*, *vahelduvvooluallikas*, *vahelduvvoolugeneraatorid*, *vahelduvvooluenergiavõrk*, *vahelduvvoolumootor*, *vahelduvvoolu-valgustusvõrk*.

Сложные слова от *põhi*: *põhimõte*, *põhimõtteline*, *põhitüübiks*, *põhinõuetele*, *põhisuurus*, *põhisagedus*, *põhimagnetväli*, *põhivool*, *põhiühik*.

Условные вероятности +2 и +4 позиции при H_{3+} :

+2 позиция		+4 позиция	
A=0,455	E=0,039	E=0,404	K=0,017
Ä=0,226	I=0,033	T=0,230	H=0,008
V=0,080	O=0,031	I=0,210	N=0,006
Õ=0,079	H=0,001	J=0,059	V=0,003
Ü=0,056		A=0,042	M=0,002
		U=0,019	

Мера перепада ниже 1,5

$$(A_{+2}=0,455, E_{+4}=0,404)$$

Баз второго вида в тексте нет, т. е. $M_2=0$ и шаг алгоритма автоматически считается ложным.

III ложный шаг

$$K\Phi V_{+1}=5,4$$

Информантное подмножество состоит в основном из сложных слов, общей частью которых являются преимущественно слова *vastu*, *vahe*, *vool*, *väli*, *vähe*.

Например, сложные слова от *voolu*: *vooluallikas*, *voolujuhe*, *vooluring*, *voolutugevus*, *vooluenergia*, *voolumuutus*, *voolusuund*, *voolutarvidus*. Сложные слова от *vastu*: *vastuvõtt*, *vastuvõtja*, *vastuvõtjaam*, *vastuvõtuantenn*, *vastuvõtetud*, *vastuvõtu protsess*, *vastumõju*, *vastupidi*, *vastutusrikas* (формально сюда входят и слова с *vastutus*).

Условные вероятности графем в +2 позиции при V_{+1} :

A=0,298	O=0,204	I=0,026
Ä=0,276	Õ=0,181	E=0,015

Информантная цепочка замыкается, поскольку мера перепада ниже 1,5 ($A_{+2}=0,298$, $\bar{A}_{+2}=0,276$).

Баз второго вида в исследуемом тексте 3:

vähemal — lähemal
 väheneb — läheneb
 vähenedes — lähenedes

После шлейфа баз первого вида 165. Таким образом:

$$\frac{M_1 - M_2}{\text{к. } M_1} = \frac{165 - 3}{6,50 \cdot 165} = \frac{162}{1072,5} = 0,151.$$

IV ложный шаг

$$\text{КФ } O_{+2} = 2,9.$$

Информантное подмножество состоит главным образом из сложных слов, общей частью которых являются словоформы *vool* ($1/3$ из всех словоформ, входящих в информантное подмножество), *koogmus*, а также и *pool*, *kogu*, *soojus* и нек. др.

Условные вероятности +1 и +3 позиции при O_{+2} :

+1 позиция		+3 позиция	
V=0,319	I=0,022	O=0,594	M=0,024
K=0,208	O=0,020	S=0,088	H=0,022
P=0,170	L=0,014	R=0,053	K=0,012
T=0,058	R=0,009	L=0,049	P=0,006
J=0,055	H=0,006	T=0,048	B=0,006
S=0,046	B=0,003	G=0,034	E=0,004
M=0,042	F=0,001	N=0,030	D=0,002
N=0,027		I=0,028	

Проверка на меру перепада ($V_{+2}=0,319$, $O_{+3}=0,594$).

$$1,5 \cdot 0,319 = 0,4785$$

$$0,4785 < 0,594$$

Следовательно, алгоритм идет следующим путем:
 Медиальное разбиение:

$$\frac{0,594 + 0,001}{2} = 0,298$$

КФ $O_{+3} = 15,6$
 КФ $V_{+1} = 12,2$.

Таким образом получается информантная цепочка $I = O_{+2}$
 O_{+3} .

В информантное подмножество входят слова, имеющие общую часть

voolu (482 случая из 1000).

Например:

voolluallikas, vooluliik, voolupulss, vooluimpulss, voolugraafik, vooluring, vooluhulk, vooluthedus, voolutarbija, voolukandja, voolujuht, voolutõuge, voolusuund, vooluahel, voolujuhe, vooluharu, voolusagedus, voolufaas, vooluvõrk, voolutrafo, voolumähis.

Условные вероятности графем в +1 и +4 позиции при $O_{+2}O_{+3}$:

+1 позиция

V=0,482

K=0,133

J=0,127

P=0,085

S=0,056

M=0,040

T=0,029

I=0,021

L=0,016

H=0,007

N=0,004

+4 позиция

L=0,574

N=0,152

R=0,081

D=0,070

S=0,056

J=0,033

M=0,012

V=0,008

K=0,006

B=0,005

T=0,003

Мера перепада меньше 1,5 ($L_{+4}=0,574$, $V_{+1}=0,482$). Следовательно, информантная цепочка замыкается.

Базы второго вида:

toodetakse — täidetakse, toidetakse

koosneb — kaasneb

После шлейфа осталось 120 баз первого вида. Итак:

$$\frac{M_1 - M_2}{k \cdot M_1} = \frac{120 - 1}{6,50 : 120} = \frac{118}{780} = 0,151.$$

V ложный шаг

КФ $N_{+3}=2,6$.

В данном случае информантное подмножество образовано не через словосложение. Хотя и существует общая часть нескольких сложных слов типа ringhääling, ringtee, ringikujuline, ringitaoline, ringpermeaablus, ringjoon, ringsagedus (всего 1,2% информантного подмножества), но статистически она почти не влияет на выделение N_{+3} . Выделение $I=N_{+3}$ в данном случае больше связано с конкретным объектом описываемого предмета, чем с терминологическим словообразованием языка. Парадигма

слова võnkering занимает 26,2% всего информантного подмножества, парадигма слова ringe 7,9%, несклоняемая словоформа ainult 5,1%, парадигма слова kondensaator 11,0%. Таким образом, перечисленные выше слова занимают 50,2% всего информантного подмножества.

Условные вероятности графем +2 и +4 позиции при N_{+3} :

+2 позиция	+4 позиция	
I=0,353	K=0,268	I=0,022
Õ=0,314	D=0,228	P=0,013
O=0,149	G=0,208	A=0,013
U=0,064	N=0,125	S=0,012
A=0,061	E=0,059	U=0,004
E=0,041	T=0,046	V=0,001
A=0,010		Õ=0,001
N=0,007		
Ü=0,001		

Мера перепада ($I_{+2}=0,353$ и $Õ_{+2}=0,314$) не позволяет расширить информантную цепочку.

Баз второго вида — 1:

kantakse — kaetakse

Баз первого вида после шлейфа — 115.

$$\frac{M_1 - M_2}{k \cdot M_1} = \frac{115 - 1}{5,49 \cdot 115} = \frac{114}{892,5} = 0,127.$$

VI ложный шаг

КФ $S_{-2}=2,6$.

Информантное подмножество с S_{-2} образуется на основе структуры формообразования языка. S_{-2} входит в состав ниже следующих падежей склоняемых слов:

единственное число	множественное числс
падежи генитив (keemilise)	партитив (keemilisi)
партитив (keemilist)	иллатив (keemilistesse)
иллатив (keemilisse)	элатив (keemilistest)
элатив (keemilisest)	

Кроме того, S_{-2} входит в состав грамматического признака глагола (ksee-olevik, т. е. настоящее время имперсонала). (Таблица 4)

В информантном подмножестве S_{-2} встретилось в следующем соотношении:

Таблица 4

Форма склоняемого слова, в состав которой входит S_{-2}	Встречаемость в информантном множестве в %
генитив ед. ч.	36,4%
элатив ед. ч.	22,0%
партитив ед. ч.	12,9%
настоящее время имперсонала	9,7%
элатив мн. ч.	6,4%
партитив мн. ч.	5,8%
иллатив ед. ч.	3,4%
иллатив мн. ч.	0,4%

S_{-2} в информантном подмножестве в 1⁰/о случаев не входит в состав формообразования.

Формообразовательные компоненты в данном случае напоминают словосложение из нескольких слов, а общая их графема — общую часть сложных слов. Этим и объясняется сравнительно высокая КФ S_{-2} , так как условная вероятность общей графемы возрастает в зависимости от того, во сколько различных формообразовательных компонентов она входит.

Условные вероятности —1 и —3 позиции при S_{-2} :

—1 позиция

E=0,495
T=0,437
I=0,061
A=0,006
V=0,001

—3 позиция

I=0,282
U=0,268
E=0,188
K=0,098
A=0,059
S=0,039
V=0,020
T=0,010

L=0,009
G=0,008
R=0,007
D=0,005
M=0,002
N=0,002
O=0,002
Ö=0,001

Так как мера перепада не позволяет нам расширить информантную цепочку ($E_{-1}=0,496$, $T_{-1}=0,437$), то перейдем к базам второго вида.

Баз второго вида — 24.

1. tekkimise — tekkimine
2. analoogilise — analoogiline
3. kuumenemise — kuumenemine
4. amprise — amprine
5. omaaegse — omaaegne
6. elektrilise — elektriline
7. esialgse — esialgne
8. esimese — esimene

9. eraldumise — eraldumine
10. normaalse ← normaalne
11. niisuguse — niisugune
12. positiivse — positiivne
13. kindlasuunalise — kindlasuunaline
14. konstantse — konstantne
15. suuruse — suurune
16. kokkuleppelise — kokkuleppeline
17. siinuselise — siinusedine
18. sellise — selline
19. viimase — viimane
20. soolamise — soolamine
21. vastupidise — vastupidine
22. -voldise — -voldine
23. kasutamise — kasutamine
24. keemilise — keemiline — keemilisse — keemiliste

$$\frac{M_1 - M_2}{k \cdot M_1} = \frac{156 - 24}{6,40 \cdot 156} = \frac{132}{998,4} = 0,132.$$

Хотя редукция и в данном случае суперпороговая, все же теоретически возможно при увеличении текста получить субпороговую редукцию и шаг может оказаться истинным. Об этом говорит частичное выделение парадигмы слова

keemilise — keemiline — keemilisse — keemiliste

и потенциальная возможность выделения подобной парадигмы у большинства баз второго вида. Отдельная графема е прикрепилась бы к информантным аффиксам при проверке на унитарную членимость.

«После того как тип сформирован, необходимо произвести проверку его аффиксов и основ на унитарную членимость. Если все аффиксы парадигмы содержат одну и ту же букву (группу букв) в позиции примыкания к основе, то эта буква переводится в состав основы, и проверка на УЧ заканчивается. Если такой общности у аффиксов нет, то рассматривается последняя буква у всех основ; при наличии перепада, меньшего трех, проверка на УЧ закончена. Если же он больше трех, то из основ, содержащих наиболее вероятную для этой концевой позиции букву, она выводится в гипотетическое расширение аффиксов парадигмы. Затем снова рассматривается последняя буква — теперь уже в усеченных основах; гипотетическое расширение аффиксов продолжается до получения перепада, меньшего трех. Максимально усеченные основы после этого объявляются базами гипотетического вида, и по ним строится гипотетическое дополнительное звено информантной лестницы. Если оно пусто, то проверка на УЧ считается ничего не давшей, и тип остается без изменений; если же оно не пусто, то гипотетическое расширение

считается действительным, и достройка информантной лестницы возобновляется.

Лингвистический смысл проверки на УЧ — в выявлении агглютинативности».¹⁰

А, следовательно, базы типа

keemilisest — keemiliselst

соединились бы с предыдущими базами, и их аффиксы включились бы в общую парадигму. Мы получили бы

keemili- -se, -ne, -sse, -ste, -sest, -self.

VII ложный шаг

КФ $K_{+1}=2,5$.

Информантное подмножество состоит в основном из парадигмы слов

kõver, kõrg, kasu, kao (kadu), kiir, kuju
kesk, kondensaator, koond.

Хотя встречаются ряды слов типа:

kõrgsagedus, kõrgsageduspoolid, kõrgsagedusraud, kõrgsagedus-
vooluringid, kõrgsagedustakistuslikud, kõrgepinge, kõrgsagedus-
takistused, kõrgsageduskaod,

они статистически не влияют на состав информантного подмножества.

Условные вероятности графем в +2 позиции при K_{+1} :

A=0,236	Ä=0,024
O=0,233	Ü=0,023
U=0,154	L=0,023
E=0,146	R=0,010
Õ=0,084	V=0,003
I=0,064	

Мера перепада меньше 1,5 ($A_{+2}=0,236$, $O_{+2}=0,233$), и информантная цепочка замыкается.

Множество баз второго вида оказалось одноэлементным:

kaheks — vaheks

После шлейфа оказалось 138 баз первого вида. Пользуясь формулой меры редукции, мы получаем

$$\frac{M_1 - M_2}{k \cdot M_1} = \frac{138 - 1}{6,51 \cdot 138} = \frac{137}{898,3} = 0,152$$

$$0,152 > 0,125$$

¹⁰ Н. Д. Андреев, ук. соч., стр. 16.

VIII ложный шаг

КФ $K_{+4}=2,5$

Информантное подмножество характеризуется парадигматическими рядами сложных и несложных слов, общей частью которых является *elektr-* (занимают 42,0% всего информантного подмножества). Напр. *elektromagnetiline, elektronlamp, elektrolüüs, elektropaljundajad, elektrotehnika, elektromotoorne, elektrolüütcondensaator, elektrostaatiline;*

elektrilaengud, elektrijuhid, elektrivool, elektrijuht, elektrivooluring, elektrivõimsus, elektriväli, elektrihulk, elektrivooluallikas, elektripotentsiaal, elektrilaeng,

а также и *elektron,*
elektrood;
и т. д.

elektri,
elektriline;

Условные вероятности графем +3 и +5 позиций при K_{+4}

+3 позиция		+5 позиция	
E=0,476	N=0,027	T=0,514	K=0,026
I=0,146	U=0,025	U=0,141	L=0,020
A=0,090	Ü=0,020	S=0,076	M=0,018
K=0,084	R=0,007	E=0,072	V=0,007
S=0,064	D=0,001	I=0,049	J=0,005
T=0,030	Ö=0,001	O=0,037	N=0,003
H=0,029	O=0,001	A=0,029	K=0,002
			Õ=0,001

Информантная цепочка замыкается ($T_{+5}=0,514$, $E_{+3}=0,475$).

Базы второго вида:

1. *liikumise — liitumise*
2. *liikuvad — liiguvad, liituvad.*

После шлейфа оказалось 162 базы первого вида.

$$\frac{M_1 - M_2}{k \cdot M_1} = \frac{162 - 2}{6,50 \cdot 162} = \frac{160}{1053} = 0,151.$$

IX ложный шаг

КФ $M_{-4}=2,4$.

Информантное подмножество с $I=M_{-4}$ образуется на основе парадигмы суффикса отглагольного существительного;

Напр.

arvutamine, arvutamise, arvutamist
varjestamine, varjestamise, varjestamist
kasutamine, kasutamise, kasutamist

В информантном множестве суффикс -т- остается неизменяемой частью, меняются база и парадигма. Суффикс -т- занимает 48,0% всего информантного подмножества. (Таблица 5.)

Таблица 5

Парадигма, в которую входит суффикс -т-	% информантного подмножества, занимаемые данным суффиксом
-mine	20,6%
-mise	15,1%
-mist	10,1%
-mis-	1,3%
-misi	0,9%
	Итого 48,0%

Кроме указанного суффикса, M_{-4} является в информантном подмножестве составной частью еще некоторых других формообразующих суффиксов. Например:

-maks	} части формы сравнительной степени
-malt	
-male	
-mate	
-maid	
-mata	сложный суффикс инфинитива

и некоторые другие.

Условные вероятности графем в -5 и -3 позиции при M_{-4}

-5 позиция		-3 позиция
A=0,296	K=0,023	I=0,572
E=0,176	R=0,022	A=0,210
I=0,150	S=0,015	E=0,098
U=0,147	P=0,013	U=0,079
T=0,064	G=0,008	S=0,024
D=0,033	L=0,006	O=0,011
M=0,024	Ö=0,001	E=0,003
O=0,024	Ä=0,001	B=0,002
		N=0,001

Мера перепада ($I_{-3}=0,572$, $A_{-5}=0,296$) разрешает расширять информантную цепочку.

Итак:
$$\frac{0,572+0,001}{2} \approx 0,287$$

КФ $I_{-3}=6,57$ КФ $A_{-5}=2,99$

Образуем информантное подмножество с $I=M_4 I_3$. (Таблица 6.)

Таблица 6

Парадигма, в которую входит суффикс -m-	% %, информантного подмножества, приходящиеся на данный суффикс
-mine	37,8%
-mise	31,5%
-mist	17,7%
-mis-	1,9%
	Итого: 88,9%

Условные вероятности графем -5 и -2 позиции при $M_4 I_3$:

-5 позиция		-2 позиция
A=0,310	O=0,021	S=0,534
U=0,248	D=0,017	N=0,387
E=0,149	M=0,012	K=0,035
I=0,136	G=0,011	T=0,015
K=0,032	L=0,005	L=0,010
T=0,031	Ö=0,003	D=0,009
S=0,023	R=0,002	G=0,008
		U=0,001
		V=0,001

Информантная цепочка замыкается ($S_{-2}=0,534$, $N_{-2}=0,387$).
Множество баз второго вида оказалось одноэлементным:

arvutamist — arvutatust.

Баз первого вида после шлейфа 156.

$$\frac{M_1 - M_2}{k \cdot M_1} = \frac{156 - 1}{6,40 \cdot 156} = \frac{155}{998,4} = 0,154.$$

2. Ложные шаги, полученные на публицистических текстах

На публицистических текстах две первые по величине коррелятивные функции P_{+1} и H_{+3} дают нам два ложных шага. При P_{+1} информантная цепочка расширяется за счет словообразования. Фактически мы должны проверить и K_{+1} , который стоит под четвертым номером во множестве КФ, поскольку нет формального основания ставить его под четвертым номером.

Л о ж н ы й ш а г

КФ $P_{+1}=9,1$.

Информантное подмножество словообразовательными рядами не характеризуется. Словоформы с самыми частотными словообразовательными частями rea- и palju- занимают соответственно 7,0 и 4,3% информантного подмножества.

Условные вероятности +2 позиции при P_{+1} :

1. A=0,223	7. U=0,075
2. R=0,137	8. Ä=0,071
3. I=0,118	9. L=0,045
4. E=0,111	10. Ü=0,028
5. O=0,097	11. Ö=0,011
6. Ö=0,081	12. S=0,001

Информантная цепочка не замыкается. Медиальное разбиение:

$$\frac{0,223+0,001}{2} = 0,112.$$

КФ для верхнего подмножества:

$$\begin{aligned} \text{КФ } A &= 2,3 \\ \text{КФ } R &= 4,6 \\ \text{КФ } I &= 1,4 \end{aligned}$$

P_{+2} выделяется из-за очень низкой независимой вероятности. (Независимые вероятности: A=0,099; R=0,030).

Новое информантное подмножество с $I=P_{+1}R_{+2}$ сложных слов почти не содержит (исключения: presidendikandidaat, gruunikashallid, proletariaaditeematika). Баз второго вида нет, т. е. $M_2=0$ и шаг автоматически считается ложным.

II ложный шаг

$$\text{КФ } H_{+3} = 5,1.$$

Информантное подмножество характеризуется парадигмой слов:

tehas;	rahvas;	vähe;
lahendama;	kahe-;	juhi- (juht-);

Сложных слов сравнительно немного.

Базы второго вида:	lahendamine	— laiendamine
	lahendada	— laiendada

Мера редукции:

$$\frac{M_1 - M_2}{k \cdot M_1} = \frac{234 - 2}{6,80 \cdot 234} = 0,146.$$

III ложный шаг

$$K\Phi K_{+1} = 3,7$$

Информантное подмножество характеризуется в основном парадигмой слов

kolhoos;	keemia;
kateeder;	kauba-

Условные вероятности +2 позиции при K_{+1} :

1. A=0,380	7. Ü=0,062
2. O=0,320	8. Ä=0,050
3. U=0,250	9. V=0,016
4. E=0,144	10. R=0,016
5. I=0,100	11. L=0,010
6. Ö=0,086	12. S=0,002

Информантная цепочка замыкается.

Базы второго вида:

1. kulud — tulud
2. kasuks — tasuks
3. kahel — vahel (2)

Проверим меру редукции для предполагаемого информанта, чередующегося с предполагаемым информантом T.

Мера редукции:

$$\frac{M_1 - M_2}{k \cdot M_1} = \frac{570 - 2}{6,80 - 570} = \frac{568}{3548,8} = 0,148.$$

Ясно, что мера редукции для $I_2 = V$ (базы kahel — vahel (2)) будет еще выше, поскольку I_2 имеет только одну базу, а при уменьшении количества баз мера редукции повышается.

3. Причины возникновения ложных шагов

Обобщая, таким образом, лингвистические причины возникновения ложных шагов на материале эстонского языка, можно сделать некоторые общие выводы:

1. Количество ложных шагов зависит от подъязыка.
2. Количество ложных шагов является определенной характеристикой подъязыка.
3. Общее количество ложных шагов зависит от места слово-сложения в системе словообразования языка. Такие ложные шаги можно называть грамматическими.
4. Ложные шаги могут быть вызваны своеобразием аффиксации языка. В таком случае несколько различных по значению аффиксов имеют общий элемент, позиционная условная вероят-

ность которого возрастает в зависимости от того, во сколько различных аффиксов он входит.

5. Наступает момент, когда ложные шаги связаны не со словообразованием или словоизменением, а с конкретным объектом описываемого подъязыком предмета, т. е. с речевой вероятностью словоформы.

Такие ложные шаги можно называть предметными, подъязыковыми, т. е. неграмматическими, поскольку они связаны непосредственно только с объектом описания.

6. По причинам возникновения ложные шаги бывают:

I. Грамматические $\left\{ \begin{array}{l} 1. \text{ Словообразовательные} \\ 2. \text{ Формообразовательные} \end{array} \right.$

II. Неграмматические

7. Грамматические ложные шаги характеризуют одновременно язык и подъязык. Неграмматические ложные шаги образованы речевой вероятностью подъязыка и характеризуют только подъязык.

III. СОДЕРЖАНИЕ ИНФОРМАНТНОГО ПОДМНОЖЕСТВА И ПОРЯДОК ВЫДЕЛЕНИЯ ИНФОРМАНТНЫХ АФФИКСОВ ПРИ ИСТИННОМ ШАГЕ АЛГОРИТМА

1. Истинный шаг алгоритма, полученный на радиоэлектронных текстах

Лингвистическая интерпретация информантного подмножества

$$КФ E_{-1} = 2,3$$

Информантное подмножество при E_{-1} состоит из баз со следующими формальными аффиксами: -se, -te, -de, -le, -ne, -me. (Таблица 7).

При остальных 3,1% информантного подмножества графемы в -2 позиции при E_{-1} формального грамматического значения в себе не носят и имеет смысл разобрать словоформы с информантом относительно к падежам. E_{-1} как часть формы падежей:

номинатив ед. ч.:	alumiiniumjuhe (1) kuulike (2) lühike (2) peenike (2) ristlõige (1) süsihape (2)
генитив ед. ч.:	ristlõike (3) väävelhappe (2)

Слова, формально оканчивающиеся в номинативе и генитиве ед. ч. на -e

генитив мн. ч.

партитив мн. ч.

alasurve (3)
 Karlsruhe (1)
 Punaarmee (1)
 talituspinge (1)
 varasalve (1)
 ülesurve (2)
 eeskujulike (1)
 isetegevuslike (1)
 ulatuslike (1)
 vedelike (1)
 paberitükke (1)
 ulatuslikke (1)
 vajalikke (1)

Таблица 7

Формальный аффикс, частью которого является И	Грамматические формы, в которые входит предполагаемый информант	% информантного подмножества, занимаемый каждой из грамматических форм	Общий % информантного подмножества, занимаемый данным формальным аффиксом
-se	1. Генитив ед. ч. (kauguse, keemilise) 2. Настоящее время имперсонала (tehakse) 3. Иллатив ед. ч. (ajastusse, tegevusse) 4. Партитив мн. ч. (plastmasse)	24,2% 3,7% 2,3% 0,1%	31,4%
-te	1. Генитив мн. ч. (positiivsete, negatiivsete) 2. Генитив ед. ч. (ülevaate) 3. Партитив мн. ч. (elemente)	17,1% 0,3% 0,1%	17,5%
-de	1. Генитив мн. ч. (voolude) 2. Генитив ед. ч. (tuhande) 3. Часть основ несклоняемых слов (juurde)	15,1% 0,2% 0,1	15,4%

-ne	1. Номинатив ед. ч. (keemiline, konstantne)	14,9%	15,3%
	2. Слова, формально оканчивающиеся в номинативе и генитиве ед. ч. на -ne. (lihtaine, liitaine)	0,2%	
	3. Партитив мн. ч. (reaktsioone)	0,1%	
	4. Генитив ед. ч. (neljakümne)	0,1%	
-le	1. Аллатив ед. ч. (maksimaalsele, omadusele)	9,7%	11,2%
	2. Аллатив мн. ч. (takistustele, harudele)	1,3%	
	3. Партитив мн. ч. (molekule, detaile)	0,2	
-me	1. Генитив ед. ч. (juhe — juhtme tase — taseme)	3,6	6,1%
	2. Окончание 1 лица мн. ч. глагола (настоящего и прошедшего времени)	2,4%	
	3. Партитив мн. ч. (telegramme)	0,1	

Итого: 96,9%

Соответственно с этим распределяются условные вероятности графем в —2 позиции при E_{-1} .

Условные вероятности графем в —2 позиции при E_{-1} :

T=0,175	K=0,016
S=0,314	V=0,006
D=0,154	P=0,004
N=0,153	G=0,002
L=0,112	H=0,002
M=0,061	E=0,001

Информантная цепочка удлиняется, поскольку

$$1,5 \cdot 0,175 = 0,2625$$

$$0,2625 < 0,314$$

Проведем медиальное разбиение:

$$\frac{0,314+0,001}{2} = 0,158$$

Вычисляем коррелятивные функции:

$$K\Phi S = 4,2$$

$$K\Phi T = 2,3$$

Образуем новой информантной цепочкой $S_{-2} E_{-1}$ информантное подмножество. Полученное информантное подмножество состоит главным образом из баз, стоящих в генитиве ед. ч. (Таблица 8.)

Таблица 8

Грамматический состав информантного подмножества

% % в информантном подмножестве	Информант с предыдущей графемой	Грамматические формы, в которые входит предполгаемый информант	% информантного подмножества, занимаемый данной грамматической формой
27,9%	-use	часть падежной формы генитива ед. ч.	69,4%
27,0%	-ise		
4,0%	-vse		
2,5%	-lse		
2,3%	-ase		
2,0%	-rse		
1,4%	-ese		
0,9%	-gse		
0,7%	-dse		
0,6%	-tse		
0,1%	-kse		
19,9%	-kse	настоящее время имперсонала	19,9%
10,6%	-sse	1. -sse падежное окончание иллатива ед. ч.	8,9%
		2. -desse, -tesse падежное окончание иллатива в соединении с признаком множественности	1,5%
		3. -e часть формы партитива мн. ч.	0,2%
0,1%	-tse	Словообразующий суффикс глаголов	0,1

Базы с парадигмой И₁

1. aktiivtakistu-	И ₁ ,	s(1) ¹¹ , sega(1), sest(1), si(1), ste(1)
2. alalisvoolu- takistu-	И ₁ ,	sega(1), seks(1), sest(1)
3. alumi-	И ₁ ,	sel(1)
4. arendami-	И ₁ ,	seks(1), sest(1)
5. arvuli-	И ₁ ,	ne(1), selt(9), ste(1)
6. aurami-	И ₁ ,	ne(1)
7. avastu-	И ₁ ,	stega(1)
8. dielektrili-	И ₁ ,	ne(2)
9. elektrijuhtivu-	И ₁ ,	s(1)
10. elektrili-	И ₁ ,	ne(20), sed(7), seks(1), sele(1), selt(9), ses(3), si(2), sesse(1), st(7), ste(7), stelt(1)
11. elektromag- netili-	И ₁ ,	ne(1), sed(1), seks(1), sel(1), si(2), st(1), ste(7)
12. elektromo- toor-	И ₁ ,	ne(23), seks(3), sest(2), set(4), sete(5)
13. eneses-	И ₁ ,	t(1)
14. eraldumi-	И ₁ ,	ne(2), sena(1)
15. erinimeli-	И ₁ ,	sed(4), selt(3), st(2), ste(3)
16. eritakistu-	И ₁ ,	s(2), sed(1), sega(2), seks(1)
17. esialg-	И ₁ ,	ne(1), sega(1), seid(1), sesse(1)
18. esialgses-	И ₁ ,	∅(1), ed(3), eks(1), el(2), ena(3), est(1), tel(1)
19. esime-	И ₁ ,	ne(3), sed(3), seks(1), sel(2), sele(3), sena(3), sest(1), stel(2)
20. esimes-	И ₁ ,	ele(3)
21. ferromag- netili-	И ₁ ,	ne(1), sed(1), sest(1), si(1), ste(1), stel(1)
22. horisontaal-	И ₁ ,	selt(1), sest(1)
23. hõõrumi-	И ₁ ,	sega(1), sel(4)
24. igapäeva-	И ₁ ,	sesse(1)
25. intensiivsu-	И ₁ ,	s(1)
26. isetühjenemi-	И ₁ ,	ne(2)
27. jooni-	И ₁ ,	s(5), sel(10), selt(1)
28. juhtivu-	И ₁ ,	sed(1), seks(1), ste(1)
29. järgmi-	И ₁ ,	ne(4), sed(1), ses(1), selt(1), sesse(1), stest(1)

¹¹ Цифра в скобках указывает на встречаемость данного информантного аффикса с данной базой.

30. järjestiku-	H ₁ ,	ne(1), seks(1), sest(1), st(1), ste(20), stel(1), stele(2)
31. kaasaeg-	H ₁ ,	ses(2)
32. kadumi-	H ₁ ,	sel(2)
33. kasutami-	H ₁ ,	ne(2), seks(2), sel(8), sele(2), seta(1), st(2)
34. katkemi-	H ₁ ,	sega(1)
35. kaugu-	H ₁ ,	s(1), sel(1), sele(1), sest(1)
36. keemili-	H ₁ ,	ne(14), sed(1), sel(3), selt(4), ses(1), sest(2), si(6), sse(2), st(3), ste(5), steks(1), stes(1), stest(1)
37. keemilis-	H ₁ ,	ed(1), el(3), elt(4), es(1), i(6), t(3), te(5), teks(1), tes(1), test(1)
38. keeruli-	H ₁ ,	ne(1)
39. keraamili-	H ₁ ,	ne(1), sed(1), si(2), sse(1)
40. keraamilis-	H ₁ ,	e(1), ed(1), i(2)
41. kiiru-	H ₁ ,	sega(1)
42. kinni-	H ₁ ,	ne(1), ses(1)
43. kodumai-	H ₁ ,	ste(1)
44. kogumahtuvu-	H ₁ ,	s(8), seks(1), sena(1)
45. kogutakistu-	H ₁ ,	s(1), sega(2), seks(1), st(1)
46. kokkuleppeli-	H ₁ ,	ne(3), sed(1), sele(2), selt(3)
47. konstant-	H ₁ ,	ne(2), set(2)
48. konvent- sionaal-	H ₁ ,	sele(1), set(2)
49. koormami-	H ₁ ,	sega(1), sel(7), st(1)
50. koormatavu-	H ₁ ,	s(6), sega(1)
51. koormu-	H ₁ ,	s(2), sed(1), sega(1), seta(1)
52. koormus- takistu-	H ₁ ,	sele(1)
53. koostis-	H ₁ ,	es(3), est(3)
54. korruti-	H ₁ ,	s(1), sega(1), sena(1)
55. kuumenemi-	H ₁ ,	ne(1)
56. kuumutami-	H ₁ ,	sel(1)
57. kõrvali-	H ₁ ,	ne(2), sest(2), st(1), ste(1)
58. küllastu-	H ₁ ,	seni(1)
59. laadimi-	H ₁ ,	ne(2), seks(2), sel(11)
60. lahti-	H ₁ ,	ses(1)
61. lahu-	H ₁ ,	sega(1), ses(2), sesse(3), sest(6), stamisel(3)
62. lahuses-	H ₁ ,	ø(2), t(6)
63. leeli-	H ₁ ,	ste(3)
64. leidmi-	H ₁ ,	sega(2), seks(3)
65. liig-	H ₁ ,	ne(1), selt(2), sete(1)

66. liikumi-	H ₁ ,	ne(5), sega(1), seks(3), sel(4), sele(2), sena(1), sest(1), st(4)
67. lõunapoolu-	H ₁ ,	s(4), seks(1), sele(2)
68. läbitavu-	H ₁ ,	sega(9), st(1)
69. lühi-	H ₁ ,	ajalise(1), ajaliste(1), kesel(1), s(1), stamine(1), stamisega (1), ste(2)
70. lühiajali-	H ₁ ,	ste(1)
71. lülitami-	H ₁ ,	seks(2)
72. lülitu-	H ₁ ,	ses(2)
73. magneetimi-	H ₁ ,	sel(2)
74. magnetili-	H ₁ ,	ne(15), sed(6), seks(1), selt(3), sest(1), si(4), sse(1), st(1), ste(4), stele(1)
75. magnetilis-	H ₁ ,	e(1), ed(6), eks(1), elt(3), est(1), i(4), t(1), te(4), tele(1)
76. mahtuvu-	H ₁ ,	s(26), sed(2), sega(10), seks (1), sele(1), sest(4), si(1), st(6), ste(2), stega(3)
77. maksimaal-	H ₁ ,	ne(17), sed(1), sele(2), selt(4), set(1), setest(1)
78. matemaatili-	H ₁ ,	sed(1), selt(4), ste(1), stes(1)
79. mehaanili-	H ₁ ,	ne(3), st(3)
80. milli-	H _L ,	ne(1), sed(2), seks(1), selt(1), sest(1), ste(4)
81. minimaal-	H ₁ ,	ne(1), sele(1)
82. mitmekülg-	H ₁ ,	sem(1)
83. mitmesugu-	H ₁ ,	seid(1), ste(1), stel(1)
84. muundumi-	H ₁ ,	ne(1)
85. muutumi-	H ₁ ,	ne(1), sel(1), sele(1), st(2)
86. möötimi-	H ₁ ,	ne(1), seks(2)
87. mähi-	H ₁ ,	sega(1), sesse(1), sest(1), st(2)
88. negatiiv-	H ₁ ,	ne(18), sed(6), seid(2), seks(4), sel(3), sele(10), selt(7), semaks(1), semat(1), set(6), sete(2), seteks(1), sus(1)
89. neutraal-	H ₁ ,	sel(1), ses(2)
90. niisugu-	H ₁ ,	ne(2), sed(1), seid(1), sel(1), ses(3), ste(1), stest(1)
91. nimetu-	H ₁ ,	sed(1), st(1)
92. normaal-	H ₁ ,	ne(2), sel(1), selt(1), ses(1), sest(3), setes(1)
93. nõrgenemi-	H ₁ ,	sel(1), sele(1), st(1)
94. nähtu-	H ₁ ,	s(6), sed(1), sel(1), ses(1), si(2), st(2), ste(3)

95. omaaeg-	H₁ ,	ne(1)
96. omadu-	H₁ ,	sed(3), sega(2), seks(1), sele (1), si(4), ste(1), stega(4), stelt(1), stes(1)
97. omavaheli-	H₁ ,	ne(1), ste(1)
98. osatähtsu-	H₁ ,	ste(1)
99. paksu-	H₁ ,	ne(1)
100. paralleel- lülitu-	H₁ ,	ses(2)
101. pikku-	H₁ ,	ne(1), s(2), sega(1), sel(1) sest(2)
102. pingeli-	H₁ ,	ne(1)
103. plaatide- vaheli-	H₁ ,	ne(2)
104. polaarsu-	H₁ ,	s(1)
105. poolu-	H₁ ,	s(6), sed(3), sega(4), sele(3), selt(2), st(1), ste(1), stesse (1), stest(1),
106. positiiv-	H₁ ,	ne(14), sed(8), seid(1), seks(5), sel(6), se(6), selt(14), ses (1), sest(1), set(10), sete(1), seteks(1), sus(1)
107. praktili-	H₁ ,	seks(1), selt(8), ste(1)
108. primaarmähi-	H₁ ,	s(1), sesse(1), sest(1), st(2)
109. pulgakujuli-	H₁ ,	si(1)
110. puudumi-	H₁ ,	sel(2), st(1)
111. põhjapoolu-	H₁ ,	s(2), sega(1), seks(1), sete(1), selt(2)
112. raadiotööstu-	H₁ ,	s(1), ses(1)
113. ringikujuli-	H₁ ,	ne(1), st(1)
114. ruumili-	H₁ ,	st(1)
115. sagedu-	H₁ ,	s(2), sest(1), st(3)
116. samanimeli-	H₁ ,	ne(1), sed(4), selt(1), si(1), st(1), ste(1)
117. samatüübili-	H₁ ,	ste(1), stest(1)
118. seadi-	H₁ ,	s(1), st(1)
119. seadu-	H_i ,	s(2), sest(1), st(3)
120. segalülitus-	H₁ ,	es(5)
121. sekundaar- mähises-	H₁ ,	t(1)
122. selli-	H₁ ,	ne(3)
123. siinuseli-	H₁ ,	ne(1), st(1)
124. sisetakistu-	H₁ ,	s(14), sed(1), sega(1), seks(1), sel(3), sest(2), st(1), ste(1)
125. sisselülitami-	H₁ ,	sel(3)
126. sooju-	H₁ ,	seks(8), sena(2), st(4)
127. suhteli-	H₁ ,	ne(5), sed(1), selt(7)

128. suuru-	H ₁ ,	ne(6), s(12), sed(4), sega(9), sele(2), selt(2), sest(6), si(2), st(7), ste(5), stega(2), steks(1), stest(2)
129. sõltuvu-	H ₁ ,	ses(1)
130. takistu-	H ₁ ,	s(33), sed(2), sega(7), seks(1), sel(2), sele(2), ses(1), sest (1), si(4), st(9), ste(6), stele (2), stest(1).
131. tarvit-	H ₁ ,	amiseks(1), usel(1)
132. tavalis-	H ₁ ,	selt(5), st(1), ste(1), stes(1), stest(1)
133. teadu-	H ₁ ,	salal(1), ste(1)
134. tegevus-	H ₁ ,	ulatus(1)
135. tehnilis-	H ₁ ,	seks(1), selt(1), ste(2)
136. tekkimis-	H ₁ ,	ne(4), seks(1), sel(1), sele(1), st(1)
137. telegrafeeri-	H ₁ ,	st(1)
138. teoreetilis-	H ₁ ,	ne(1)
139. teostamis-	H ₁ ,	seks(1)
140. tera-	H ₁ ,	ses(1)
141. tihedus-	H ₁ ,	s(1), sele(1)
142. tolleaeg-	H ₁ ,	sed(1)
143. traatmähi-	H ₁ ,	ses(1)
144. tugevus-	H ₁ ,	s(5), sega(9), st(4)
145. tulemus-	H ₁ ,	sed(1), seks(1), sena(1), si(1)
146. tähtsus-	H ₁ ,	s(1), st(3)
147. vahekaug-	H ₁ ,	s(3), st(1)
148. vahelis-	H ₁ ,	ne(4), ses(2), sest(1)
149. valitsus-	H ₁ ,	s(1)
150. vastasnimelis-	H ₁ ,	ne(2), st(1)
151. vastassuunalis-	H ₁ ,	sed(1)
152. vastastikus-	H ₁ ,	ne(3), sed(1), sel(1)
153. vastupidi-	H ₁ ,	ne(4), sed(2), sel(1), selt(1)
154. viimasis-	H ₁ ,	ne(3), sel(1), sest(1), stes(1)
155. voolamis-	H ₁ ,	sest(1), st(1)
156. võimsus-	H ₁ ,	s(15), sed(2), sega(5), seks(1), sest(1), si(1), st(4), ste(2)
157. võnkumis-	H ₁ ,	sest(1), st(1)
158. võrdlemis-	H ₁ ,	seks(1)
159. võrd-	H ₁ ,	lemise(1), lemiseks(1), ne(34), sed(3), seks(1), selt(1), set (1), sete(5), setes(1)
160. väike-	H ₁ ,	sel(1), sele(2), ses(1), si(1), ste(1), stel(1), stest(1)
161. väli-	H ₁ ,	ne(2), selt(3), ses(1), sesse(1)

162. väljatihedu-	H₁,	s(7), sega(1), st(1)
163. väljatugevu-	H₁,	s(4), sed(1), sega(1), seks(1), sest(2), st(4), ste(6), stel(1)
164. värvu-	H₁,	st(1)
165. väärtu-	H₁,	s(7), sed(1), sega(2), seks(1), seni(3), sest(2), si(1)
166. üheampri-	H₁,	ne(1), st(1)
167. ühendami-	H₁,	ne(2), seks(1), sel(4), st(1)
168. ühendus-	H₁,	ses(1)
169. ühtla-	H₁,	selt(3)
170. ükstei-	H₁,	sega(1), st(1)
171. ülekoormami-	H₁,	st(1)
172. üleliiduli-	H₁,	selt(1)
173. ülemäära-	H₁,	ne(1)
174. ümberpaik- nemi-	H₁,	st(1)
175. -ampri-	H₁,	∅(1), ne(1), st(1)
176. -kord-	H₁,	seks(1), selt(1)
177. -voldi-	H₁,	ne(4), ses(1), st(1)

Базы I вида, полученные из информантного подмножества после шлейфа:

1. asetatakse (2)	25. pendesse (2)
2. antakse (4)	26. nimetatakse (32)
3. anumasse (4)	27. nimetatakse (32)
4. enesse (2)	28. peenikese (2)
5. esialgse (3)	29. primaarse (3)
6. happese (2)	30. Päikese (2)
7. hüstereesisilmuse (3)	31. rakendamise (3)
8. igasse (3)	32. rakendatakse (2)
9. juhatakse (2)	33. raskuse (4)
10. kantakse (2)	34. saadakse (3)
11. kasutatakse (63)	35. samasse (2)
12. kilovatise (4)	36. sellesse (8)
13. kokkupuutumise (2)	37. teadlase (3)
14. kondensaatorisse (8)	38. tekitatakse (2)
15. kontrollitakse (2)	39. tähistatakse (5)
16. kujutise (2)	40. vaadeldakse (2)
17. laetakse (2)	41. valguse (3)
18. laetuse (2)	42. valmistatakse (6)
19. loetakse (4)	43. vastassuunalise (2)
20. lülitatakse (3)	44. villase (2)
21. millesse (2)	45. vähesuse (2)
22. moodustatakse (2)	46. värvilise (2)
23. mõõdetakse (10)	47. õhukese (2)
24. märgitakse (8)	48. ühendatakse (2)

49. -lülituse (4)

50. -oomise (3)

51. -ruutmeetrise (2)

Таким образом, баз первого вида 228 (177+51).

Условные вероятности информантных аффиксов с базами, имеющими предполагаемый информантный аффикс -se:

1. — ne	=0,01092	36. — ena	=0,00012
2. — s	=0,00923	37. — ele	=0,00012
3. — st	=0,00453	38. — le	=0,00012
4. — selt	=0,00424	39. — esse	=0,00012
5. — sel	=0,00395	40. — ega	=0,00012
6. — sega	=0,00334	41. — stamisel	=0,00012
7. — sed	=0,00325	42. — tamisel	=0,00012
8. — seks	=0,00268	43. — e	=0,00012
9. — ste	=0,00268	44. — vahe	=0,00008
10. — sest	=0,00218	45. — eks	=0,00008
11. — sele	=0,00191	46. — stelt	=0,00008
12. — si	=0,00139	47. — seta	=0,00008
13. — ses	=0,00139	48. — samat	=0,00008
14. — set	=0,00107	49. — sus	=0,00008
15. — ed	=0,00064	50. — sem	=0,00004
16. — sete	=0,00056	51. — tel	=0,00004
17. — es	=0,00052	52. — ni	=0,00004
18. — i	=0,00052	53. — tes	=0,00004
19. — est	=0,00048	54. — teks	=0,00004
20. — t	=0,00044	55. — test	=0,00004
21. — sena	=0,00040	56. — se	=0,00004
22. — stest	=0,00040	57. — ksel	=0,00004
23. — te	=0,00036	58. — tele	=0,00004
24. — stel	=0,00032	59. — stest	=0,00004
25. — sesse	=0,00028	60. — stes	=0,00004
26. — stega	=0,00024	61. — semaks	=0,00004
27. — seid	=0,00024	62. — stesse	=0,00004
28. — stes	=0,00024	63. — ksel	=0,00004
29. — sse	=0,00020	64. — salae	=0,00004
30. — el	=0,00020	65. — ulates	=0,00004
31. — steks	=0,00016	66. — amiseks	=0,00004
32. — lt	=0,00016	67. — stamine	=0,00004
33. — stele	=0,00016	68. — stamisega	=0,00004
34. — elt	=0,00012	69. — skiiirus	=0,00004
35. — seni	=0,00012	70. — Ø морфема	=0,00004

Медиальное разбиение:

$$\frac{0,01092+0,00004}{2} = 0,00548$$

КФ для верхнего подмножества:

1. КФ-NE=2,69

2. КФ-S=0,12

Таким образом, предполагаемым вторым информантным аффиксом является — NE.

Базы с парадигмой И₁, И₂

1. -ampri-	И ₁ , И ₂ ,	
2. arvuli-	И ₁ , И ₂ ,	
3. aurami-	И ₁ , И ₂ ,	
4. dielektrili-	И ₁ , И ₂ ,	
5. elektrili-	И ₁ , И ₂ ,	sed(7), seks(1), sele(1), sele(1), selt(9), ses(3), si(2), sse(1), st(7), ste(7), stelt(1)
6. elektromagnetili-	И ₁ , И ₂ ,	sed(1), seks(1), sel(1), si(2), st(1), ste(7)
7. elektromotoor-	И ₁ , И ₂ ,	seks(3), sest(2), set(4), sete(5)
8. eraldumi-	И ₁ , И ₂ ,	sena(1)
9. esialg-	И ₁ , И ₂ ,	sega(1), seid(1), sesse(1)
10. esime-	И ₁ , И ₂ ,	sed(3), seks(1), sel(2), sele(3), sena(3), sest(1), stel(2)
11. ferromagnetili-	И ₁ , И ₂ ,	sed(1), sest(1), si(1), ste(1), stel(1)
12. isetühjenemi-	И ₁ , И ₂	
13. järgmi-	И ₁ , И ₂ ,	ne(4), sed(1), ses(1), selt(1), sse(1), stest(1)
14. järjestiku-	И ₁ , И ₂ ,	seks(1), sest(1), st(1), ste(20), stel(1), stele(2)
15. kasutami-	И ₁ , И ₂ ,	seks(2), sel(8), sele(2), seta(1), st(2)
16. keemili-	И ₁ , И ₂ ,	sed(1), sel(3), selt(4), ses(1), sest(2), si(6), sse(2), st(3), steks(1), stes(1), stest(1)
17. keeruli-	И ₁ , И ₂	
18. keraamili-	И ₁ , И ₂ ,	sed(1), si(2), sse(1)
19. kinni-	И ₁ , И ₂ ,	ses(1)
20. kokkuleppeli-	И ₁ , И ₂ ,	sed(1), sele(2), selt(3)
21. konstant-	И ₁ , И ₂ ,	set(2)
22. kuumenemi-	И ₁ , И ₂	
23. kõrvali-	И ₁ , И ₂ ,	sest(2), st(1), ste(1)
24. laadimi-	И ₁ , И ₂ ,	seks(2), sel(11)
25. liig-	И ₁ , И ₂ ,	selt(2), sete(1)
26. liikumi-	И ₁ , И ₂ ,	sega(1), seks(3), sel(4), sele (2), sena(1), sest(1), st(4)

27. magnetili-	H₁, H₂ ,	sed(6), seks(1), selt(3), sest(1), si(4), sse(1), st(1), ste(4), stele(1)
28. maksimaal-	H₁, H₂ ,	sed(1), sele(2), selt(4), set(1), setest(1)
29. mehhaanili-	H₁, H₂ ,	st(3)
30. milli-	H₁, H₂ ,	sed(2), seks(1), selt(1), sest(1), ste(4)
31. minimaal-	H₁, H₂ ,	sele(1)
32. muundumi-	H₁, H₂	
33. muutumi-	H₁, H₂ ,	sel(1), sele(1), st(2)
34. mõõtmisi-	H₁, H₂ ,	seks(2)
35. omavaheli-	H₁, H₂ ,	ste(1)
36. negatiiv-	H₁, H₂ ,	sed(6), seid(2), seks(4), sel(3), sele(10), selt(7), semaks(1), semat(1), set(6), sete(2), seteks(1), sus(1)
37. niisugu-	H₁, H₂ ,	sed(1), seid(1), sel(1), ses(3), ste(1), stest(1)
38. normaal-	H₁, H₂ ,	sel(1), selt(1), ses(1), sest(3), setes(1)
39. ringikujuli-	H₁, H₂ ,	sed(1), st(1)
40. paksu-	H₁, H₂	
41. pikku-	H₁, H₂ ,	s(2), sega(1), sel(1), sest(2)
42. pingeli-	H₁, H₂	
43. plaatidevaheli-	H₁, H₂	
44. positiiv-	H₁, H₂ ,	sed(8), seid(1), seks(5), sel(6), sele(6), selt(14), ses(1), sest(1), set(10), sete(1), seteks(1), sus(1)
45. samanimeli-	H₁, H₂ ,	sed(4), selt(1), si(ī), st(1), ste(1)
46. selli-	H₁, H₂	
47. siinusele-	H₁, H₂ ,	st(1)
48. suhteli-	H₁, H₂ ,	sed(1), selt(7)
49. suuru-	H₁, H₂ ,	s(12), sed(4), sega(9), sele(2), selt(2), sest(6), si(2), st(7), ste(5), stega(2), steks(1), stest(2)
50. tekkimi-	H₁, H₂ ,	seks(1), sel(1), sele(1), st(1)
51. teoreetili-	H₁, H₂	
52. vaheli-	H₁, H₂ ,	ses(2), sest(1)
53. vastanimeli-	H₁, H₂ ,	st(1)
54. vastastiku-	H₁, H₂ ,	sed(1), sel(1)
55. vastupidi-	H₁, H₂ ,	sed(2), sel(1), selt(1)
56. viima-	H₁, H₂ ,	sel(1), sest(1), stes(1)

57. võrd-	И₁, И₂ ,	sed(3), seks(1), selt(1), set(1), sete(5), setes(1)
58. väli-	И₁, И₂ ,	selt(3), ses(1), sesse(1)
59. üheampri-	И₁, И₂ ,	st(1)
60. ühendami-	И₁, И₂ ,	seks(1), sel(4), st(1)
61. ülemäära-	И₁, И₂	
62. -voldi-	И₁, И₂ ,	ses(1), st(1)

Мера редукции:

$$\frac{M_1 - M_2}{k \cdot M_1} = \frac{228 - 62}{6,50 \cdot 228} = 0,112$$

Редукция субпороговая, следовательно, предполагаемые информантные аффиксы оказываются истинными. Вычисляем условные вероятности остальных предполагаемых информантных аффиксов с базами, имеющими информантные аффиксы И₁ и И₂. После медиального разбиения вычисляем КФ для элементов верхнего подмножества.

Условные вероятности информантных аффиксов с базами, имеющими И₁, И₂:

1. -selt	=0,00268	17. -stel	=0,00020
2. -ste	=0,00214	18. -seid	=0,00020
3. -sed	=0,00206	19. -sesse	=0,00016
4. -sel	=0,00206	20. -stele	=0,00012
5. -st	=0,00171	21. -steks	=0,00012
6. -sele	=0,00127	22. -stes	=0,00008
7. -seks	=0,00123	23. -setes	=0,00008
8. -set	=0,00096	24. -seteks	=0,00008
9. -sest	=0,00080	25. -sus	=0,00008
10. -si	=0,00080	26. -semat	=0,00008
11. -sete	=0,00064	27. -stega	=0,00008
12. -ses	=0,00060	28. -stelt	=0,00008
13. -s	=0,00056	29. -seta	=0,00004
14. -sega	=0,00048	30. -skiiirus	=0,00004
15. -stest	=0,00024	31. -semaks	=0,00004
16. -sse	=0,00020	32. -sena	=0,00004

Медиальное разбиение

$$\frac{0,00268 + 0,00004}{2}$$

КФ для элементов верхнего подмножества:

$$1. \text{ КФ - SELT} = \frac{0,00268}{0,074 \cdot 0,104 \cdot 0,058 \cdot 0,075} = 80,08$$

$$2. \text{КФ} - \text{STE} = \frac{0,00214}{0,074 \cdot 0,075 \cdot 0,104} = 3,708$$

$$3. \text{КФ} - \text{SED} = \frac{0,00206}{0,074 \cdot 0,104 \cdot 0,038} = 7,045$$

$$4. \text{КФ} - \text{SEL} = \frac{0,00206}{0,074 \cdot 0,104 \cdot 0,058} = 4,616$$

$$5. \text{КФ} - \text{ST} = \frac{0,00171}{0,074 \cdot 0,075} = 30,81$$

Таким образом, предполагаемый $I_3 = -\text{SELT}$.

Базы с парадигмой I_1, I_2, I_3

1. elektrili-	$I_1, I_2, I_3,$	sed(7), seks(1), sele(1), ses(3), si(2), sse(1), st(7), ste(7), stelt(1)
2. järgmi-	$I_1, I_2, I_3,$	sed(1), ses(1), sse(1), stest(1)
3. keemili-	$I_1, I_2, I_3,$	sed(1), sel(3), ses(1), sest(2), si(6), sse(2), st(3), steks(1), stes(1), stest(1)
4. kokkuleppeli-	$I_1, I_2, I_3,$	sed(1), sele(2)
5. liig-	$I_1, I_2, I_3,$	sete(1)
6. magnetili-	$I_1, I_2, I_3,$	sed(6), seks(1), sest(1), si(4), sse(1), st(1), ste(4), stele(1)
7. maksimaal-	$I_1, I_2, I_3,$	sed(1), sele(2), set(1), setest(1)
8. milli-	$I_1, I_2, I_3,$	sed(2), seks(1), sest(1), ste(4)
9. negatiiv-	$I_1, I_2, I_3,$	sed(6), seid(2), seks(4), sel(3), sele(10), semaks(1), semat(1), set(6), sete(2), seteks(1), sus(1)
10. normaal-	$I_1, I_2, I_3,$	sel(1), ses(1), sest(3), setes(1)
11. positiiv-	$I_1, I_2, I_3,$	sed(8), seid(1), seks(5), sel(6), sele(6), ses(1), sest(1), set(10), sete(1), seteks(1), sus(1)
12. samanimeli-	$I_1, I_2, I_3,$	sed(4), si(1), st(1), ste(1)
13. suhteli-	$I_1, I_2, I_3,$	sed(1)
14. suuru-	$I_1, I_2, I_3,$	s(12), sed(4), sega(9), sele(2), sest(6), si(2), st(7), ste(5), stega(2), steks(1), stest(2)

15. vastupidi-	И₁, И₂, И₃ ,	sed(2), selt(1)
16. võrd-	И₁, И₂, И₃ ,	sed(3), seks(1), set(1), sete (5), setes(1)
17. väli-	И₁, И₂, И₃ ,	ses(1), sesse(1)

Мера редукции для И₃:

$$\frac{M_1 - M_2}{k \cdot M_1} = \frac{62 - 17}{6,50 \cdot 62} = \frac{45}{403} \approx 0,112.$$

Следовательно, И₃ = -SELT.

Условные вероятности информантных аффиксов с базами, имеющими И₁, И₂, И₃:

1. -sed	=0,00171	16. -seid	=0,00012
2. -ste	=0,00107	17. -seteks	=0,00012
3. -sele	=0,00094	18. -stest	=0,00008
4. -st	=0,00076	19. -setes	=0,00008
5. -set	=0,00072	20. -sus	=0,00008
6. -seks	=0,00052	21. -semat	=0,00008
7. -s	=0,00048	21. -semat	=0,00008
8. -si	=0,00048	22. -stega	=0,00004
9. -sega	=0,00036	23. -stek	=0,00004
10. -sest	=0,00032	24. -stes	=0,00004
11. -sel	=0,00056	25. -stele	=0,00004
12. -ses	=0,00016	26. -sesse	=0,00004
13. -sete	=0,00016	27. -semaks	=0,00004
14. -sse	=0,00016	28. -sesse	=0,00004
15. -stest	=0,00012	29. -stelt	=0,00004

$$\frac{0,00171 + 0,00004}{2} \approx 0,00088$$

КФ для верхнего подмножества:

1. КФ — SED	$\frac{0,00171}{0,074 \cdot 0,104 \cdot 0,038} = 5,848$
2. КФ — STE	$\frac{0,00107}{0,074 \cdot 0,075 \cdot 0,104} = 1,897$
3. КФ — SELE	$\frac{0,00094}{0,074 \cdot 0,104 \cdot 0,058 \cdot 0,104} = 28,08$

Предполагаемый И₄ = -SELE.

Базы с парадигмой И₁, И₂, И₃, И₄

1. elektrili-	И₁, И₂, И₃, И₄ ,	sed(7), seks(1), ses(3), si(2), sse(1), st(7), ste(7), stelt(1)
---------------	---	--

2. kokku- leppeli-	$\mathbf{И}_1, \mathbf{И}_2, \mathbf{И}_3, \mathbf{И}_4,$	sed(1)
3. maksimaal-	$\mathbf{И}_1, \mathbf{И}_2, \mathbf{И}_3, \mathbf{И}_4,$	sed(1), set(1), setest(1)
4. negatiiv-	$\mathbf{И}_1, \mathbf{И}_2, \mathbf{И}_3, \mathbf{И}_4,$	sed(6), seid(2), seks(4), sel(3), semaks(1), semat(1), set(6), sete(2), seteks(1), sus(1)
5. positiiv-	$\mathbf{И}_1, \mathbf{И}_2, \mathbf{И}_3, \mathbf{И}_4,$	sed(8), seid(1), seks(5), sel(6), ses(1), sest(1), set(10), sete(1), seteks(1), sus(1)
6. suuru-	$\mathbf{И}_1, \mathbf{И}_2, \mathbf{И}_3, \mathbf{И}_4,$	s(12), sed(4), sega(9), sest(6), si(2), st(7), ste(5), stega(2), steks(1), stest(2)

Мера редукции для $\mathbf{И}_4$:

$$\frac{M_1 - M_2}{k \cdot M_1} = \frac{17 - 6}{6,50 \cdot 17} = \frac{11}{182} = 0,061.$$

Таким образом, $\mathbf{И}_4 = \text{-SELE}$.

Высчитаем условные вероятности информантных аффиксов с базами, имеющими $\mathbf{И}_1, \mathbf{И}_2, \mathbf{И}_3, \mathbf{И}_4$, а также и КФ для элементов верхнего подмножества (после медиального разбиения).

1. -sed	= 0,00107	12. -seid	= 0,00012
2. -set	= 0,00068	13. -sete	= 0,00012
3. -st	= 0,00056	14. -seteks	= 0,00008
4. -ste	= 0,00048	15. -sus	= 0,00008
5. -s	= 0,00048	16. -semat	= 0,00008
6. -seks	= 0,00040	17. -stega	= 0,00008
7. -sel	= 0,00032	18. -semaks	= 0,00004
8. -sega	= 0,00032	19. -sest	= 0,00004
9. -ses	= 0,00016	20. -steks	= 0,00004
10. -si	= 0,00016	21. -sse	= 0,00004
11. -stest	= 0,00012	22. -stelt	= 0,00004

Медиальное разбиение:

$$\frac{0,00107 + 0,00004}{2} = 0,00056$$

КФ для верхнего подмножества:

$$1. \text{ КФ—SED} = \frac{0,00107}{0,074 \cdot 0,104 \cdot 0,038} = 3,66$$

$$2. \text{ КФ—SET} = \frac{0,00068}{0,074 \cdot 0,104 \cdot 0,075} = 1,18$$

Базы с парадигмой И₁, И₂, И₃, И₄, И₅

1. elektrili-	И ₁ , И ₂ , И ₃ , И ₄ , И ₅ ,	seks(1), ses(3), si(2), ssé (1), st(7), ste(7), stelt(1)
2. maksimaal-	И ₁ , И ₂ , И ₃ , И ₄ , И ₅ ,	set(1), setest(1)
3. negatiiv-	И ₁ , И ₂ , И ₃ , И ₄ , И ₅ ,	seid(2), seks(4), sel(3), semaks(1), semat(1), set (6), sete(2), seteks(1), sus(1)
4. positiiv-	И ₁ , И ₂ , И ₃ , И ₄ , И ₅ ,	seid(1), seks(5), sel(6), ses (1), sest(1), set(10), sete (1), seteks(1), sus(1)
5. suugu-	И ₁ , И ₂ , И ₃ , И ₄ , И ₅ ,	s(12), sega(9), sest(6), si(2), st(7), ste(5), stega(2), steks(1), stest(2)

Мера редукции для И₅:

$$\frac{M_1 - M_2}{k \cdot M_1} = \frac{6 - 5}{6,50 \cdot 6} = \frac{1}{39} = 0,026$$

Следовательно, И₅ = -SED.

Условные вероятности информантных аффиксов с базами, имеющими И₁, И₂, И₃, И₄, И₅:

1. -set	=0,00064	12. -sus	=0,00008
2. -st	=0,00056	13. -stest	=0,00008
3. -ste	=0,00048	14. -steks	=0,00008
4. -s	=0,00048	15. -semat	=0,00008
6. -seks	=0,00040	16. -stega	=0,00008
7. -sel	=0,00032	17. -semaks	=0,00004
8. -sega	=0,00032	18. -sest	=0,00004
9. -ses	=0,00016	19. -steks	=0,00004
10. -si	=0,00016	20. -sse	=0,00004
11. -seid	=0,00012	21. -stelt	=0,00004

Медиальное разбиение:

$$\frac{0,00064 + 0,00004}{2} = 0,00034$$

КФ для верхнего подмножества:

$$1. \text{ КФ—SET} = \frac{0,00064}{0,074 \cdot 0,104 \cdot 0,075} = 1,109$$

$$2. \text{ КФ—ST} = \frac{0,00056}{0,074 \cdot 0,075} = 0,1009$$

$$3. \text{КФ—STE} = \frac{0,00048}{0,074 \cdot 0,075 \cdot 0,104} = 0,8315$$

$$4. \text{КФ—S} = \frac{0,00048}{0,074} = 0,006486$$

$$5. \text{КФ—SEKS} = \frac{0,00040}{0,074 \cdot 0,104 \cdot 0,039 \cdot 0,074} = 18,01$$

Предполагаемый $\mathbf{И}_6 = \text{-SEKS}$.

Базы с парадигмой $\mathbf{И}_1, \mathbf{И}_2, \mathbf{И}_3, \mathbf{И}_4, \mathbf{И}_5, \mathbf{И}_6$

- | | | |
|--------------|---|--|
| 1. elektri- | $\mathbf{И}_1, \mathbf{И}_2, \mathbf{И}_3, \mathbf{И}_4, \mathbf{И}_5, \mathbf{И}_6,$ | ses(3), si(2), sse(1), st(7), ste(7), stelt(1) |
| 2. negatiiv- | $\mathbf{И}_1, \mathbf{И}_2, \mathbf{И}_3, \mathbf{И}_4, \mathbf{И}_5, \mathbf{И}_6,$ | seid(2), sel(3), semaks(1), semat(1), set(6), sete(2), seteks(1), sus(1) |
| 3. positiiv- | $\mathbf{И}_1, \mathbf{И}_2, \mathbf{И}_3, \mathbf{И}_4, \mathbf{И}_5, \mathbf{И}_6,$ | seid(1), sel(6), ses(1), sest(1), set(10), sete(1), seteks(1), sus(1) |

Мера редукции для предполагаемого $\mathbf{И}_6$:

$$\frac{M_1 - M_2}{K \cdot M_1} = \frac{5 - 3}{6,50 \cdot 5} = \frac{2}{32,5} = 0,061$$

Таким образом $\mathbf{И}_6 = \text{-SEKS}$.

Условные вероятности информантных аффиксов с базами, имеющими $\mathbf{И}_1, \mathbf{И}_2, \mathbf{И}_3, \mathbf{И}_4, \mathbf{И}_5, \mathbf{И}_6$:

- | | | | |
|----------|-----------|-------------|-----------|
| 1. -set | = 0,00064 | 8. -seteks | = 0,00008 |
| 2. -sel | = 0,00036 | 9. -sus | = 0,00008 |
| 3. -st | = 0,00028 | 10. -semat | = 0,00008 |
| 4. -ste | = 0,00028 | 11. -si | = 0,00008 |
| 5. -ses | = 0,00016 | 12. -semaks | = 0,00004 |
| 6. -sete | = 0,00016 | 13. -sest | = 0,00004 |
| 7. -seid | = 0,00012 | 14. -sse | = 0,00004 |
| | | 15. -stelt | = 0,00004 |

Медиальное разбиение:

$$\frac{0,00064 + 0,00004}{2} = \frac{0,00068}{2} = 0,00034$$

КФ для верхнего подмножества:

$$1. \text{-set} = \frac{0,00064}{0,074 \cdot 0,104 \cdot 0,058} = 1,11$$

$$2. \text{-sel} = \frac{0,00036}{0,074 \cdot 0,104 \cdot 0,058} = 0,81$$

Предполагаемым информантным аффиксом I_7 является -SET.

Базы с парадигмой $I_1, I_2, I_3, I_4, I_5, I_6, I_7$

1. negatiiv- $I_1, I_2, I_3, I_4, I_5, I_6, I_7$, seid(2), sel(3), semaks(1), semat(1), sete(2), seteks(1), sus(1)
2. positiiv- $I_1, I_2, I_3, I_4, I_5, I_6, I_7$, seid(1), sel(6), ses(1), sest(1), sete(1), seteks(1), sus(1)

После того как в информантном подмножестве остались две базы, назовем аффиксы, общие для обеих баз, информантными и дадим им номера информантных аффиксов по убывающей величине их условных вероятностей.

Условные вероятности аффиксов по их убывающей величине:

1. -set	=0,00064	6. -sus	=0,00008
2. -sel	=0,00036	7. -semat	=0,00008
3. -seid	=0,00012	8. -semaks	=0,00004
4. -sete	=0,00012	9. -sest	=0,00004
5. -seteks	=0,00008	10. -ses	=0,00004

Таким образом мы получили следующие информантные аффиксы:

I_7	= -SET	I_{11}	= -SETEKS
I_8	= -SEL	I_{12}	= -SUS
I_9	= -SEID	I_{13}	= -SEMAT
I_{10}	= -SETE		

Так как выделять больше аффиксов из информантного подмножества уже невозможно, то полученными информантными аффиксами кончается работа алгоритма по выделению парадигмы первого морфологического типа.

Следовательно, первый морфологический тип, полученный на текстах по радиоэлектронике, характеризуется следующими информантными аффиксами:

I_1	= -SE	I_8	= -SEL
I_2	= -NE	I_9	= -SEID
I_3	= -SELT	I_{10}	= -SETE
I_4	= -SELE	I_{11}	= -SETEKS
I_5	= -SED	I_{12}	= -SUS
I_6	= -SEKS	I_{13}	= -SEMAT
I_7	= -SET		

2. Лингвистические комментарии к I морфологическому типу (на радиоэлектронных текстах)

- I. Выделенные алгоритмом информантные аффиксы оказались лингвистически окончаниями склоняемых слов. К первому морфологическому типу относятся склоняемые слова, оканчивающиеся на следующие графемы в нижеуказанных главных падежах:

Единственное число:	Множественное число:
номинатив -ne	-sed
генитив -se	-sete
партитив -set	-seid

- II. Информантный аффикс -semat является окончанием партитива единственного числа от сравнительной степени. Следовательно, слова, принадлежащие к I морфологическому типу, должны иметь степени сравнения.
- III. Третье требование к словам, принадлежащим к I морфологическому типу, предъявляет информантный аффикс -sus. На -sus могут оканчиваться в номинативе единственного числа имена существительные. Следовательно, при помощи информантного аффикса -sus базы должны образовать имя существительное (в традиционном смысле этого слова).
- IV. Итак: к I морфологическому типу относятся слова типа (взяты внеалгоритмическим способом):

abstraktne	kollektiivne
agressiivne	musikaalne
aktuaalne	solidaarne

3. Истинный шаг алгоритма, полученный на публицистических текстах

КФ $D_{-1} = 3,7$

Берем $v = 30\,000$.

Словоформ с D_{-1} оказалось 3797.

Условные вероятности графем в -2 позиции при D_{-1}

U = 0,303	R = 0,012
I = 0,298	Ö = 0,004
A = 0,208	O = 0,003
E = 0,129	L = 0,001
N = 0,015	

Так как $U = 0,303$ и $I = 0,298$, информантная цепочка замыкается.

Лингвистические комментарии информантного подмножества

Обращает на себя внимание довольно равномерное расположение условных вероятностей графем в -2 позиции при D_{-1} . Это объясняется способностью D_{-1} выступать в качестве грамматического признака множественности как при склоняемых словах, так и при глаголе, а также частью признака и при причастиях прошедшего времени на $-tud$, $-nud$.

После шлейфа мы имеем 936 баз I вида. В тексте оказалась 481 база II вида.

При базах II вида D_{-1} является:

I. Признаком множественности

1. В номинативе мн. ч. (79,0%);
2. В 3 л. мн. ч. прош. вр. глагола (*astusid, jõudsid*);
3. В 3 л. мн. ч. наст. вр. глагола (*asuvad*).

II. Полным формальным окончанием

1. В пассиве мн. ч. (*muljeid, teateid*);
2. Во 2 л. ед. ч. наст. вр. глагола (*jõuad, lakkad, nägid*).

III. Частью формального признака

1. Причастия на $-tud$ (*nimetatud*);
2. Причастия на $-dud$ (*õeldud*);
3. Причастия на $-nud$ (*lõpetanud*).

IV. Частью омонимичных форм

1. а) 2 л. ед. ч. глагола
б) 3 л. мн. ч. глагола (*tulid*);
2. а) Признаком множественности в номинативе мн. ч.
б) Полным формальным окончанием (*komsõmolitööd, partitiva ед. ч. koostööd*);
3. Признаком множественности в
а) номинативе мн. ч. (*sügavad, teatavad*,
б) 3 л. мн. ч. наст. вр. *töötavad, puuduvad*);
4. а) признаком множественности в
номинативе мн. ч.
б) 2 л. ед. ч. наст. вр. глагола (*kohtud, kallad*);
5. Частью корня (*brigaad, kevad, seekord*).

При базах II вида встретилось 495 предполагаемых информантных аффиксов, из которых в верхнее подмножество вошло Ø-окончание с условной вероятностью 0,0265.

Базы с парадигмой И₁, И₂

1. aasta-	И ₁ , И ₂	ga (4), id (7), il (2), ks (6), l (1), le (1), plaani (3), plaaniks (1), päeva (2), päevale (2), s (3), st (4), t (16), te (7), tel (4)
2. abinõu-	И ₁ , И ₂	
3. agentuuri-	И ₁ , И ₂	
4. agronoomi-	И ₁ , И ₂	dele (1)
5. ajakirjaniku-	И ₁ , И ₂	
6. ajalehe-	И ₁ , И ₂	artiklitest (1), ga (1), s (2)
7. alusetu-	И ₁ , И ₂	ks (1)
8. analüüsi-	И ₁ , И ₂	de (1), lehed (1), lehte (1), ta (1), ti (1)
9. asetäitja-	И ₁ , И ₂	
10. asjatu-	И ₁ , И ₂	lt (3), ndmatult (1), ndmatus (1)
11. asutuse-	И ₁ , И ₂	ga (1)
12. asuva-	И ₁ , И ₂	s (1), st (2), t (1)
13. autoinspektori-	И ₁ , И ₂	te (2), tel (1), tele (1)
14. autori-	И ₁ , И ₂	ga (1), le (1), te (1), teks (1), tel (2), test (1)
15. avalduse-	И ₁ , И ₂	le (1), s (2)
16. avaliku-	И ₁ , И ₂	
17. baasi-	И ₁ , И ₂	dega (1), l (1), le (1), st (1)
18. brigaadi-	И ₁ , И ₂	de (1), des (1), le (1), päevad (1), s (2)
19. delegatsiooni-	И ₁ , И ₂	
20. dissertatsiooni-	И ₁ , И ₂	de (3)
21. ebaõige-	И ₁ , И ₂	id (1)
22. eeskuju-	И ₁ , И ₂	ks (1), l (1)
23. ehitatava-	И ₁ , И ₂	
24. elamu-	И ₁ , И ₂	tes (1)
25. endise-	И ₁ , И ₂	lt (2)
26. entusiastii-	И ₁ , И ₂	
27. episoodi-	И ₁ , И ₂	des (1)
28. esimehe-	И ₁ , И ₂	le (1)
29. esimese-	И ₁ , И ₂	ks (2), l (6), le (4), s (2)
30. esindaja-	И ₁ , И ₂	ga (1), id (2), le (1), te (1)
31. etapi-	И ₁ , И ₂	l (1)
32. ettepaneku-	И ₁ , И ₂	id (4), l (1), lt (2), tega (1)
33. ettevõtte-	И ₁ , И ₂	is (1), s (1)
34. grupi-	И ₁ , И ₂	
35. fakti-	И ₁ , И ₂	de (1), dest (1)
36. fašistliku-	И ₁ , И ₂	
37. haava-	И ₁ , И ₂	tud (3)
38. halli-	И ₁ , И ₂	
39. hooldaja-	И ₁ , И ₂	

40. hoone-	H₁, H₂ ,	id(7), kesi(1), l(2), le(3), s(2), stamine(1), statud(1)
41. igapäevase-	H₁, H₂ ,	id(1)
42. informatsiooni- agentuuri-	H₁, H₂ ,	de(1)
43. inimese-	H₁, H₂ ,	l(1), le(2), st(2)
44. inseneri-	H₁, H₂ ,	de(1), del(1), dele-mehhanisaa- toritele(1)
45. isiku-	H₁, H₂ ,	id(2), l(1), test(1)
46. isoleeriva-	H₁, H₂	
47. istungi-	H₁, H₂ ,	järgul(1), l(3)
48. jaama-	H₁, H₂ ,	s(2), st(1)
49. juhataja-	H₁, H₂ ,	id(1)
50. juhendaja-	H₁, H₂ ,	le(2)
51. jääva-	H₁, H₂	
52. kaadri-	H₁, H₂ ,	st(1), t(3)
53. kahju-	H₁, H₂ ,	ks(2), le(1)
54. kangelase-	H₁, H₂	
55. karistuse-	H₁, H₂	
56. kateedri-	H₁, H₂ ,	s(5), te(1), tes(1)
57. katse-	H₁, H₂ ,	baasidega(1), baasist(1), jaa- madest(1), jaamas(1), kivi (1), tada(1), tatava(1)
58. kaupluse-	H₁, H₂ ,	komisjonid(1), komisjonidele (1), le(2)
59. kirja-	H₁, H₂ ,	kandjad(1), likest(1), nduse(1), s(2), saatjad(1), vahetusse(1)
60. klassijuhataja-	H₁, H₂	
61. kogukonna-	H₁, H₂ ,	le(1), s(1)
62. kokkuleppe-	H₁, H₂ ,	id(2)
63. kollektiivi-	H₁, H₂ ,	de(1), ga(2), l(1), le(2), s(3), st(1)
64. kolmanda-	H₁, H₂ ,	ks(2), l(2), s(1)
65. komandöri-	H₁, H₂ ,	na(2)
66. komisjoni-	H₁, H₂ ,	l(1), le(1)
67. kommunisti-	H₁, H₂	
68. kommunistliku-	H₁, H₂ ,	lt(2), st(2)
69. komsomolitöö-	H₁, H₂ ,	s(1)
70. konverentsi-	H₁, H₂	
71. kooli-	H₁, H₂ ,	de(7), des(5), desse(1), ga(1), hoone(1), lastena(1), maja (1), noored(1), pingist(1), praktikas(1), s(12), st(1), sööklad(1), tamiseks(1), vä- lise(1), välised(2), välisest (1)

72. koolivälise-	H₁, H₂	st(1)
73. koopia-	H₁, H₂	
74. koosseisu-	H₁, H₂	
75. koostöö-	H₁, H₂	ks(2), le(1)
76. korrespondendi-	H₁, H₂	
77. kostüümi-	H₁, H₂	
78. kursuse-	H₁, H₂	
79. kvartali-	H₁, H₂	s(8), te(1)
80. kõrge-	H₁, H₂	le(2), m(1), ma(2), male(1), mat(2), mate(4), mates(1), t(1)
81. käiva-	H₁, H₂	tele(1)
82. külanõukogu-	H₁, H₂	s(2)
83. küsimuse-	H₁, H₂	le(5), s(1)
84. laboratooriumi-	H₁, H₂	des(1), s(1)
85. lahtise-	H₁, H₂	
86. lakka-	H₁, H₂	
87. lepingu-	H₁, H₂	te(1)
88. lihtsa-	H₁, H₂	id(1), lt(6), maks(1)
89. liikme-	H₁, H₂	ga(1), id(4), ks(6), l(4), pileti (1), te(3), tel(1), test(1)
90. loengu-	H₁, H₂	id(10), line(2), lises(1), st(2), t(7), te(2), tega(1), tel(2), tele(1)
91. loosungi-	H₁, H₂	t(1), te(1)
92. maa-ala-	H₁, H₂	le(1)
93. majandi-	H₁, H₂	juh(1), juht(1), juhtide(1), s(6), t(1), te(8), tega(1), tes (7), test(1)
94. materjali-	H₁, H₂	de(5), dega(2), dest(1), le(1)
95. mehhanisaatori-	H₁, H₂	le(1), te(1)
96. meistri-	H₁, H₂	st(1), t(1), te(1)
97. metoodilise-	H₁, H₂	
98. ministri-	H₁, H₂	te(3)
99. mitme-	H₁, H₂	id(6), l(3), le(1), poolsed(1), s(2), st(1), sugused(3), sugu- seid(6), suguste(3), sugus- tesse, tahulise(1), te(5), tel(1), tes(2), tesse(1)
100. mõjuva-	H₁, H₂	
101. mõlema-	H₁, H₂	le(1), te(2)
102. naise-	H₁, H₂	
103. niisuguse-	H₁, H₂	id(6), ks(1), na(1), sse(1)
104. noore-	H₁, H₂	le(2), lt(1), m(3), mat(1)
105. nõukogu-	H₁, H₂	de(38), desse(1), l(1), s(1), sid (1)
106. nädala-	H₁, H₂	l(1), lehes(1), st(1), t(2)

107. näite-	H₁, H₂	id(6), ks(23)
108. odava-	H₁, H₂	ma(1), maid(1)
109. organisatsiooni-	H₁, H₂	de(3), dega(1), ga(1), le(1)
110. otsese-	H₁, H₂	le(2), lt(3)
111. otsuse-	H₁, H₂	ga(1), id(1), s(2), st(1)
112. palga-	H₁, H₂	tõendeid(1)
113. palju-	H₁, H₂	de(4), deks(1), dele(2), des(4), desse(1), dest(1), gi(2), kan- natanud(1), ndamiseks(1), nemine(1), sid(2), st(2)
114. paranduse-	H₁, H₂	
115. parema-	H₁, H₂	ks(1), l(4), t(1), te(4), telt(1), tést(1)
116. parima-	H₁, H₂	ks(1), s(1), tel(1)
117. partei-	H₁, H₂	koosolekud(1), l(1), organisat- sioonis(1), töö(1)
118. passi-	H₁, H₂	ga(1), kandjad(1), le(1)
119. perekonna-	H₁, H₂	
120. pikaajalise-	H₁, H₂	
121. pildi-	H₁, H₂	l(3), stasin(1), stavad(1)
122. plaani-	H₁, H₂	kohaseks(1), le(3), organitelt (1), päraseks(1), pärasest(1), s(2)
123. planeeritu-	H₁, H₂	
124. poliitilise-	H₁, H₂	
125. poole-	H₁, H₂	ks(1), l(1), li(1), s(4), st(1)
126. praeguse-	H₁, H₂	s(1), st(2)
127. praktilise-	H₁, H₂	
128. probleemi-	H₁, H₂	de(4), dega(1), dest(1), ga(3), ks(2)
129. projekterija-	H₁, H₂	
130. protsendi-	H₁, H₂	
131. puhta-	H₁, H₂	id(1), lt(1)
132. puuduliku-	H₁, H₂	
133. päeva-	H₁, H₂	de(2), ga(2), kodu(1), kodus (1), korda(3), korras(3), ks(1), l(4), lgi(1), problee- me(1), probleemidest(1), st (1), sündmisi(1), sündmus- tega(1)
134. raamatukogu-	H₁, H₂	
135. rahva-	H₁, H₂	ga(1), id(1), kohtu(1), le(4), majanduse(1), majandusha- rude(1), naljand(1), pidusid (1), rikkad(1), rohked(1), s(7), ste(7)
136. rahvusvahelise-	H₁, H₂	le(1)

137. raketi-	H₁, H₂	
138. raudtee-	H₁, H₂	de(1), del(1), lased(1), sse(1)
139. riigi-	H₁, H₂	ametnikud(1), juhid(1), le(4), peadele(2), s(4)
140. ringi-	H₁, H₂	de(1), desse(2), töös(1)
141. ringkonna-	H₁, H₂	
142. rongi-	H₁, H₂	de(2)
143. ruumi-	H₁, H₂	dega(1), des(1), s(1), st(2)
144. saagi-	H₁, H₂	ks(1), maade(1), se(1), st(3)
145. saanu-	H₁, H₂	
146. saavutatu-	H₁, H₂	
147. sammu-	H₁, H₂	ga(1), l(1)
148. sarnaneva-	H₁, H₂	
149. seadme-	H₁, H₂	id(1), te(1)
150. sealse-	H₁, H₂	id(1), te(1)
151. seina-	H₁, H₂	le(3), lehed(1)
152. sellise-	H₁, H₂	id(4), l(2)
153. seminari-	H₁, H₂	del(1), st(1)
154. senise-	H₁, H₂	s(1), st(5)
155. silma-	H₁, H₂	le(2), moondused(1), nähtavad (1), paistvate(1), s(3)
156. soovija-	H₁, H₂	
157. spetsialisti-	H₁, H₂	de(10), dele(1), dest(1), l(1)
158. suure-	H₁, H₂	ks(1), l(3), lamud(1), le(1), m(3), ma(2), mad(1), maid (3), maks(4), mal(1), male (1), mat(4), ndada(1), ndamise(1), ndamiseks(1), ndasid(1), ndati(1), ndatud (1), ndavad(1), neb(4), ne- mine(1), nenud(3), nes(2), päraane(1), pärased(2), pära- seid(1), päraselt(1), s(4), st(2)
159. suurema-	H₁, H₂	id(3), ks(4), l(1), le(1), t(4)
160. sõna-	H₁, H₂	de(1), ga(3), gi(1), raamatu (1), võtjad(3), võtjate(1), võtnud(2), võtu(1), võtud(1), võttude(2)
161. sõnavõtu-	H₁, H₂	
162. sügava-	H₁, H₂	id(1), l(1), lt(1), maks(1)
163. sündmuse-	H₁, H₂	
164. tarbetu-	H₁, H₂	id(1)
165. teadlase-	H₁, H₂	
166. teadusliku-	H₁, H₂	ks(1), le(2), st(1)
167. tehase-	H₁, H₂	ga(1), id(8), l(1), lt(2), sse(2), st(2)

168. tehnilise-	$\mathbb{H}_1, \mathbb{H}_2$	lt(1), s(1), st(1)
169. tehtu-	$\mathbb{H}_1, \mathbb{H}_2$	le(1), st(1)
170. teise-	$\mathbb{H}_1, \mathbb{H}_2$	ks(4), l(2), le(2), lt(2), s(3), st(2)
171. televiisori-	$\mathbb{H}_1, \mathbb{H}_2$	omanikud(1), st(1)
172. tellimuse-	$\mathbb{H}_1, \mathbb{H}_2$	l(1)
173. teose-	$\mathbb{H}_1, \mathbb{H}_2$	id(2)
174. terve-	$\mathbb{H}_1, \mathbb{H}_2$	
175. toimetuse-	$\mathbb{H}_1, \mathbb{H}_2$	l(1), s(1)
176. trahvi-	$\mathbb{H}_1, \mathbb{H}_2$	
177. traktori-	$\mathbb{H}_1, \mathbb{H}_2$	sti(1), stid(1), stidest(1), test(1)
178. traktoristi-	$\mathbb{H}_1, \mathbb{H}_2$	dest(1)
179. trassi-	$\mathbb{H}_1, \mathbb{H}_2$	
180. tubli-	$\mathbb{H}_1, \mathbb{H}_2$	l(1)
181. tuhande-	$\mathbb{H}_1, \mathbb{H}_2$	id(3), ni(1)
182. tulnu-	$\mathbb{H}_1, \mathbb{H}_2$	
183. tunni-	$\mathbb{H}_1, \mathbb{H}_2$	ga(1), l(1), s(1), stada(1), staja(1), stas(2), stati(2), statud(1), stavad(1), stuseks (2)
184. tähtaja-	$\mathbb{H}_1, \mathbb{H}_2$	
185. tänase-	$\mathbb{H}_1, \mathbb{H}_2$	id(2)
186. tänava-	$\mathbb{H}_1, \mathbb{H}_2$	l(2), lahingud(1), le(1), lt(1), s(7), sse(1), st(2), te(1), tel (1), test(1)
187. töölise-	$\mathbb{H}_1, \mathbb{H}_2$	l(1)
188. tööpassi-	$\mathbb{H}_1, \mathbb{H}_2$	de(1), s(2)
189. töötaja-	$\mathbb{H}_1, \mathbb{H}_2$	id(7), ks(1), t(7), te(14), tega (1), tele(2), test(1)
190. töötava-	$\mathbb{H}_1, \mathbb{H}_2$	id(1), le(1)
191. türklase-	$\mathbb{H}_1, \mathbb{H}_2$	
192. uurimise-	$\mathbb{H}_1, \mathbb{H}_2$	
193. uurimuse-	$\mathbb{H}_1, \mathbb{H}_2$	
194. vabariigi-	$\mathbb{H}_1, \mathbb{H}_2$	ga(2), st(1)
195. vabariikliku-	$\mathbb{H}_1, \mathbb{H}_2$	
196. vaenlase-	$\mathbb{H}_1, \mathbb{H}_2$	
197. vajaliku-	$\mathbb{H}_1, \mathbb{H}_2$	ks(1), maks(1)
198. valge-	$\mathbb{H}_1, \mathbb{H}_2$	
199. vastava-	$\mathbb{H}_1, \mathbb{H}_2$	id(3), lt(2), sse(1), te(2)
200. vastuolu-	$\mathbb{H}_1, \mathbb{H}_2$	
201. vastutava-	$\mathbb{H}_1, \mathbb{H}_2$	id(1)
202. vedurijuhi-	$\mathbb{H}_1, \mathbb{H}_2$	
203. viimase-	$\mathbb{H}_1, \mathbb{H}_2$	id(2), ks(1), l(10)
204. viiva-	$\mathbb{H}_1, \mathbb{H}_2$	
205. võimu-	$\mathbb{H}_1, \mathbb{H}_2$	de(1), kad(1)
206. võitleja-	$\mathbb{H}_1, \mathbb{H}_2$	id(2)

207. väeosa-	И₁, И₂ ,	de(1), des(1), st(1)
208. väikese-	И₁, И₂	
209. õiguse-	И₁, И₂	
210. õpetaja-	И₁, И₂ ,	id(3), te(2), tel(1), tele(1)
211. õpilase-	И₁, И₂ ,	ga(1)
212. õpiringi-	И₁, И₂ ,	desse(2)
213. õppejõu-	И₁, И₂ ,	dude(2), duest(1)
214. õppeplaani-	И₁, И₂	
215. õeldu-	И₁, И₂ ,	le(1)
216. ühiskondliku-	И₁, И₂ ,	ks(1), lt(1), s(2), st(1)
217. üksiku-	И₁, И₂ ,	id(5), l(1), te(3), tel(1)
218. üleliidulise-	И₁, И₂ ,	lt(1)
219. ülesande-	И₁, И₂ ,	id(4), ks(5)
220. ületamatu-	И₁, И₂	
221. ülevaafe-	И₁, И₂	
222. žurnaali-	И₁, И₂	

Проверка на меру редукции:

$$\frac{M_1 - M_2}{k \cdot M_1} = \frac{481 - 222}{6,70 \cdot 481} = \frac{259}{3222,7} = 0,081$$

$$k = D \frac{1}{1 + 0,02 \log v}; \quad D = 7,93; \quad v = 30000; \quad k = 6,70$$

Редукция субпороговая, следовательно, $I_1 = -D$; $I_2 = \emptyset$ окончание, и расширение парадигмы продолжается.

Условные вероятности предполагаемых информантных аффиксов с базами, имеющими **И₁, И₂**:

1. -id	=0,0042	20. -desse	=0,0002
2. -l	=0,0038	21. -dest	=0,0002
3. -s	=0,0037	22. -dega	=0,0002
4. -de	=0,0030	23. -tele	=0,0002
5. -le	=0,0024	24. -sse	=0,0002
6. -ks	=0,0024	25. -suguseid	=0,0002
7. -te	=0,0023	26. -ma	=0,0002
8. -st	=0,0018	27. -dele	=0,0001
9. -t	=0,0015	28. -maid	=0,0001
10. -lt	=0,0010	29. -mate	=0,0001
11. -ga	=0,0009	30. -neb	=0,0001
12. -tel	=0,0005	31. -tega	=0,0001
13. -des	=0,0005	32. -plaani	=0,0001
14. -tes	=0,0004	33. -gi	=0,0001
15. -mat	=0,0003	34. -tud	=0,0001
16. -test	=0,0003	35. -na	=0,0001
17. -m	=0,0002	36. -nanud	=0,0001
18. -ste	=0,0002	37. -korda	=0,0001
19. -maks	=0,0002	38. -korras	=0,0001

39.	-võtjad	=0,0001	86.	-kohaseks
40.	-del	=0,0001	87.	-kohtu
41.	-sid	=0,0001	88.	-komisjonid
42.	-suguste	=0,0001	89.	-komisjonide
43.	-sugused	=0,0001	90.	-koosolekud
44.	-lehed	=0,0001	91.	-lahingud
45.	-päeva	=0,0001	92.	-lamud
46.	-päevale	=0,0001	93.	-lased
47.	-juhid	=0,0001	94.	-lastena
48.	-il	=0,0001	95.	-lehes
49.	-nemine	=0,0001	96.	-lehte
50.	-stavad	=0,0001	97.	-lgi
51.	-stas	=0,0001	98.	-li
52.	-stuseks	=0,0001	99.	-likest
53.	-stati	=0,0001	100.	-lisse
54.	-line	=0,0001	101.	-maade
55.	-statud	=0,0001	102.	-mad
56.	-ndamiseks	=0,0001	103.	-maja
57.	-male	=0,0001	104.	-majandusharude
58.	-nes	=0,0001	105.	-majanduse
59.	-välised	=0,0001	106.	-mates
60.	-pärased	=0,0001	107.	-moondused
61.	-dude	=0,0001	108.	-naljand
62.	-võtnud	=0,0001	109.	-ndamise
63.	-peadele	=0,0001	110.	-ndasid
64.	-võttudes	=0,0001	111.	-ndati
65.	-artiklitest	=0,0000	112.	-ndatud
66.	-ametnikud		113.	-ndavad
67.	-baasidega		114.	-ndamatus
68.	-baasist		115.	-ndamatult
69.	-dele-mehhanisaatoritele		116.	-nduse
70.	-dudest		117.	-nevad
71.	-hoone		118.	-ni
72.	-is		119.	-noored
73.	-jaamadest		120.	-nähtavad
74.	-jaamas		121.	-omanikud
75.	-juht		122.	-organisatsioonid
76.	-juhtide		123.	-organitelt
77.	-järgul		124.	-paistvate
78.	-kandjad		125.	-pidusid
79.	-kandjaid		126.	-pileti
80.	-kannatanud		127.	-pingist
81.	-kat		128.	-plaaniks
82.	-kesi		129.	-poolsed
83.	-kivi		130.	-praktikas
84.	-kodu		131.	-probleeme
85.	-kodus		132.	-probleemidest

133. -päevad	153. -sööklad
134. -pärane	154. -sündernusi
135. -päraseid	155. -sündernustega
136. -päraseks	156. -ta
137. -päraselt	157. -tada
138. -pärasest	158. -tahulise
139. -raamatu	159. -tamiseks
140. -rikkad	160. -tatava
141. -rohked	161. -teks
142. -saatjad	162. -telt
143. -se	163. -tesse
144. -stada	164. -ti
145. -staja	165. -tõendeid
146. -stamine	166. -töö
147. -stasin	167. -vahetusse
148. -statud	168. -võtjate
149. -sti	169. -võtu
150. -stid	170. -võtud
151. -stidest	171. -välise
152. -sugustesse	172. -välisest

Медиальное разбиение и КФ для верхнего подмножества:

$$\frac{0,0042+0,0000}{2} = 0,0021$$

$$1. \text{ КФ -ID} = \frac{0,0024}{0,087 \cdot 0,038} = 1,27$$

$$2. \text{ КФ -L} = \frac{0,0038}{0,058} = 0,07$$

$$3. \text{ КФ -S} = \frac{0,0037}{0,074} = 0,05$$

$$4. \text{ КФ -DE} = \frac{0,0030}{0,038 \cdot 0,104} = 0,769$$

$$5. \text{ КФ -LE} = \frac{0,0024}{0,075 \cdot 0,104} = 0,398$$

$$6. \text{ КФ -KS} = \frac{0,0042}{0,039 \cdot 0,074} = 0,831$$

$$7. \text{ КФ -TE} = \frac{0,0023}{0,058 \cdot 0,104} = 0,269$$

Базы с парадигмой И₁, И₂, И₃

1. aasta-	И ₁ , И ₂ , И ₃ ,	ga(4), il(2), ks(6), l(43), le(1), plaani(3), plaaniks (1), päeva(2), päevale(2), s(3), st(4), t(16), te(7), tel(4)
2. ebaõige-	И ₁ , И ₂ , И ₃	
3. esindaja-	И ₁ , И ₂ , И ₃ ,	ga(1), le(1), te(1)
4. ettepaneku-	И ₁ , И ₂ , И ₃ ,	l(1), lt(2), tega(1)
5. hoone-	И ₁ , И ₂ , И ₃ ,	kesi(1), l(2), le(3), s(2), stamine(1), statud(1)
6. igapäevase-	И ₁ , И ₂ , И ₃	
7. isiku-	И ₁ , И ₂ , И ₃ ,	l(1), test(1)
8. juhataja-	И ₁ , И ₂ , И ₃	
9. kokkuleppe-	И ₁ , И ₂ , И ₃	
10. lihtsa-	И ₁ , И ₂ , И ₃ ,	lt(6), maks(1)
11. liikme-	И ₁ , И ₂ , И ₃ ,	ga(1), ks(6), l(4), pileti(1), te(3), tel(1), test(1)
12. loengu-	И ₁ , И ₂ , И ₃ ,	line(2), lises(1), st(2), t(7), te(2), tega(1), tel(2), tele (1)
13. mitme-	И ₁ , И ₂ , И ₃ ,	l(3), le(1), poolsed(1), s(2), st(1), sugused(3), sugu- seid(6), suguste(3), sugus- tesse(1), tahulise(1), te(5), tel(1), tes(2), tesse(1)
14. niisuguse-	И ₁ , И ₂ , И ₃ ,	ks(1), na(1), sse(1)
15. näite-	И ₁ , И ₂ , И ₃ ,	ks(23)
16. otsuse-	И ₁ , И ₂ , И ₃ ,	ga(1), s(2), st(1)
17. puhta-	И ₁ , И ₂ , И ₃ ,	lt(1)
18. rahva-	И ₁ , И ₂ , И ₃ ,	ga(1), kohtu(1), le(4), ma- jandusharude(1), naljand (1), pidusid(1), rikkad(1), rohked(1), s(7), ste(7)
19. suurema-	И ₁ , И ₂ , И ₃ ,	ks(4), l(1), le(1), t(4)
20. sügava-	И ₁ , И ₂ , И ₃ ,	l(1), lt(1), maks(1)
21. tarbetu-	И ₁ , И ₂ , И ₃	
22. tehase-	И ₁ , И ₂ , И ₃ ,	ga(1), l(1), lt(2), sse(2), st(2)
23. teose-	И ₁ , И ₂ , И ₃	
24. tuhande-	И ₁ , И ₂ , И ₃ ,	ni(1)
25. tänase-	И ₁ , И ₂ , И ₃	
26. töötaja-	И ₁ , И ₂ , И ₃ ,	ks(1), t(7), te(14), tega(1), tele(2), test(1)
27. töötava-	И ₁ , И ₂ , И ₃ ,	le(1)
28. vastava-	И ₁ , И ₂ , И ₃ ,	lt(2), sse(1), te(2)

29. vastutava-	И₁, И₂, И₃	
30. viimase-	И₁, И₂, И₃ ,	ks(1), l(10)
31. võitleja-	И₁, И₂, И₃	
32. õpetaja-	И₁, И₂, И₃	
33. üksiku-	И₁, И₂, И₃ ,	l(1), te(3), tel(1)
34. ülesande-	И₁, И₂, И₃ ,	ks(5)
35. ületamatu-	И₁, И₂, И₃	
36. ülevaate-	И₁, И₂, И₃	

Проверка на меру редукции:

$$\frac{M_1 - M_2}{k \cdot M_1} = \frac{222 - 36}{6,70 \cdot 222} = \frac{186}{1487,4} = 0,125$$

Условные вероятности предполагаемых информантных аффиксов с базами, имеющими **И₁, И₂, И₃**

1. -l	=0,0022	23. -päeva	=0,0001
2. -ks	=0,0016	24. -päevale	=0,0001
3. -te	=0,0013	25. -tes	=0,0001
4. -t	=0,0011	26. -test	=0,0001
5. -s	=0,0005	27. -kesi	=0,0000
6. -lt	=0,0005	28. -kohtu	
7. -le	=0,0004	29. -majanduse	
8. -pileti	=0,0004	30. -majandusharude	
9. -ga	=0,0003	31. -na	
10. -st	=0,0003	32. -naljand	
11. -tel	=0,0003	33. -ni	
12. -ste	=0,0002	34. -pidusid	
13. -suguseid	=0,0002	35. -plaaniks	
14. -plaani	=0,0001	36. -poolused	
15. -sse	=0,0001	37. -rikkad	
16. -sugused	=0,0001	38. -rohked	
17. -suguste	=0,0001	39. -stamine	
18. -tega	=0,0001	40. -statud	
19. -tele	=0,0001	41. -sugustesse	
20. -il	=0,0001	42. -tahulise	
21. -line	=0,0001	43. -tesse	
22. -maks	=0,0001	44. -test	

Медиальное разбиение и КФ для верхнего подмножества:

$$\frac{0,0022 + 0,0000}{2} = 0,0011$$

$$1. \text{ КФ } -L = \frac{0,0022}{0,058} = 0,038$$

$$2. \text{ КФ } -KS = \frac{0,0016}{0,039 \cdot 0,074} = 0,561$$

$$3. \text{ КФ -TE} = \frac{0,0013}{0,075 \cdot 0,104} = 0,167$$

$$4. \text{ КФ -T} = \frac{0,0011}{0,075} = 0,015$$

Базы с парадигмой **И₁, И₂, И₃, И₄**

1. aasta-	И₁, И₂, И₃, И₄	ga(4), il(2), l(43), le(1), plaani(3), plaaniks(1), päeva(2), päevale(2), s(3), st(4), t(16), te(7), tel(4)
2. liikme-	И₁, И₂, И₃, И₄	ga(1), l(4), pileti(1), te(3), tel(1), test(1)
3. niisuguse-	И₁, И₂, И₃, И₄	na(1), sse(1)
4. näite-	И₁, И₂, И₃, И₄	
5. suurema-	И₁, И₂, И₃, И₄	l(1), le(1), t(4)
6. töötaja-	И₁, И₂, И₃, И₄	t(7), te(14), tega(1), tele(2), test(1)
7. viimase-	И₁, И₂, И₃, И₄	l(1)
8. ülesande-	И₁, И₂, И₃, И₄	

Мера редукции:

$$\frac{M_1 - M_2}{K \cdot M_1} = \frac{36 - 8}{6,70 \cdot 36} = \frac{28}{241,2} = 0,112$$

Условные вероятности предполагаемых информантных аффиксов с базами, имеющими **И₁, И₂, И₃, И₄**:

1. -l	=0,0019
2. -t	=0,0009
3. -te	=0,0008
4. -ga	=0,0002
5. -tel	=0,0002
6. -st	=0,0001
7. -s	=0,0001
8. -plaani	=0,0001
9. -le	=0,0001
10. -il	=0,0001
11. -päeva	=0,0001
12. -päevale	=0,0001
13. -tele	=0,0001
14. -plaaniks	=0,0000
15. -na	
16. -sse	
17. -tega	
18. -tele	

Медиальное разбиение:

$$\frac{0,0019+0,0000}{2} = 0,0009$$

КФ для верхнего подмножества:

$$1. \text{ КФ } -L = \frac{0,0019}{0,058} = 0,033$$

$$2. \text{ КФ } -T = \frac{0,0009}{0,075} = 0,012$$

Базы с парадигмой **И₁**, **И₂**, **И₃**, **И₄**, **И₅**

- | | | |
|-------------|--|--|
| 1. aasta- | И₁ , И₂ , И₃ , И₄ , И₅ , | ga(14), il(2), le(1), plaani(3),
plaaniks(1), päeva(2), päe-
vale(2), s(3), st(4), t(16),
te(7), tel(4) |
| 2. liikme- | И₁ , И₂ , И₃ , И₄ , И₅ , | ga(1), pileti(1), te(3), tel(1),
test(1) |
| 3. suurema- | И₁ , И₂ , И₃ , И₄ , И₅ , | le(1), t(4) |
| 4. viimase- | И₁ , И₂ , И₃ , И₄ , И₅ | |

Мера редукции для предполагаемого **И₅**:

$$\frac{M_1 - M_2}{k \cdot M_1} = \frac{8 - 4}{6,70 \cdot 8} = \frac{4}{53,6} = 0,074$$

Условные вероятности предполагаемых информантных аффиксов с базами, имеющими **И₁**, **И₂**, **И₃**, **И₄**, **И₅**:

- | | |
|---------------|----------|
| 1. -t | = 0,0007 |
| 2. -te | = 0,0003 |
| 3. -ga | = 0,0002 |
| 4. -tel | = 0,0002 |
| 5. -s | = 0,0001 |
| 6. -plaani | = 0,0001 |
| 7. -st | = 0,0001 |
| 8. -le | = 0,0001 |
| 9. -il | = 0,0001 |
| 10. -päeva | = 0,0001 |
| 11. -päevale | = 0,0001 |
| 12. -plaaniks | = 0,0001 |
| 13. -pileti | = 0,0001 |
| 14. -test | = 0,0000 |

Медиальное разбиение и КФ для верхнего подмножества:

$$\frac{0,0007+0,000}{2} = 0,0004$$

$$K\Phi - T = \frac{0,0007}{0,075} = 0,009$$

Базы с парадигмой **И₁, И₂, И₃, И₄, И₅, И₆**

1. aasta- **И₁, И₂, И₃, И₄, И₅, И₆**, ga(4), il(2), le(1), plaani(3), plaaniks(1), päeva(2), päevale(2), s(4), te(7), tel(4)
2. suurema- **И₁, И₂, И₃, И₄, И₅, И₆**, le(1)

Информантное множество состоит из двух баз. Назовем аффиксы, общие для обеих баз, информантными.

Следовательно, $I_7 = -LE$.

Таким образом, I морфологический тип характеризуется следующими информантными аффиксами:

$$\begin{aligned} I_1 &= -D \\ I_2 &= -\emptyset \\ I_3 &= -ID \\ I_4 &= -KS \\ I_5 &= -L \\ I_6 &= -T \\ I_7 &= -LE \end{aligned}$$

4. Лингвистические комментарии к I морфологическому типу (на публицистических текстах)

- I. Первым информантным аффиксом, на основе которого образовано информантное подмножество, является $-D$. D_{-1} является признаком множественности всех склоняемых слов, а также и частью омонимии падежей: партитива ед. ч. ($-D_{-1}$ является падежным окончанием) и номинатива мн. ч. ($-D_{-1}$ является признаком множественности)

у слов типа maа — maad
töö — tööd
kantselei — kantseleid

Так как $-D_{-1}$ может быть полным окончанием, а также и частью окончания у глагольных форм, то принадлежность его к склоняемым словам (для выражения признака множественности) выявляется при содействии I_2 . Однако, являясь единым признаком множественности для всех склоняемых слов, D_{-1} не дает никакой информации о типах внутри склоняемых слов.

- II. Зная, что номинатив множественного числа образован от

IV. ОБЩИЕ ВЫВОДЫ

Таким образом, в настоящей работе использовалась статистико-комбинаторная модель языка для изучения грамматической и статистической структуры эстонского языка.

1. Каждый язык, а также и подъязык (как мы увидели), по-своему реагирует на алгоритм статистико-комбинаторного моделирования языков. Это своеобразие, с одной стороны, как было сказано выше, зависит от типологии языков той группы, к которой принадлежит исследуемый язык, а с другой стороны — от самого конкретного языка. Именно это последнее обстоятельство дает нам различные исходы и в родственных языках.

Эстонский язык является первым из финно-угорской семьи языков, к которому применяется метод статистико-комбинаторного моделирования языка¹³. По этой причине представлялся правильным следующий подход к работе. Работая по алгоритму модели языка и обнаруживая полученные алгоритмом модели данные, интересные лингвисту для анализа языка, но не подлежащие дальнейшему анализу алгоритмом модели (из-за универсальности модели), лингвист обязан идти по пути, указанному ему самим языком. Иначе работа его может потерять многое из своей ценности и превратиться отчасти в иллюстративный материал модели на данных используемого им языка. Таким образом лингвист должен не только проверить работу модели на материале определенного языка, но зафиксировать и те дополнительные результаты, которые не предписано обработать алгоритмом, и порою отойти от всеобщего хода алгоритма. Под отходом от алгоритма понимается не переделывание алгоритма, а одновременно 1) работа над алгоритмом и 2) следование за языковыми данными, не подлежащими анализу алгоритмом, но интересными в лингвистическом смысле.

2. Примером влияния типологии языка на ход алгоритма статистико-комбинаторного моделирования языка может служить понятие устойчивости, взятое из алгоритма¹⁴. По содержанию понятие устойчивости выросло из той закономерности в языке, которую мы называем словоизменением. Логически вполне закономерно ожидание того, что если в языке существует словоизменение, то признаков, по которым идет классификация (то есть аффиксов), несравнимо меньше, чем классифицируемых элементов (т. е. баз). В противном случае в языке не было бы регулярного словоизменения и язык состоял бы из одноэлемент-

¹³ В Венгрии стали применять данный метод к венгерскому языку с ноября 1963 года, тогда как статистико-комбинаторным моделированием эстонского языка автор занимается с ноября 1962 года.

¹⁴ О понятии устойчивости и устойчивого класса см. Н. А. Андреев, ук. соч., стр. 12.

ных множеств (т. е. словоформ), каждое из которых характеризовалось бы своим единственным элементом. Следовательно: с увеличением количества словоформ на достаточно обширных текстах (когда алгоритм уже выделил все основные аффиксы) наступает момент, с которого наращивается множество баз при сравнительно неизменяемом множестве аффиксов. С этого момента мы фактически подбираем к лексически высокочастотным базам парадигму, так как редко встречаемые (но с потенциально полной парадигмой) базы фактически алгоритмом почти сразу исключаются. Рассмотрим данное теоретическое положение на материале русского и эстонского языков. Берем словоизменение склоняемых слов, как наиболее высокочастотное явление словоизменения в обоих языках. При подсчете не учитывается возможное совпадение отдельных падежных форм при одном и при различных типах склонения.

В русском языке

6 падежей (ед. 6 + мн. 6) = 12
 типов склонения 30 = 30

Итого $12 \times 20 = 360$ потенциально возможных аффиксов

В эстонском языке

14 падежей (ед. 14 + мн. 14) = 28
 7 падежей во мн. числе имеют
 регулярное двойное словоизме-
 нение = 7

типов склонения 35 = 90

Итого $35 \times 90 = 3150$ потенциально возможных аффиксов.

При этом надо учесть изменение базы в эстонском языке.
 Ср.

riik — riigi,
 hammas — hamba.

Около половины склоняемых слов имеют внутренние изменения баз, и таким образом количество баз удваивается, следовательно, вероятность базы уменьшается. В таких условиях возникновение устойчивого класса хотя не исключается, но его возникновение требует, несомненно, больше алгоритмического времени, и, практически возникая, он будет малозлементным. По сравнению с русским языком, возникновение устойчивого класса на материале эстонского языка требует увеличения текста в несколько раз, и, следовательно, эта работа может выполняться лишь машинным способом. Практически: хотя и нужно учесть возможность появления устойчивого класса, работа

алгоритма идет по-аффиксно. Серьезно может устойчивость повлиять на ускорение работы алгоритма, очевидно, лишь при машинном способе обработки текста, хотя эту возможность надо проверить на практике.

Как видно из предыдущего примера, появление и непоявление устойчивости связано с типологией обрабатываемых языков.

3. Типы, выделенные на основе эстонских текстов методом статистико-комбинаторного моделирования языка, позволяют несомненно в дальнейшем получить интересные данные о самой структуре и типологических особенностях эстонского языка. Одним из фактов, касающихся структуры языка, является то, что морфологический тип в эстонском языке может объединять в себе как прилагательное, так и существительное. Это соответствует данным языка, поскольку на морфологическом уровне в эстонском языке существительные от прилагательных формально не отличаются. Этот факт сам по себе не новый, он зафиксирован и в орфографическом словаре эстонского языка, где типовым словом существительного может оказаться прилагательное и наоборот. Но в данном случае под различительными признаками подразумеваются смысловые; так, в качестве типового слова для слова *aasta* (год) дано *kõrge* (высокий) («высота» было бы — *kõrgus*).

4. С точки зрения сопоставительного языкознания модель может служить как бы эталоном. Сопоставлению подвергаются лишь те языковые явления в сопоставляемых языках, в зависимости от которых даются различные ходы алгоритма и получаются различные данные. Т. е. модель раскрывает типологическую структуру языка, и типологический подход на данном уровне является наиболее продуктивным средством сравнения языков.

«Прежде всего нам представляется, что этот подход (подразумевается типологический подход — Х. Х.) должен основываться на идее абстрактного моделирования языка. Этот подход возник в связи с потребностями практики, а именно: практики сопоставления алгоритмов для машинного перевода с одного языка на другой. Дело в том, что работы по осуществлению машинного перевода с одного языка на другой уже выходят за рамки частичных типологических штудий, представляя собой фрагмент универсальной типологической схемы, которую можно будет считать построенной, как только в нашем распоряжении окажутся алгоритмы всех возможных переводов от одного языка к другому или же языку-посреднику».¹⁵

¹⁵ З. М. Волоцкая, Т. Н. Молошная, Т. М. Николаева, И. И. Ревзин, Т. В. Цивьян. Об одном подходе к типологии славянских языков. — Славянское языкознание. V международный съезд славистов, Изд. АН СССР, М., 1963, стр. 511.

Наоборот, имея моделированные по определенной модели языки, сравнительно легко будет построить переводческие алгоритмы. Таким образом подчеркивается взаимосвязь между машинным переводом и моделированием языков.

Построение математических моделей языков является, несомненно, более точным и общим подходом к языку. Однако сами построенные модели приобретают свою ценность и право на существование постольку, поскольку они работоспособны в языке. Язык является таким образом той объективной мерой, при помощи которой дается оценка на ценность моделей. Несомненно, наибольшую ценность имеют те математические модели, которые способны вырабатывать большее количество языковой информации.

СПИСОК ТЕКСТОВ, ОБРАБОТАННЫХ АЛГОРИТМОМ СТАТИСТИКО-КОМБИНАТОРНОГО МОДЕЛИРОВАНИЯ ЯЗЫКА

Независимые вероятности графем

1. E. Alt ja E. Jakoobi, Raadiokorrastaja käsiraamat. — Tallinn, 1960, стр. 5, 24, 57, 231.
2. I. Elgas, Raadioseadmed. — Tallinn, 1953, стр. 165—166.
3. V. Halliksoo, Transistoride kasutamine raadioseadmetes. — Tallinn, 1962, стр. 54—55.
4. V. Halliksoo, Transistoride tunnussarjade ülesvõtmine. — Tehnika ja Tootmine, 1960, 3, стр. 35.
5. A. Isotamm, Raadioamatööri käsiraamat. — Tallinn, 1950, стр. 18, 156—157, 190—192, 381—382, 514.
6. A. Kaskneem, Lühilainete levimisest. — Tehnika ja Tootmine, 1961, 2, стр. 32.
7. A. Laevnik, Värvuskaugnägemine. — Tehnika ja Tootmine, 1961, 1, стр. 29.
8. H. Pedusaar, Logaritmiline väljundimõõtja heliseadmetele. — Tehnika ja Tootmine, 1960, 10, стр. 16.
9. H. Pedusaar, Konverter stereofooniliste raadioseadmete vastuvõtuks. — Tehnika ja Tootmine, 1962, 4, стр. 20.
10. H. Pedusaar, Raadiopraktika. — Tallinn, 1961, стр. 10, 16, 131.
11. A. Pihlo, Raadiohäirete vähendamisest. — Tehnika ja Tootmine, 1962, 6, стр. 27.
12. A. Raave, Lihtsad taskuvastuvõtjad. — Tehnika ja Tootmine, 1962, 9, стр. 29.
13. R. Sari, Plasmatorvik. — Tehnika ja Tootmine, 1961, 9, стр. 34.
14. A. Sillart, Televisioonitehnika. — Tallinn, 1960, стр. 25, 90, 108—109, 122.
15. Toimetuse artikkel, Elektromagnetilised filtrid. — Tehnika ja Tootmine, 1962, 8, стр. 28.
16. Toimetuse artikkel, Kooli või käitise radiofitseerimisest. — Tehnika ja Tootmine, 1961, 12, стр. 23—24.
17. Toimetuse artikkel, Televisioonisaadete ülekandevõrk. — Tehnika ja Tootmine, 1961, 7, стр. 26.
18. R. Uustalu, Lihtne, kuid küllalt kvaliteetne akustiline agregaat. — Tehnika ja Tootmine, 1960, 10, стр. 37.

Средняя длина слова

1. A. Isotamm, Raadioamatööri käsiraamat. — Tallinn, 1958, стр. 259—262.
2. H. Pedusaar, Elektro- ja raadiotehnika algajaile. — Tallinn, 1959, стр. 215—220, 278—282, 316—324.
3. A. Sillart, Televisioonitehnika. — Tallinn, 1961, стр. 210—216.

Средняя длина фразы

1. E. Alt, E. Jakoobi, Raadiokorrastaja käsiraamat. — Tallinn, 1960, стр. 5—89.
2. A. Isotamm, Raadioamatööri käsiraamat. — Tallinn, 1950, стр. 5—81.
3. H. Pedusaar, Elektro- ja raadiotehnika algajaile. — Tallinn, 1959, стр. 7—65, 179—246.
4. H. Pedusaar, Raadiopraktika. — Tallinn, 1961, стр. 81—104, 118—183.

Позиционные вероятности графем на радиоэлектронных текстах

1. R. Hollmann, Uldine elektrotehnika. — Tallinn, 1954, стр. 129—133, 745—751.
2. H. Pedusaar, Elektro- ja raadiotehnika algajaile. — Tallinn, 1959, стр. 179—186.

Ложные шаги алгоритма на радиоэлектронных текстах

1. A. Isotamm, Raadioamatööri käsiraamat. — Tallinn, 1950,

стр. 5—148 (I ложный шаг),
стр. 5—61 (III ложный шаг),
стр. 5—112 (V ложный шаг),
стр. 13—80 (VII ложный шаг),
стр. 13—79 (VIII ложный шаг),
стр. 13—144 (IX ложный шаг)

2. H. Pedusaar, Elektro- ja raadiotehnika algajaile. — Tallinn, 1959,

стр. 108—183 (II ложный шаг),
стр. 7—71 (IV ложный шаг),
стр. 7—61 (VI ложный шаг).

Истинный шаг алгоритма на радиоэлектронных текстах

1. H. Pedusaar, Elektro- ja raadiotehnika algajaile. — Tallinn, 1959, стр. 13—16.

Позиционные вероятности графем на публицистических текстах

1. Газета «Edasi», 1964, 10 (4350), стр. 1, Lühidalt vabariigist. — Tiheneb side teooria ja praktika vahel. —
2. Газета «Edasi», 1964, 18 (4358), стр. 2, Sõna on meie väliskommentaatoril. — Mida arvate kunstiteoste laenutamisest. —
3. Газета «Edasi», 1964, 20 (4360), стр. 3.
4. Газета «Edasi», 1964, 21 (4361), стр. 1, Võtkem paber ja pliiaats. Arvutagem! — Sünniaeg: detsember 1963. —
5. Газета «Edasi», 1964, 25 (4365), стр. 3.

Ложные шаги алгоритма на публицистических текстах

1. Все статьи, использованные для вычисления позиционных вероятностей графем.

2. Газета «Edasi», 1964, 24 (4364), стр. 1—4.
3. Газета «Edasi», 1964, 35 (4375), стр. 1—4.

Истинный шаг алгоритма на публицистических текстах

1. Все статьи, использованные для вычисления позиционных вероятностей графем.

2. Газета «Edasi», 1964, 10 (4350), стр. 1—3.
3. Газета «Edasi», 1964, 28 (4368), стр. 1—3.
4. Газета «Edasi», 1964, 36 (4376), стр. 1—2.
5. Газета «Edasi», 1964, 37 (4377), стр. 1—3.
6. Газета «Edasi», 1964, 68 (4408), стр. 1—3.

ESIMESE MORFOLOOGILISE TÕUBI PARADIGMA SAAMINE ERINEVATE ALLKEELTE KORRAL EESTI MORFOLOOGIA STATISTILIS-KOMBINATOORSEL MODELLEERIMISEL

Resümee

Käesolev töö on eesti keele erinevate allkeelte võrdleva uurimise esimene katse. Töö aluseks on võetud raadioelektronika ning publitsistika-alased tekstid, mis on töödeldud vastavalt keele statistilis-kombinatoorse modelleerimise algoritmile. Võrreldakse grafeemide tinglikke tõenäosusi positsioonide järgi sõnas, tähtede korrelatiivsete funktsioonide hulka (ülemise allhulga jaoks), algoritmi «väärsumme», algoritmi «summe».

Põhilises osas on kasutatud ainult algoritmi käiku, algoritmi «väärsummade» korral ning informanthulga puhul aga lingvistilist kommenteerimist.

Töö lõpus on antud üldistavad järeldused eesti keele tüpologia osa kohta keele statistilis-kombinatoorse modelleerimise algoritmi täiendamises (püsiva klassi mõiste sõltuvus keele tüpoloogiast) ning samuti ka eesti keele ja indoeuroopa keelte tüpoloogiliste erinevuste kohta.

OBTAINING OF THE PARADIGM FOR THE FIRST MORPHOLOGICAL TYPE OF ESTONIAN LANGUAGE BY STATISTICAL-COMBINATORIAL MODELLING IN CASE OF DIFFERENT SUBLANGUAGES

Summary

The present paper is a first attempt in the comparative study of different sublanguages in Estonian language. The work is based on radioelectronics and publicistics texts treated according to the algorithm of the statistical-combinatorial modelling of a language.

The following problems have been studied comparatively: relative changes in grapheme occurrences according to their positions in a word, the values of correlative function for different letters (within the upper subsets), «false» steps of the algorithm, «decisive» steps of the algorithm.

In case of «false» steps of the algorithm as well as concerning the informant letter, linguistic commentaries have been supplied.

The paper contains generalized conclusions on the contribution of the typology of Estonian language to improving the algorithm of the statistical-combinatorial modelling (the dependence of the notion of an invariable class on the typology of a language). Conclusions are also drawn on typological differences between Estonian and Indo-European languages.

ОГЛАВЛЕНИЕ

I. ЗАВИСИМОСТЬ ХОДА АЛГОРИТМА ОТ ПОДЪЯЗЫКА. ХОД АЛГОРИТМА ПРИ РАЗЛИЧНЫХ ПОДЪЯЗЫКАХ.	3
1. Условные вероятности графем по позициям.	6
2. Независимые вероятности графем.	6
3. Множество коррелятивных функций букв по убывающей величине.	7
II. ЛОЖНЫЕ ШАГИ АЛГОРИТМА.	9
1. Ложные шаги, полученные на радиоэлектронных текстах.	9
2. Ложные шаги, полученные на публицистических текстах.	22
3. Причины возникновения ложных шагов.	24
III. СОДЕРЖАНИЕ ИНФОРМАНТНОГО МНОЖЕСТВА И ПОРЯДОК ВЫДЕЛЕНИЯ ИНФОРМАНТНЫХ АФФИКСОВ ПРИ ИСТИННОМ ШАГЕ АЛГОРИТМА.	25
1. Истинный шаг алгоритма, полученный на радиоэлектронных текстах.	25
2. Лингвистические комментарии к I морфологическому типу (на радиоэлектронных текстах).	45
3. Истинный шаг алгоритма, полученный на публицистических текстах.	45
4. Лингвистические комментарии к I морфологическому типу (на публицистических текстах).	60
IV. ОБЩИЕ ВЫВОДЫ.	62
Список текстов, обработанных алгоритмом статистико-комбинаторного моделирования языка.	65
Esimese morfoloogilise tüübi paradigma saamine erinevate allkeelte korral eesti morfoloogia statistilis-kombinatoorsel modelleerimisel. Resümee	67
Obtaining of the paradigm for the first morphological type of Estonian language by statistical-combinatorial modelling in case of different sublanguages. Summary	68

Тартуский государственный университет
ЭССР, г. Тарту, ул. Юликооли, 18

Х. А. Хольм

**ВЫДЕЛЕНИЕ ПАРАДИГМЫ ПЕРВОГО
МОРФОЛОГИЧЕСКОГО ТИПА НА РАЗЛИЧНЫХ
ПОДЪЯЗЫКАХ ПРИ СТАТИСТИКО-
КОМБИНАТОРНОМ МОДЕЛИРОВАНИИ
ЭСТОНСКОЙ МОРФОЛОГИИ**

На русском языке, резюме на эстонском
и английском языках

Ответственный редактор Х. Рятсеп

Корректор А. Правдин

Сдано в набор 9/1 1965. Подписано к печати
13/V 1965 г. Форм. бум. 60×90, 1/16. Печ. л. 4,5.

Учетно-издат. л. 4,5. Тираж 500 экз. МВ-04670.

Заказ № 196.

Типография им. Ханса Хейдеманна, ЭССР,
г. Тарту, ул. Юликооли, 17/19. I.

Цена 30 коп.