

TARTU RIIKLIKU ÜLIKOOLI

# TOIMETISED

УЧЕННЫЕ ЗАПИСКИ  
ТАРТУСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА  
ACTA ET COMMENTATIONES UNIVERSITATIS TARTUENSIS

518

LINGVOSTATISTILISI UURIMUSI  
SOOME-UGRI KEELTE ALALT

ЛИНГВОСТАТИСТИЧЕСКИЕ ИССЛЕДОВАНИЯ  
ПО ФИННО-УГОРСКИМ ЯЗЫКАМ

TÖID KEELESTATISTIKA ALALT

V

ТРУДЫ ПО ЛИНГВОСТАТИСТИКЕ

TARTU RIIKLIKU ÜLIKOOLI TOIMETISED  
УЧЕННЫЕ ЗАПИСКИ  
ТАРТУСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА  
ACTA ET COMMENTATIONES UNIVERSITATIS TARTUENSIS  
ALUSTATUD 1893.a. VIHK 518 ВЫПУСК ОСНОВАНЫ В 1893.г.

LINGVOSTATISTILISI UURIMUSI  
SOOME-UGRI KEELTE ALALT

ЛИНГВОСТАТИСТИЧЕСКИЕ ИССЛЕДОВАНИЯ  
ПО ФИННО-УГОРСКИМ ЯЗЫКАМ

TÖID KEELESTATISTIKA ALALT

V

ТРУДЫ ПО ЛИНГВОСТАТИСТИКЕ

TARTU 1980

Toimetuskolleegium: U. Kaasik, J. Soontak (vastutav toimetaja), J. Tuldava (esimees), A. Valmet, T.-R. Viitso, A. Villup

Редакционная коллегия: Ю. Каазик, Я. Соонтак (отв. редактор), Ю. Тулдава (председ.), А. Валмет, Т.-Р. Вийтсо, А. Виллуп.

Kogumiku "Tõid keelestatistika alalt" viies väljaanne on pühendatud uurimustele soome-ugri keelte statistika alal. Prof. P. Saukkonen (Oulu) annab ülevaate soome keele kohta ilmnud keelestatistilistest töödest Soomes. Dr. F. Papp (Debrecen) kirjutab samalaadilisest uurimistööst ungari keele alal Ungaris. Dr. T. Zsilka (Nitra) uurib ungari laulekeele morfoloogiat keelestatistilisest aspektist. J. Tuldava (Tartu) käsitleb oma artiklis eesti keele sõnavara foneetilis-grafeemseid mõtteid (tähtede ja häälikute, sealhulgas algus- ja lõpptähtede, resp. -häälikute, sagedusi tekstis ja sõnastikus). Ü. Kaasiku (Tartu) ja J. Tuldava ühisartiklis vaadeldakse sõnalõpu grafemotaktikat eestikeelses tekstis, mis võib osutada kasulikuks teksti automaattöötluse küsimuste lahendamisel.

Пятый выпуск сборника "Труды по лингвостатистике" посвящен исследованиям по статистике финно-угорских языков. Проф. П. Саукконен (Оулу) дает обзор о лингвостатистических работах по финскому языку в Финляндии. Д-р Ф. Папп (Дебрецен) пишет о лингвостатистических исследованиях по венгерскому языку в Венгрии. Д-р Т. Зилка (Нитра) исследует морфологическую структуру поэзии (на материале венгерского языка) с помощью лингвостатистических методов. Доц. Ю. Тулдава (Тарту) рассматривает фонетико-графемные измерения эстонской лексики (частотность букв и звуков, в том числе начальных и конечных букв и звуков в тексте и словаре). Доц. Ю. Каазик (Тарту) и Ю. Тулдава в своей совместной статье исследуют графемотактику конца слова эстонского языка с целью решения некоторых проблем автоматической переработки текста.

This fifth issue of "Works on Linguostatistics" is devoted to the statistical study of Finno-Ugric languages. Professor P. Saukkonen (Oulu) gives a survey of linguostatistical studies of the Finnish language published in Finland. Dr. F. Papp (Debrecen) writes on analogical works on the Hungarian language in Hungary. Dr. T. Zsilka (Nitra) investigates the morphology of poetic language with the help of statistical methods (on the material of the Hungarian language). J. Tuldava (Tartu) presents an analysis of the phonetic-graphemic parameters of Estonian lexica (the frequency of letters and sounds as well as the initial and final letters and sounds in text and vocabulary). The article by Ü. Kaasik and J. Tuldava (both Tartu) deals with the word-ending graphemotactics of Estonian from the point of view of automatic text analysis.

## STATISTICAL LINGUISTIC RESEARCH IN FINLAND

=====

Pauli Saukkonen

Linguistic research in Finland began to utilize statistical methods more widely in the 1960s when synchronous linguistics gained more ground. The statistical tendency was related to the influence and expansion of automatic data processing on the one hand and social and behavioural sciences on the other. It is significant that the first statistical accounts of language in many fields were written by educationalists who had a definite applicatory goal. Prior to the 1960's, only a few detached calculations of various linguistic phenomena existed. In his book Suomen kielen rakenne ja kehitys, Lauri Hakulinen presented some calculations of the structure of Finnish phonology and vocabulary. The first noteworthy scholar to apply quantitative methods was Arvid Genetz, who, as early as the late 19th century, used statistical calculations to support his comparative conclusions.

In the contemporary monographs, quantification of phenomena is rather the rule than an exception. Hence, for example, all the doctoral dissertations published in the 1970's contain statistics of some kind. The doctoral dissertation by Terho Itkonen in 1964 has served as a model in this respect. In the following review, however, I shall label as statistical linguistic studies only those comprehensive papers or plans in which statistical methods receive the main emphasis. I shall classify them into phonetic-phonological, morphological, lexical, syntactic and textual studies.

### Phonetic-phonological studies

Phonetics was the branch of linguistic research first to adopt the exact numerical methods of natural sciences and it has, quite understandably, also, the longest history of statistical studies. Antti Sovijärvi in his doctoral disserta-

tion Die gehaltenen, geflüsterten und gesungenen Vokale und Nasale der finnischen Sprache (1938) already applied advanced statistical formulas to his calculations. Of the subsequent investigators, Antti Iivonen and Jaakko Lehtonen have been particularly interested in the statistical treatment of phonetic phenomena.

The following studies of phonological phenomena based on statistical treatment can be listed in chronological order. Švaa-analyysi savolais- ja kaakkoismurteiden rajalta (The schwa analysis at the boundary between the Savo dialects and the southeastern dialects) by Pauli Saukkonen (1967) examines the variation of the schwa vowel from the sociolinguistic point of view. Pentti Leino has made a comprehensive statistical analysis of folklore, *Strukturaalinen alkusointu suomessa* (Structural alliteration in Finnish) (1970), and written a paper on alliteration in the proverbial sayings of Norwegian Lapps (*Kalevalaseuran vuosikirja* 51, 1971). Heikki Leskinen has studied the dialects of Karelian immigrants and their intermixture, using a phonological approach and quantitative methods (*Virittäjä* 1971, 1974), and he is conducting a wider project on this subject in the University of Jyväskylä. Jaakko Pesonen's paper *Sanamuodot ja niiden kirjainrakenne suomenkielisessä sanomalehtitekstissä* (Word forms and their graphemic structure in Finnish newspapers) (Department of special education, Jyväskylä 1971) contains information on the frequency of Finnish graphemes and some grapheme combinations. Vilho Setälä has compiled corresponding grapheme and phoneme statistics from the Finnish translation of the New Testament and some other reference texts (*Suomen kielen dynamiikka*, *Suomi* 116:3, 1972). Matti Pääkkönen has written the most exhaustive account of the distribution of graphemes in standard Finnish, using the material of fiction, radio language, press language and non-fiction (3, 130, 382 graphemes) collected by the Department of Finnish and Lappish, University of Oulu. Some general statistics from this work were published in *Suomalais-ugrilaisen Seuran Aikakauskirja* 72 (1973), but the bulk of the results, which deal with the in-

terdependence between grapheme frequencies and extralinguistic context in 58 different text classes, await publication. The most recent studies of the phonological structure of standard Finnish are those of Jorma Kajava, Erkki Saarela and Jorma Kekäläinen, the authors of Suomen kielen diagrammeihin perustuva optimikoodi (Diagrammatic optimal code of Finnish), a report of the Department of Instrumental Engineering, University of Oulu (1977), and Bernoulli and Markov source entropies and redundancies of the Finnish language, a report of the Department of Mathematics (1978), as well as that of Kaisa Häkkinen, whose article in Sananjalka 19 and licenciate thesis (Turku University 1978) also include a comparison between the phonological structures of Finnish, Hungarian and Vogul.

Manuscripts containing statistical analyses of phonological and other aspects of language have been and will continue to be written by students of several universities. It is here only possible to refer to the bibliography *Fennistiikan opinnäytteitä* (Student papers in Fennistic studies) edited by Juhani Paronen (Castrenianumin toimitteita 16, 1977).

### Morphological studies

The first notable morphological study utilizing statistical theory, particularly information theory, was *Die Entwicklung der morphologischen Methode im Lappischen* by Mikko Korhonen (Finnisch-ugrische Forschungen XXXVII, 1969). In accordance with the project being carried out in the Department of Finnish and Lappish, University of Oulu, methods of automatic data processing have been applied to comparative studies of the frequencies of inflection and word formation in the different text classes of standard Finnish by students of Finnish. The results of these studies are being published in a series of reports by the department. The compound analysis of the material provided by *Nykysuomen sanakirja* (Dictionary of Current Finnish), also accomplished as student assignments, has been summarized by Pauli Saukkonen in *Suomalais-ugrilaisen Seuran Toimituksia* 150 (1973). Seppo Räsänen has

published his statistical observations on the use of local cases by Finnish poets of the 1950's in *Sananjalka* 16 (1974). The frequencies of single inflectional forms in different text classes can be seen from *Nykysuomen saneiston yleisyys-tilastoa saneenloppuisessa aakkosjärjestyksessä* (Statistics of the word occurrences of present-day Finnish in a reverse alphabetical order), a concordance by Pauli Saukkonen only available in a few complete copies (Oulun yliopiston suomen ja saamen kielen laitoksen tutkimusraportteja 9, 1977), which contains more than 100,000 different word forms. Frequency studies of the stocks of morphemes used by different groups of speakers will also be published in this series. In the University of Joensuu, Alpo Räsänen is pursuing studies on the translative verbs of Finnish, including the quantitative examination of translative verbs with different derivational suffixes. In *Kotimaisten kielten tutkimuskeskus* (The research center of domestic languages), Tuomo Tuomi is working on a statistic-phonotactic structural study of the word stems of standard Finnish based on the reverse dictionary compiled by him.

### Lexical studies

Statistical studies are most numerous in the lexical field. The first large-scale, purely statistical lexical studies of Finnish were published by Kalevi Kajava and Sirkka-Liisa Särkilähti. Kajava published frequency analyses of the vocabulary in elementary textbooks in *Virittäjä* 1962. Särkilähti determined the part of speech distribution and the type/token ratio of the vocabularies of different writers and the Kalevala (*Acta Universitatis Tamperensis* A 18 and 26, 1967 and 1969, *Virittäjä* 1969), utilized discriminatory analysis (in her doctoral dissertation *Marja-Liisa Vartiön kertomataide* (The narrative art of Marja-Liisa Vartiö) 1973), and applied correlation and factor analyses (*Mainostaja* 1973). Minor statistics have also been compiled by Marjatta Reimu (*Virittäjä* 1963).

Pertti Virtaranta was the first to find out word frequencies in spoken Finnish (Sanojen ja omistusliitteisten muotojen esiintymistiheydestä hämäläismurteissa (Frequency of words and possessive forms in the Häme dialects), Acta Universitatis Tamperensis A 26, 1969). Opettajan sanastokirja (Teacher's vocabulary book) (1970) was the first actual frequency dictionary of Finnish. It is based on written texts produced by school children, and it has been compiled by Juhani Karvonen, Kyllikki Röman, Annika Takala and Oiva Ylimentalo. School compositions also form the basis of the study by Heikki Leskinen, Ilkka Savijärvi and Tauno Särkkä published in Jyväskylän yliopiston suomen kielen laitoksen julkaisuja n:o 8 (1974), which deals with the development of the literary expression of 16 pupils throughout their entire school attendance from the viewpoints of, e.g., part of speech distribution and lexical variety, turnover and uniformity. Irmeli Pääkkönen surveyed the vocabulary of beginning fifth-graders on the basis of the secondary-school entrance examination, publishing her findings in Äidinkielen Opettajain Liiton vuosikirja XXII (1975) and her licentiate thesis (Oulun yliopisto 1973).

The word lengths and the quantitative relations between parts of speech emerged as the central factors in the studies of readability first introduced in Finland by Osmo A. Wiio. The book on advertising language by Jarl Louhija also uses the readability approach (Kauppakorkeakoulun julkaisuja C:11:14, 1972). Variations in the popularity of different Christian names were statistically analyzed by Pentti Leino (Kalevalaseuran vuosikirja 52, 1972). Part of speech distributions are further presented by Esko Vierikko in his book Puhekieli ja ilmaisu (Spoken language and expression) (1972) and his licentiate thesis on parliamentary language (Oulun yliopisto 1974). Seppo Räsänen calculated certain word frequencies in Kullervo and Nummisuutarit by Aleksis Kivi (Acta Universitatis Tamperensis A 63, 1975).

Since the 1960's, the Department of Finnish and Lappish in the University of Oulu has been working on a comprehensive comparative and quantitative lexical analysis of 58 different text classes of standard Finnish. Student papers have been written covering, e.g., the different parts of speech. Since 1975, results of different studies have been published as compact statistics in a series of reports by the department. Two frequency dictionaries have been compiled using the same material. A student's Glossary of Finnish by Michael Branch - Antero Niemikorpi - Pauli Saukkonen (1978) contains the 2000 most frequent words of the language of the press and fiction translated into English, French, German, Hungarian, Russian and Swedish. Suomen kielen taajuussanasto (Finnish frequency dictionary) by Pauli Saukkonen, Marjatta Haipus, Antero Niemikorpi and Helena Sulkala (1979) contains frequency information on the 12,000 most common words of the language of fiction, the radio, the press and non-fiction. The word frequencies of the whole material of over 400,000 word occurrences are presented for the different word occurrences in the report no. 9 by Pauli Saukkonen mentioned above.

In 1979 Osmo Ikola will publish a paper on the relative frequencies of the parts of speech in the Finnish dialects and written Finnish. His basic material consists of the dialect texts collected from 13 localities where the southwestern dialects are spoken and kept in the syntactic files of the University of Turku, and his ultimate total material comprises over 100,000 words. The syntactic files coded and stored according to the methods of automatic data processing under the supervision of Ikola permit statistical lexical and morphological studies of many kinds.

In the same way, the materials processed for automatic output in Kotimaisten kielten tutkimuskeskus, such as Index Agricolaensis and the microcards of current Finnish (two corpora of approximately half a million word occurrences of press and literature language) prepared under the guidance of Esko Koivusalo, and the corrected topographic general index

of the dictionary of Finnish dialects (over 400,000 entries) by Tuomo Tuomi provide data ready to be used in statistical analyses. Tuomi is preparing a statistical study of the lexical uniformity of the dialects on the basis of the last mentioned material. Esko Koivusalo also intends to carry out statistical calculations of neological vocabulary.

### Syntactic studies

Large-scale syntactic projects with statistical objectives are being conducted in the universities of Tampere and Turku. Paavo Siro was the one to initiate computerized analysis of the syntax of standard Finnish in the University of Tampere in 1960's. The material consists of 500,000 words of periodical and newspaper texts analyzed for 20 variables. At the same time, Osmo Ikola founded the files of dialectal syntax in the University of Turku. The material comprises 120 one-hour recordings transcribed and provided with morphological and syntactic codes and a reference material of 15 written texts. A more accurate description of the files founded in 1967 and their potential uses is given in the article by Ikola to be published in Proceedings of the XIIth International Congress of Linguists. The third large project is the aforementioned project of standard Finnish launched in the University of Oulu in 1967, in which syntactic analysis is emphasized equally to lexical and morphological analysis. The first syntactic summary published in the series of reports by the Department of Finnish and Lappish was the study of coordination by Marjatta Haipus in 1975. The material of press language and the material of parliamentary language collected by Esko Vierikko were used to prepare a sentence formula survey with its statistics and illustrations on the basis of part of speech - sentence component combinations.

The first statistical analysis of the sentence structures of Finnish was, however, the study *Lausetajun kehityksestä* (On the development of sentence perception) by Inkeri Laurinen (1955), which dealt with the syntactic development of chil-

dren's language. The next was Göran Karlsson's statistical analysis of number congruence between the subject and the predicate (*Congressus secundus internationalis fenno-ugratarum I*, 1965). The abovementioned study by Heikki Leskinen, Ilkka Savijärvi and Tauno Särkkä, *Koululaisten kirjallisen ilmaisutaidon kehityksestä* (1974), provides a quantitative description of the structure of written compositions as sections, sentences, clauses and word occurrences. Tauno Särkkä is continuing the study of the development of the literary expressive skill of pupils and school instruction, using material collected in the whole country as well as textbook material, and examining sentence and clause structures in particular. Ilkka Savijärvi has made quantitative observations on redundance associated with negation in Finnish dialects (*Jyväskylän yliopiston suomen kielen ja viestinnän laitoksen julkaisuja 14*, 1977). The textual project being carried out by Auli Hakulinen and Fred Karlsson also involves some syntactic points.

### Textual study

Text-linguistic analyses of the composition of a text from sentences are so far scarce, which explains the absence of statistical analyses in this field. There is, however, a textual project supported by the Academy of Finland being carried out by Fred Karlsson and Auli Hakulinen as well as Maria Vilkuna in Turku. The material consists of 15,000 sentences of Finnish press and non-fiction texts, which are to be analyzed by clauses for 59 syntactic and/or textual parameters. The objective is a computerized output and cross-tabulation of the frequencies of various items. Olli Järvikoski, on the other hand, has approached text from a psycholinguistic and semantic point of view: he uses a group of test subjects and the methods of factor analysis in studying the dimensions of text perception on the basis of the subjective rating scales proposed by Carroll and partly also Osgood (*Rakenteita*, Turun yliopisto 1978).

Stylistic studies are typically quantitative studies of texts. Stylistics, however, aims at quantification of all the afore-said components of text from the phonological to the textual. (It might be pointed out here that a review of the quantitative methods of stylistics has been written by Sirkka-Liisa Särkilähti (AFinLa:n julkaisuja 19, 1977).) The studies listed above may hence include a stylistic aspect. It is the project of the University of Oulu in particular which aims ultimately at discriminating and describing styles with quantitative methods. The situation in sociolinguistically oriented studies is the same. The project of spoken language, which is being conducted by Heikki Paunonen in the University of Tampere and also carried out in the universities of Helsinki, Jyväskylä and Turku, is similar in this respect.

In conclusion it can be said that statistical analysis of the Finnish language has been undertaken in all the pertinent institutions in Finland in the 1970's. The statistical methods have not been applied equally to the other Finno-Ugric languages. Methodologically, the works have generally been simple frequency calculations. In the future, it will be possible to use them as the basis for more advanced techniques. Not all the means available have been used yet, nor has the computer been utilized in a versatile manner. The most advanced linguistic computer programs in Finland have been produced by Kimmo Koskeniemi (see e.g. Kielitieteellinen tietojenkäsittely LINUS-ohjelmiston avulla, Helsingin yliopiston laskentakeskus. Research Reports 2, 1978). It is possible that the statistical research into language, which is popular at present, will, however, lose ground in the future, because adequate basic information will soon be secured, on the one hand, and linguists will grow tired of such research, on the other. This is already taking place in the model disciplines.

## СТАТИСТИКО-ЛИНГВИСТИЧЕСКИЕ ИССЛЕДОВАНИЯ В ФИНЛЯНДИИ

Паули Саукконен

### Резюме

Статистические методы стали более широко применять в лингвистических исследованиях в Финляндии в шестидесятых годах, в период бурного развития синхронической лингвистики. Начало лингвостатистических исследований было тесно связано с распространением и влиянием идей автоматической переработки информации, а также с влиянием социальных и поведенческих наук. Характерно еще то, что первые лингвостатистические работы были выполнены педагогами, которые руководствовались определенными прикладными целями.

В данном обзоре к лингвостатистическим исследованиям причислены лишь те работы, в которых доминируют статистические методы. Эти работы подразделены на фонетико-фонологические, морфологические, лексические, синтаксические и текстовые исследования. В итоге можно сказать, что статистический анализ финского языка проводится в семидесятых годах во всех соответствующих научно-исследовательских центрах Финляндии, и особое внимание в них уделяется лексической статистике. В исследованиях по другим финно-угорским языкам статистические методы применялись финскими лингвистами редко, причем подобные работы представляют собой простые частотные вычисления.

## ЛИНГВОСТАТИСТИКА И ВЕНГЕРСКИЙ ЯЗЫК

Ференц Папп

Q. В данной статье мы попытаемся решить две задачи: ознакомить а) с историей лингвостатистики в Венгрии, б) представить некоторые статистические количественные показатели на разных уровнях венгерского языка. Задачу а) мы считаем решенной в основном тем, что в качестве приложения дадим довольно длинный список литературы - работы венгерских ученых и практиков в этой области (там же помещены и некоторые, немногочисленные, иностранные исследования, касающиеся венгерского языка в этой области). Русский перевод каждой единицы (если она написана по-венгерски) считается достаточным ориентиром для читателя, интересующегося этой проблемой. В результате этого основной текст статьи посвящен главным образом решению задачи б); план изложения материала по этому вопросу изображен схематически на табл. I. При этом, как в случае а), так и случае б) нужно сделать одну серьезную оговорку: работы и результаты, касающиеся стилистической статистики, не рассматриваются по двум причинам: I) это почти что особая область внутри самой лингвостатистики с многочисленными работами; 2) уже имеется сравнительно новая работа, в первой сноске которой приводится богатая библиография по данному вопросу, см /0/.

Таблица I

Подход \ Область	Диахрония		Синхрония	
	текст	инвентарь	текст	инвентарь
Звуки, фонемы, буквы	I	2	3	
Лексемы, слово-формы, словоупотребления		4	5	6
	7	8	9	10
Морфология	II		12	13
Синтаксис			14	

I. Какова характеристика венгерских текстов в историческом плане на нижнем уровне, т.е. на уровне звуков-фонем? Можно ли количественно-статистически утверждать что-либо о венгерских текстах в этом отношении? Как показано особенно в /93/ и /94/, для венгерских текстов характерна постепенная "веляризация". Так, в Надгробной речи (XII в.) насчитывается 40% велярных гласных (100% = все гласные текста), в собрании сочинений поэта XVI в. Б. Балаши - 42%, а в произведениях второй половины XIX в. до наших дней - 48%. К этим глобальным данным необходимо добавить следующее:

(1) "Велярные гласные" - "палатальные гласные": это коренная оппозиция для системы венгерского языка, так как гармония гласных строится именно на этой оппозиции.

(2) У нас сплошные данные (обработка полного собрания стихов некоторых поэтов объемом в несколько миллионов фонем) только со второй половины XIX века; относительно XIII-XV вв. и относительно XVII-XVIII вв. - интерполяция, но можно ожидать, что общая тенденция именно и соответствует веляризации с некоторыми колебаниями в местах, еще точно не измеренных.

(3) Если провести экстраполяцию относительно текстов до XII века, то можно предположить, что древневенгерские тексты были еще более палатальными (в них содержалась еще большая пропорция палатальных гласных), что приводит к антиномии, рассматриваемой несколько более подробно ниже (п. 4). Подобная экстраполяция - единственная возможность, так как вряд ли можно предположить, что вдруг обнаружится большое количество древневенгерских памятников (к тому же венгерская почва - не новгородского типа, где можно найти в хорошей сохранности большое количество грамот).

(4) При велярности - палатальности текстов необходимо учитывать и жанр. Так, поэзия Б. Балаши "более архаична" (= более палатальна), а его переписка "более современная" (= более велярна). В текстах XX в., где в качестве контроля были приведены и тексты газетные, технические и т.п., такая разница уже не наблюдается.

(5) Среди структурно-языковых причин веляризации мы пока предполагаем такие, как появление весьма частотного определенного артикля (как раз в эпоху Балаши), этот артикль велярной гармонии (a, az); появление велярных вариантов прилеп, относящихся как самостоятельные слова к палатальным (этот процесс произошел в результате потери самостоятельности этих

слов и под влиянием гармонии гласных); появление большого количества славянских, немецких, латинских заимствований велярной гармонии (более подробно об этом см. ниже, п. 4).

(6) Процесс веяризации остановился не позже второй половины XIX в. (возможно, и значительно раньше - у нас в этой полосе еще нет данных).

По другим важным показателям (так, в частности, по пропорции гласных к согласным в текстах) еще не найдены подобные диахронно-количественные изменения в истории венгерского языка.

2. Что касается инвентаря фонем в диахронно-количественном плане, то издавна известны некоторые тривиальные факты, не нуждающиеся в статистической обработке вследствие ограниченного количества фонем (ограниченности инвентаря фонем): утрата некоторых из них, появление новых (как /ц/, /ж/, /ф/ и т.д.). Однако в работе /92/ дана оригинальная и многообещающая попытка статистического обоснования надежности этимологического восстановления тех или иных фонем в общей финно-угорской структуре.

3. Как охарактеризовать современное состояние венгерского языка (совокупность венгерских текстов) на уровне фонем?

Из п. I явствует, что с точки зрения велярности-палатальности достигнуто равновесие с некоторым преобладанием палатальных гласных (52% от всех гласных) и что это состояние длится уже по крайней мере сто лет или даже дольше. В неколичественном, структурном отношении это означает, на наш взгляд, что гармония гласных в полном расцвете. (Истинность этого утверждения можно показать на обратном. Если, допустим, оказалось бы, что одних гласных - велярных или палатальных, неважно, - 98% в текстах, в других - 2%, то это в структурном смысле означало бы, что закон гармонии гласных в этом языке не играет слишком большой роли). Если к этому синхронному состоянию добавить тенденцию веяризации и предположить, что эта тенденция еще активна (что однако вовсе не точно), то можно сказать, что венгерский язык развивается в сторону еще большего палатально-велярного равновесия.

Другой важный количественный показатель на этом уровне - пропорция гласных к согласным, показывающая, как мы видели, довольно большое постоянство за истекшие века. В синхронном смысле копий тут ломалось немало: авторам начала века, да порою и наших дней, казалось, что а) в венгерском языке срав-

нительно часто встречаются гласные и б) чем больше пропорция гласных в тексте, тем краше язык. (Ср. в этой связи уже с самими заглавиями /17 и /18/: "красота", "бессмертие"). Нужно указать, что с подобной как бы идеологизацией этих простых цифр мы не встречались в русской лингвостатистической литературе. Однако а) венгерский язык - язык очень маленького народа, дрожащего почти за все время своего существования за "бессмертие"; б) названные работы появились неслучайно после проигранной первой мировой войны со всеми ее последствиями. Из-за причины б) мы бы даже не удивились, если бы, допустим, и в немецкой лингвостатистической литературе появились в эту эпоху некоторые цифры под националистическим лозунгом. В свете новейших исследований /0/, /31/, /43/ и др. мы должны установить в противовес этой традиции следующее:

(1) В венгерских художественных, поэтических текстах пропорция гласных к согласным такая же, или еще несколько ниже, чем, скажем, в русском литературном языке: на основе обработки современных венгерских текстов в несколько миллионов фонем оказалось, что в них 41,38% от всех фонем - гласные, 58,62% - согласные, а в одном отрывке Тургенева пропорция гласных составляет 42,20%.

(2) Однако лингвостатистика-"националисты" явно не учли одного важного фактора - длительность отдельных гласных. Как показывают новейшие экспериментально-фонетические исследования, средняя длина венгерского гласного (суммируя краткие и долгие согласные) примерно вдвое больше, нежели средняя длина русского гласного (суммируя подударные и безударные гласные), так как под одной и той же (или даже несколько меньшей) пропорцией встречаемости венгерских гласных в тексте кроется примерно вдвое больше вокального элемента. (То есть, тем самым мы как бы реабилитировали результаты этих исследователей, но см. дальше).

(3) Оказалось, что в нехудожественных текстах пропорция гласных сигнификативно выше, чем в художественных; что внутри художественных текстов в прозе пропорция гласных выше, чем в поэзии. К тому же в нехудожественных текстах и доля долгих гласных была выше, чем в художественных, т.е. нехудожественные венгерские тексты на самом деле более "вокальны", нежели художественные. Так, если взять соотношение гласных к согласным в полном собрании стихотворений поэта Э. Ади (начало XX в.) за 100%, то другие венгерские поэты в своих стихах близки к

нему: упомянутый поэт XVI в. имеет 99,71%, поэты А. Йозеф и М. Бабич - 101,30%; Тургенев - 105,80%; проза того же Э. Ади (выборка в несколько сот тысяч фонем) содержит 106,38% и, наконец, технические-газетные тексты содержат около 111% гласных. Так что в результате частичной реабилитации старых результатов относительно большей "вокальности" венгерского языка получается обратная картина: оказывается, это нехорошо, а была бы желательна как раз большая пропорция согласных в эстетических целях. Однако мы не намерены делать подобных выводов: мы вовсе не знаем, что такое хорошо и что такое плохо в красоте языка. (Исследования А.П. Буравлева как будто показывают, что представители разных языков как бы "опportunисты": каждому кажется красивым именно то, что часто встречается в его родном языке).

В современном венгерском языке насчитывается 64 фонемн. (Это "кавказское" количество вызвано тем, что как среди гласных, так и среди согласных долгота - фонематический признак, краткие и долгие фонемы считаются особо: /a/ - /a:/, /b/ - /b:/ и т.д.).

В табл. 2 показана частота встречаемости наиболее употребительных фонем у четырех авторов. (Нарушая синхронность этого раздела мы включили сюда также одного автора XVI и одного - XIX века. В общем, как видно, последний и в этом отношении оказывается вполне современным автором). Как явствует, первые три фонемы покрывают одну четверть, первые восемь - половину всего текста в каждом из четырех случаев. Сходная картина наблюдается и в других обследованных нами как художественных, так и нехудожественных текстах; исключения единичны и по объему незначительны. Под количественными данными кроются и здесь, естественно, некоторые структурные особенности. Так, например, фонема /t/ встречается в морфеме винительного надежа прошедшего времени и т.д., фонема /k/ - прилепе множественного числа у существительных, первого лица (ед. и мн. ч.) у глаголов, фонема /ш/- в первом лице ед. ч. в притяжательной форме существительных и у глаголов, и т.л.

4. Пунктир в табл. I между первой и второй строками обозначает, что здесь речь идет о некоторых смежных вопросах, одинаково касающихся фонетического и лексического уровней. Так, здесь мы должны говорить о связи между инвентарем фонем и инвентарем лексем в историческом плане. В этой обла-

Таблица 2

	XVI в. Balassi 100% = 271776			XIX в. Arany 100% = 497955			XX в. Ady 100% = 552788			XX в. Бабиц 100% = 405975		
A	B	C	D	B	C	D	B	C	D	B	C	D
1	e	12,81	12,81	e	10,55	10,55	e	10,02	10,02	e	10,09	10,09
2	a	7,74	20,54	a	9,46	20,01	a	9,07	19,09	a	9,45	19,55
3	n	6,75	27,29	t	6,16	26,16	n	6,01	25,10	t	5,96	25,50
4	t	6,51	33,80	n	5,69	31,85	t	5,73	30,83	n	5,67	31,17
5	m	5,64	39,44	l	5,25	37,10	k	5,55	36,38	l	5,25	36,43
6	l	5,22	44,66	k	4,59	41,69	l	5,15	41,52	k	4,98	41,41
7	k	4,92	49,58	r	4,39	46,09	m	4,87	46,39	m	4,61	46,02
8	i	4,62	54,20	m	4,39	50,47	o	4,60	51,00	r	4,42	50,44

Примечания к таблице:

A = порядковый номер по убывающей частоте, B = фонема, C = процент, D = кумулятивный процент. Цифры под названием авторов означают количество обработанных фонем. В случае Балашши было обработано полное собрание его сочинений, а в случае Араия - три его поэмы (Тольди, Вечер Тольди, Потерянная конституция) и все его маленькие стихотворения, в случае Ади и Бабица - полное собрание их стихотворений.

сти есть интересное исследование /85/. Автором предлагается оригинальный метод применения исчисления вариаций и тем самым установления необходимого количества фонем для создания того или иного словарного состава. При этом в исчислении учитываются также известные ограничения в применении фонем при создании лексем (последний слог открытий, нет согласных групп, состоящих из трех или более элементов, гласные подчиняются гармонии гласных и т.п.), характерных для финно-угорской системы. Так, получается, что для системы, в которой предлагается гармония гласных и имеются лексемы двусложные, для создания (построения) 8000 лексем необходимо иметь 13 фонем; если усложнить систему тем, что в конце слова может стоять только один из двух гласных и предполагается наличие 24 000 лексем, то для этого нужно иметь 21 фонему и т.д.

В этом случае можно поставить и гораздо более простые вопросы. Так, по работам /31/, /79/, /92/ нам известно, что среди 6000 корневых слов, зарегистрированных в современном венгерском языке, распределение по происхождению примерно таково: финно-угорского происхождения 10%, результат венгерского развития 8%, турецкого происхождения 5%, славянского - 9%, немецкого - 5% и т.д. В связи со звуковым составом представителей разных этимологических пластов отметим два явления:

(1) Доля гласных в корневых словах разных пластов: латино-греческого происхождения 42%, славянского - 41%, турецкого - 40%, финно-угорского - 40% и т.д. То есть мы вернулись к роковому вопросу о вокальном характере на этот раз не текстов, а словаря. И оказывается, что даже лексемы славянского происхождения в своей совокупности и современной венгерской огласовке содержат большую долю гласных, нежели, допустим, лексемы финно-угорского происхождения.

(2) Доля велярных гласных в разных этимологических пластах: финно-угорский содержит 55%, турецкий - 59%, венгерский внутренний - 65%, немецкий - 72%, славянский - 76%, латино-греческий - 81%; то есть, чем моложе этимологический пласт, тем больше в нем велярных гласных. А в исходном положении финно-угорские и турецкие элементы содержали почти одинаковую долю велярных и палатальных гласных, но откуда тогда предполагаемая нами палатальность древневенгерских текстов? Эту антиномию мы пока не можем разрешить.

5. На грани лексикологии и фонетики-фонологии стоит так-

же синхронный вопрос о длине словоупотреблений в текстах. В табл. 3 показаны некоторые результаты из этой области. Как

Таблица 3

i	Процент словоупотреблений, состоящих из i фонем			
	Балашин 100% = 53889	Арань 100% = 104901	Ади 100% = 115568	Бабич 100% = 88060
1	2,47	8,54	8,88	10,30
2	17,56	12,12	13,23	12,27
3	14,57	14,60	13,42	14,10
4	12,18	12,93	11,47	12,97
5	16,22	17,32	17,61	17,23
6	10,59	11,29	11,90	11,79
7	8,27	9,19	8,44	8,73
8	7,11	6,31	6,57	6,15

Примечания: как и в табл. 2, цифры под авторами означают количество обработанных словоупотреблений, i = количество фонем.

видно из таблицы, словоупотребления по длине распределяются с двумя максимумами. В материале XVI в. эти два максимума помещены у двусложных и пятисложных словоупотреблений, а в текстах "современных" (начиная с XIX в.) — у трехсложных и у пятисложных, покрывающих вместе почти треть всего текста. Средняя длина словоупотреблений в современных текстах колеблется между 4,6 и 4,8 фонем (у Балашина несколько выше — 5,0, но там учитывалось все его наследие, а не только стихи). Ясно, что этот показатель в большой мере зависит от жанра: в современных научных и газетных текстах средняя длина значительно выше — около 6,7 фонем на словоупотребление. Но этот критерий не оказался решающим: средняя длина словоупотреблений в некоторых произведениях нашего поэта-современника Ф. Юаса равна также 6,7. (Этот поэт изображает в исследованных стихотворениях страшные последствия атомной бомбы в длинных сложных словах-существительных и в образованных от них прилагательных). Словоупотребления, состоящие из 9 и более фонем, встречаются и в исследованных нами текстах, но в сокращенной табл. 3 эти цифры не приводятся, они составляют всего несколько процентов исследованных текстов.

6. Количественная характеристика современного словарного состава на этом фонологическом уровне представляет собой распределение длины лексем в словаре. В приложении к работе /40/ приводятся цифровые данные по этому вопросу на основе современного академического толкового словаря венгерского языка. Оказывается, в этом словаре большинство лексем в словарной форме состоит из 6-10 букв (учитывая характер венгерской графики, это несколько меньше фонем), лексемы такой длины занимают почти две трети всего словаря. Максимум у восьмибуквенных лексем - они одни занимают 15% всего состава.

7. Статистическая обработка памятников (как и диалектных текстов) в наши дни должна производиться на вычислительных машинах. В настоящее время в Дебрецене подготовлен к машинной обработке памятник середины XV в. "кодекс Йокаи". Публикация результатов ожидается в ближайшем будущем.

8. Несмотря на расхождения в подходе, в работах /78/ и /80/ ставится отчасти общая задача: связь между такими категориями, как часть речи, морфологический состав и этимология. Оказывается, что в области корневых (простых, неприводных) лексем наиболее древнюю группу и представляют глаголы (почти 40% простых глаголов - финно-угорского происхождения), и наоборот, наиболее подвижная, молодая группа - существительные (среди простых существительных всего 7% финно-угорского происхождения). В отношении действия разных словообразовательных (в широком смысле этого слова) средств языка по частям речи можно сказать следующее. Среди 14 000 глаголов академического словаря 56% образовано префиксацией (с суффиксами или без них), 40% бесприставочных (с суффиксами или без них); большинство (76%) глаголов образовано суффиксами. Среди 31 000 существительных 55% состоит из двух корней (т.е. сложное слово с суффиксами или без них); однокорневых (с суффиксами или без них) насчитывается 34%; 42% существительных не содержат в себе суффикса (при этом они могут быть простые или сложные); 34% имеют суффикс в конце (простого или сложного) слова. Прилагательные занимают как бы промежуточное место: свыше 90% состава простых или сложных из двух корней (только с меньшей пропорцией сложных, чем у существительных); в то же время большинство (76%) образовано суффиксами.

9. Синхронное состояние венгерских текстов на уровне лексем. Сюда относятся весьма традиционные исследования в

области лексикостатистики.

В настоящее время мы располагаем одним частотным словарем венгерского языка, составленным на материале газетных текстов середины 30-ых годов /47/, и частотным списком творчества поэта XX в. Д. Южаса /64/. Ввиду библиографической редкости /47/ и в самой Венгрии, мы разрешаем себе воспроизвести полностью две таблицы из этой книги (см. стр.334 и 321 соответ.). Как видно из табл. 4, автором (прилежным стенографистом) вручную обработаны тексты с общим объемом почти в 400 000 словоупотреблений; в табл. 5 показано, что в этих

Таблица 4

i	Количество словоупотр. с частотой i	Процент ко всем словоупотр.
I	15171	3,81
2 - 4	22317	5,61
5 - 9	20250	5,09
10 - 24	33573	8,44
25 - 49	31981	8,03
50 - 99	39115	9,83
100 - 999	124516	31,29
1000 -	111033	27,90
Всего	397956	100,00

Таблица 5

i	Количество лексем с частотой i	Процент ко всем лексемам
I	15171	48,82
2 - 4	8513	27,39
5 - 9	3130	10,07
10 - 24	2237	7,20
25 - 49	920	2,96
50 - 99	563	1,81
100 - 999	520	1,67
1000 -	24	0,08
Всего	31078	100,00

текстах было обнаружено всего около 30 000 лексем (слов-типов). Вот начало частотного списка (см. стр. 55), в скобках дано количество словоупотреблений, в которых представлена данная лексема: a (артиклъ) (42383), az (артиклъ) (14688), éв 'я' (10746), hogу (союз: 'что, чтобы') (7066), is 'тоже' (3969), nem 'не, нет' (3844), ez 'этот' (2832), egy (неопределенный артиклъ и числительное 'один') (2472) и т. д. Первые по частоте полнзначные слова: magyar 'венгр, венгерский' (1746), nagy 'большой' (1417), pengő 'денежная единица в тридцатые годы' (770), év 'год' (706), tud 'знать, уметь' (691), kormány 'правительство' (689), doktor 'доктор (чаще, очевидно, как "чин", обозначение полностью оконченного университетского образования)', mond 'сказать' (672), új 'новый' (667), akar 'хотеть' (651) и т. д. Видна большая зависимость самых употребительных лексем при довольно большом объеме обработанного материала от жанра (газетные тексты). В этой работе содержится также ценные данные о длине словоупотреблений, о связи между длиной и частотой встречаемости и т. п. только с этим стилистическим ограничением и с ограниченностью лингвистически продуманной структуры. Рассматриваемая работа также не лишена некоторого национализма: автора смущает наличие иностранных слов, в разных аспектах он разбирает их особо. Даже лингвисту нелегко порою решить, что считать иностранным элементом в лексике, еще труднее сделать это лингвисту.

По работе /64/ (и по словарю поэзии Д. Кхаса, составленного тем же автором) известно, что объем словаря (количество лексем) этого поэта равен 11606, а объем написанных им словоупотреблений - примерно 150 000. Частотный список и у него начинается, естественно, с семантически пустых слов: a (около 11100), az (ок. 3900), éв (3370), is (1946), nem (1161) и т. д. Вот наиболее употребительные полнзначные слова в его поэзии: ég 'небо' (649), élet 'жизнь' (619), szív 'сердце' (597), nagy 'большой' (582), szép 'красивый' (563), magyar 'венгр, венгерский' (504). На основе частичной обработки словаря мы располагаем предварительными материалами и относительно некоторых других венгерских поэтов. Так, по /58/ нам известно, что по обработанной выборке наиболее употребительными лексемами у современника Д. Кхаса, А. Тоты (Tóth Árpád) оказались такие, как búв 'печальный', lélek 'душа', jó 'хороший, добрый', halk 'тихий' и т. д.; у Л. Сабо (Sza-

bó lórinс) - tud 'знать, уметь', lát 'видеть', Isten 'бог', pénz 'деньги', akar 'хотеть', szeret 'любить'. На основе первых двух томов словаря Петёфи (на основе полного собрания сочинений, стихи, переводы, переписка и т.д.) его словарь, по нашим подсчетам, будет содержать более 20 тыс. лексем, а его тексты окажутся длиной примерно в 500 тыс. словоупотреблений (ср. с подсчетами-предсказаниями одного из составителей словаря: /56/). Среди наиболее употребительных лексем Петёфи безусловно окажутся такие, как mond 'сказать' (1742), lát 'видеть' (1693), jó 'хороший, добрый' (1518), ember 'человек, мужчина' (1248), isten 'бог' (1175), megy 'идти, gehen' (1045), jön 'идти, kommen' (983) и т.д. Ясно, что все эти данные - предварительные и нуждаются в дальнейшей проверке.

Институтом языкознания Академии наук Венгрии координируется большая коллективная работа по составлению современного частотного словаря венгерского языка (руководитель: Д.Сепе). Общий корпус всего в несколько миллионов словоупотреблений (из разных жанров) планируется на машинную обработку.

Еще одно заключительное замечание относительно использования приведенных в этом пункте статистических данных. Лемматизация для венгерского языка кажется еще труднее, чем для некоторых других, окружающих нас языков. Дело не только в разделяемых приставках (ср. со сходным положением в немецком языке, которая несколько затрудняет только машинную обработку) и в большом количестве сложных слов (ср. также с положением в немецком или английском языке: автора-стенографиста /47/ весьма увлекло и это свойство венгерского языка и его данными иногда трудно пользоваться именно вследствие слишком детального учета этого фактора). Но имеются и дублетные формы (см. выше: a--az, év--a 'и'), свободно варьируемые или зависящие от окружения. Одними авторами они рассматривались как отдельные, другими - как одна лексема; есть ряд падежных окончаний и послелогов, регулярно снабжаемых личными прилепками, они трактуются тоже у разных авторов по-разному. Нам неясно, насколько это влияет на возможное использование статистических данных, на их сравнение с данными, почерпнутыми из других языков; однако осторожность при их использовании необходима.

10. Лексемы как таковые, не в их органической совокупности в текстах, а взятые в инвентаре, в совокупности словаря - тоже могут дать некоторые интересные результаты.

Так, в /63/ мы предлагаем эмпирическую формулу относительно распределения лексем в венгерском академическом словаре по количеству их значений. (Грубо говоря, оказывается, что половина лексем в словаре имеет по одному значению, четверть - по два, одна восьмая часть - по три и т.д.). Интересно было бы посмотреть, каково положение с другими толковыми словарями, составленными для других языков и другими коллективами.

Рассмотрим несколько более сложный вопрос: как связана многозначность лексемы с ее мотивированностью или степенью сложности морфологического строения? В /77/ мы приходим к выводу, что появление суффиксов, словосложение и т.п. (т.е. повышение мотивированности лексемы, сложности ее морфологической структуры) на самом деле приводит к сокращению числа значений. Так, среди глаголов (наиболее многозначная часть речи в нашем словаре) у простых производных среднее количество значений - 6, у бессуффиксных приставочных - 3, у бессуффиксных суффиксальных образований и у суффиксальных приставочных - 2 (т.е. если уже есть в глаголе суффикс, то с точки зрения значений неважно, есть ли еще приставка или нет). Среди существительных простые производные имеют в среднем 2,4 значения; сложные, состоящие из двух корней (с суффиксом или без него, неважно) - 1,3; состоящие из трех корней - 1,2; из четырех - 1,1. При этом почти во всех частях речи в каждой структурной группе преобладают элементы с одним значением (что и соответствует закономерности, изложенной в предшествующем абзаце), но среди простых производных глаголов сравнительно много элементов с двумя значениями, среди приставочных бессуффиксальных - с двумя и тремя значениями.

Имеются вопросы, граничащие здесь с морфологией. Так, оказывается, в таком русском толковом словаре, как словарь Д.Н. Ушакова, процент глаголов сигнификативно выше, чем в венгерском толковом словаре. Вызвано ли это только разными лексикографическими традициями, или кроется здесь и нечто от системы языка (см. категорию вида в русском - тогда, очевидно, глаголы в русском словаре не столь многозначны, как в венгерском, некоторые прямо с "нулем значения", где есть только указание на соответствующую видовую пару). Как выяснилось, без машинной обработки мы плохо знали наш словарь с морфологической точки зрения. Считалось, что есть слова, принад-

лежащие всегда к одной какой-нибудь части речи; встречаются слова, принадлежащие то к одной, то к другой части речи; и существуют, наконец, слова, могущие принадлежать к одной из трех частей речи. Машинная обработка венгерского толкового словаря выявила, что наличествуют также некоторые элементы, которые, по мнению составителей словаря (и, очевидно, по документации, лежащей за каждой словарной статьей), могут принадлежать к одной из четырех частей речи.

11-13. Имеются прекрасные обработки памятников венгерского языка с детально разработанной морфологией; есть также и описательные грамматики венгерского языка в современном его состоянии. Чтобы статистически осмыслить эти данные, нужны опять машины. Мы имеем в виду особенно одну машинную возможность, создание так называемых обратных конкордансов. Обратные конкордансы не очень распространены, вероятно, потому, что для английского языка они мало что дают. При обратном конкордансе, легко образуемом на ЭВМ, каждое словоупотребление (вместе со своим контекстом) расположено в порядке алфавита с конца (с правой стороны). Вследствие этого словоупотребления с одинаковыми окончаниями располагаются непосредственно друг за другом, в одном блоке. Таким образом легко дается количественный подсчет тех или иных окончаний, степень омонимичности того или иного фонетического (буквенного) окончания и т.д. В настоящее время планируется такая обработка упомянутого кодекса Йокаи; у нас в рукописи есть некоторые подобные обратные конкордансы более современных венгерских текстов. (Пользуясь случаем, отметим, что у нас готовится обратный конкорданс к Евгению Онегину, имеющий те же морфологические, морфологически-статистические цели). В нем героический предшественник этих работ составил статистику лексических и грамматических суффиксов и окончаний вручную на основе парламентских выступлений в 41 000 словоупотреблений: /66/. Всего здесь насчитывается 20 000 суффиксов лексических, 5080 суффиксов грамматических и 15 000 окончаний. (Побочный результат: в исследованных парламентских выступлениях средняя длина словоупотреблений равна 2,44 слога). Более детальными результатами этой работы трудно пользоваться вследствие неполной ясности лингвистической концепции автора.

14. Меньше всего можно сказать о статистических работах в области синтаксиса. В синхронном плане имеется большая монография /75/. В ней по разным жанрам, на большом материале

исследуются количественные показатели по длине предложений, по его сложности, структурированности и т.п.

15. В тексте нашей статьи мы не останавливались на многих вопросах лингвостатистики, затронутых некоторыми венгерскими авторами. Здесь только указываем на некоторые области и некоторые характерные черты этих исследований:

(1) Выделяется одна характерная черта лингвостатистических исследований особенно на уровне звуков-фонем-букв современного языка. А именно: их практическая направленность, установка на создание венгерской клавиатуры пишущих машин (см. /21/, /22/, /25/, /34/, но в основном то же содержали и некоторые более ранние работы, заглавия которых этого не отражали: /14/, /16//), на усовершенствования венгерской стенографии (/23/, /24/, /46/, /47/). Сама практическая цель часто оставалась недостигнутой (так, клавиатура венгерских пишущих машин такая же, как для большинства европейских латинописущих языков), но польза для науки безусловна.

(2) Среди общих вопросов, как видно из литературы, часто рассматривался вопрос о "количестве слов (= лексем) в языке", или уже: в венгерском языке (см. /48/, /50/, /51/). Формируется вывод: не менее одного миллиона, но точному подсчету это явление не поддается (из-за неясности "лемматизации", вследствие социолингвистических факторов и т.д., чтобы пользоваться современной терминологией).

(3) Мы выше вовсе не затрагивали таких больших областей, как применение статистических методов в диалектологии, в изучении детской речи, в ономастике (в частности: в антропонимике - см. /54/, /57/, /59/ и в особенности - /60/).

(4) Общие методологические вопросы лингвостатистики как бы меньше интересовали венгерских исследователей (сходное положение наблюдается и в других областях венгерской научной жизни - далеко не только в лингвистике). Есть у нас хорошие специалисты и в этом общем плане (И. Фонадь; очень полезна книга Ф. Нада /13/, вышедшая, однако, весьма малым тиражом), но скорее преобладают, как нам кажется, более маленькие, конкретные исследования, проведенные исключительно вручную, а в настоящее время (начиная примерно с начала 60-ых годов, пионеры: И. Фонадь, Б. Дёмёлки, Т. Сенде - см. /38/) больше пользуются вычислительными машинами.

БИБЛИОГРАФИЯ РАБОТ ПО ВЕНГЕРСКОЙ ЛИНГВОСТАТИСТИКЕ

0. Kemény Gábor, Stilisztika és statisztika (Zsilka Tibor új könyvéből) /Стилистика и статистика. (О новой книге Ж.Т.)/. - НУК. 78: 135-47 /1976/

Общие

1. Fónagy Iván, A hang és a szó hírértéke a költői nyelvben /Информационная стоимость звука и слова в поэтической речи/. - НУК. 62: 73-100 /1960/
2. Papp Ferenc, A matematikai módszerek alkalmazásáról nyelvtudományunkban /О применении математических методов в нашем языкознании/. - I.ОК. 17:289-306 /1961/
3. Fónagy Iván, Beszéd és valószínűség /Язык и вероятность/. - НУТ. 86: 309-20 /1962/
4. Dezső László, A szókincs és nyelvtan statisztikai vizsgálata /Статистическое изучение лексики и грамматики/. - I.ОК. 21:215-33 /1964/
5. Rényi Alfréd, Információelmélet és nyelvészet /Теория информации и лингвистика/. - АНУТ. 2: 245 - 51 /1964/
6. Büky Béla, Beszéd és valószínűség /Язык и вероятность/. - НУТЭрт. 58: 223-9 /1967/
7. Papp, F., Die Bearbeitung des ungarischen Wortschatzes auf Lochkartenmaschinen. ALH. 17: 141-72 /1967/
8. Kelemen József, Magyar szövegek statisztikai feldolgozásának problémái /Вопросы статистической обработки венгерских текстов/. - NyelvfelelDok. 61-73 /1967/
9. Petőfi S. János, A nyelvstatisztikai vizsgálatok néhány kérdése /Некоторые вопросы лингвостатистических исследований/. - NyelvfelelDok. 117-40 /1967/
10. Папп Ф., О машинной обработке одноязычных словарей. - НТИ. № 3.20-9 /1969/
11. Kelemen József, Sprachtypologie und Sprachstatistik. - Theoretical Problems of Typology and the Northern Eurasian Languages. Edited by L. Dezső and P. Hajdu. Budapest 1970. - 53-73.
12. Nagy Ferenc, Korrelációs számítás a nyelvészetben /Исчисление корреляций в языкознании/. - НУТ. 95: 342-9 /1971/
13. Nagy Ferenc, Kvantitatív nyelvészet /Квантитативная лингвистика/. Budapest 1972.

З В У К И, Ф О Н Е М Ы, Б У К В Ы

14. Simonyi Zsigmond, A betűk és hangok előfordulásának aránya /Пропорция встречаемости букв и звуков/. - Ny. 25: 325 /1896/
15. Tolnai Vilmos, Egy néhány számadat a hangokról és betűkről /Некоторые цифровые данные о звуках и о буквах/. - Ny. 35: 421-5 /1906/
16. Simonyi Zsigmond, A betűk számaránya a magyarban /Пропорция букв в венгерском языке/. - Ny. 37: 418- /1908/
17. Tolnai Vilmos, A nyelvek szépségéről /О красоте языков/. - MNy. 17: 28-32 /1921/
18. Tolnai Vilmos, Halhatatlan magyar nyelv /Бессмертный венгерский язык/. - MNy. 20: 5-9 /1924/
19. Mikes Ferenc, Új magyar hangstatisztika /Новая венгерская звуковая статистика/. - Gyvsírástudomány 16.1: 91-2 /1935/
20. Junker, H., Sprachklangbilder. - Ungarische Jahrbücher 11: 460-72 /1936/
21. Mikes Ferenc, Magyar írógépbillentyűzet /Венгерская клавиатура для пишущих машинок/. - Gyvsírástudomány 16.8: 93-6 /1936/
22. Mikes Ferenc, Betűstatisztika és írógépbillentyűzet /Статистика букв и клавиатура пишущих машинок/. - Gyvsírástudomány 17.8: 109-11 /1937/
23. Mikes Ferenc, Szó eleje - szó vége /Начало слова - конец слова/. - Gyvsírástudomány 17.8: 109-11 /1937/
24. Kaiblinger Fülöp dr., Több nyelv gyvsíróinak mérkőzése /Соревнование стенографистов разных языков/. - Gyvsírástudomány 18.2: 20-2 /1937/
25. Mikes Ferenc, A magyar írógépbillentyűzet kérdése /Вопрос о клавиатуре венгерских пишущих машинок/. - Gyvsírástudomány 18.6: 92-3 /1938/
26. Tarnóczy Tamás, A magyar magánhangzók eloszlása /Распределение венгерских гласных/. - MNy. 38: 352-7 /1942/
27. Tarnóczy Tamás, Megjegyzések és kiegészítések a magyar magánhangzók eloszlásának kérdéséhez /Примечания и дополнения по поводу распределения венгерских гласных/. - MNy. 39: 369-74 /1943/
28. Csehély Aladár, A magyar magánhangzók eloszlása /Распределение венгерских гласных/. - MNy. 39: 64-5 /1943/
29. Deme László, Statisztikai vizsgálatok a hangátvetés történetéből /Статистические исследования из истории метатезы звуков/. - MNy. 41: 55-60 /1945-6/
30. Lotz, John, Vowel Frequency in Hungarian. - Word 8:227-35 /1952/

31. Vértés, Edit, Statistische Untersuchungen über den phonetischen Aufbau der ungarischen Sprache. - ALH. 3: 125-58, 411-30 /1953/, 4: 194-224 /1954/
32. Tarnóczy Tamás, A magyar hangzókapcsolatok eloszlása Ady költői nyelve alapján /Распределение групп венгерских звуков на основе поэтического языка Ади/. - NYK. 53: 107-52 /1952/
33. Fónagy Iván, A költői nyelv hangtanából /Из фонетики поэтической речи/. - Budapest 1959
34. Dr. Nemes Zoltán, Betűpárstatisztika az írógépbillen-tűzest számára /Статистика пар букв для клавиатуры пишущих машинок/. - Gyrosírók Lapja 9:3: 74-5 /1959/
35. Deme László, Szókezdet- és szóréteg-statisztikai számítások /Подсчеты по статистике начала слова и разных слоев слов/. - MNy. 56: 458-69 /1960/
36. Fónagy Iván, A szavak hossza a magyar beszédben /Длина слов в венгерской речи/. - Ny. 84: 355-60 /1960/
37. Fónagy, Iván, Die Metaphern in der Phonetik. The Hague 1963
38. Dömölki Bálint-Fónagy Iván-Szende Tamás, Köznyelvi hangstatisztikai vizsgálatok /Звуко-статистические исследования в литературном языке/. - ANyT. 2: 117-32 /1964/
39. Farkas Vilmos, Fonémastatisztikai problémák a nyelvjárgastípus történetében /Вопросы фонемной статистики в истории типа диалекта/. - Budapest 1966
40. Papp Ferenc /szerk./, A magyar nyelv szövegmutató szótára /Обратный словарь венгерского языка/. /Статистические таблицы: стр. 541-91/. Budapest 1969
41. Szende Tamás, Anyanyelvünk hangzásáról /A hangkapcsolatok szegere/ /О звучании нашего родного языка (Роль групп звуков)/. - Ny. 95: 442-53 /1971/
42. Szende Tamás, Spontán beszédanyag gyakorisági mutatói /Частотные показатели спонтанного речевого материала/. Budapest 1973
43. Jékeli Pál-Papp Ferenc, Ady Endre összes költői műveinek fonémastatisztikája /Фонемная статистика полного собрания поэтических сочинений Э. Ади/. Budapest 1974.

См. еще в следующих работах, помещенных в других разделах:  
/ 46 /, / 72 /, / 82 /, / 84 /, / 91 /, / 92 /

### С л о в а, а н т р о п о н и м ы

44. Tolnai Vilmos, Hány szó van a magyar nyelvben? /Сколько слов в венгерском языке?/. - MNy. 14: 33 - 4 /1918/

45. Cser János, A magyar gyermek szókincsé /Словарный запас венгерского ребенка/. Budapest 1939
46. Jakab Lajos hang- és szóstatistikája, Nemes Zoltán dr. és Laszlo János dr. közreműködésével sajtó alá rendezte Traeger Ernő dr. /Статистика звуков и слов, составленная Лайошем Якобом. С помощью Н.З. и Л.Я. подготовил к печати Трэгер Эрне д-р/. Budapest 1939
47. Nemes Zoltán, Szóstatistika egymillió szótagot felölelő újságszövegek alapján /Статистика слов на основе газетных текстов в один миллион слогов/ (Без места, без даты). Szeged 1941
48. Laziczius Gyula, A magyar szókészlet nagysága /Величина венгерского словарного состава/. - MNy. 38: 65-73 /1942/
49. Ambrus Lajos, Arany János költészetének szóstatistikája /Статистика слов поэзии Я.Араня/ (Рукопись) /Budapest 1951/
50. Kelemen József, Hány szó van nyelvünkben? /Сколько слов в нашем языке?/. - NyU. 78: 68-74 /1954/
51. Vértés Edit, Szótározott szavaink száma és a szókészlet nagysága /Количество слов, зарегистрированных в наших словарях, и объем словарного состава/. - NyU. 79: 329-38 /1955/
52. Antal László, Megjegyzések a szóállomány statisztikai vizsgálatáról /Замечания о статистическом исследовании словарного состава/. - NyU. 83: 307-11 /1959/
53. Fodor István, A statisztikai módszer alkalmazásának néhány kérdése /Некоторые вопросы применения статистических методов/. - NyU. 84: 196-205 /1960/
54. Büky Béla, A fővárosi /Budapest/ keresztnévadás hatéből /Тенденции в присвоении имен в столице /Будапешт/ /. - Budapest 1961
55. J. Soltész Katalin, Guiraud statisztikai módszere a szókincs vizsgálatában /Статистический метод Гиро в исследованиях словарного состава/. - ANyU. 1: 263-72 /1963/
56. J. Soltész Katalin, Nyelystatisztikai vizsgálatok a Petőfi-szótár alapján /Лингвостатистические исследования на основе словаря языка Петёфи/. - NyU. 67: 345-54 /1965/
57. Hajdú Mihály, Adatok a XVIII - XX. század névdivatjának földrajzi vizsgálatához /Данные к географическому изучению моды на имена в XVIII - XX вв./ - MNy. 64: 2, 2-45 /1968/
58. Szende Tamás, Szóstatisztikai vizsgálatok Tóth Árpád, Juhász Gyula és Szabó Lőrinc költeményei alapján /Исследования статистики слов на основе стихотворений Т.А., Ю.Д. и С.Д./ - NyU. 92: 287-300 /1968/

59. Berényi Zsuzsanna Ágnes, XV. századi női neveink vizsgálata matematikai módszerekkel /Исследование женских имен XV-го века математическими методами/. - Névtudományi előadások: 134-8 /1969/
60. Kálmán Béla, A nevek világa /Мир имен/. Budapest 1969
61. Papp Ferenc, Az Ertelmezo Szótár legtöbb jelentéssel bíró szavairól /О словах в (венг.) Толковом словаре, обладающих наибольшим количеством значений/. - NyU. 93: 94-7 /1969/
62. S. Kádár Julia, A szókincs és a szófaj gyakoriságának alakulása 2 - 6 éves gyermekek beszédében feladatmegoldás, illetőleg kommunikáció során /Формирование словарного запаса и частоты частей речи в речи детей 2-6 лет в ходе решения задач и коммуникации/. - ANyT. 7: 149-59 /1970/
63. Papp, F.: Sur quelques lois statistiques du vocabulaire. - Communications du X<sup>e</sup> Congrès International des Linguistes. Vol. 4: 475-9. Bucuresti 1970
64. Juhász Gyula költői nyelvének gyakorisági szójegysége. Szerkeztette: Benkő László /Частотный список слов языка поэзии Д.Юхаса. Под ред. Б.Л./ - Acta Academiae Pedagogicae Szegedensis. Első rész. 95-122. Szeged 1973
65. Vértés Edit, Szám adatok az egyedi szóképzlet nagyságának megbecsüléséhez /Количественные данные к приблизительному подсчету объема индивидуального словарного состава/. - NyK. 78: 464-90 /1976/
- См. еще в следующих работах, помещенных в других разделах:  
/ 83 /, / 85 /.

Г р а м м а т и к а : м о р ф о л о г и я и с и н т а к с и с

66. Nemes Zoltán, A magyar nyelv képző-, jel- és ragstatisztikája /Статистика суффиксов, формообразующих суффиксов и окончаний венгерского языка/. - MNy. 28: 30-6, 106-11 /1932/
67. Deme László, A zsugorodás és kiegyenlítődés törvénye a parole-ban /Закон стяжения и уравнивания в речи/. - NyK. 67: 243-72 /1965/
68. Szabó Zoltán, A kalotaszegi nyelvjárás igeképző-rendszeré /Система глагольных суффиксов в диалекте Калотасег/. Budapest 1965
69. Nagy Ferenc, A képzőgyakoriság néhány jellemző vonása /Некоторые характерные черты частоты суффиксов/. - MNy. 63: 331-46 /1967/
70. Papp Ferenc, Néhány fontosabb adat szókincsünk szófaj szerinti megoszlásáról /Некоторые данные о распределении нашего словарного состава по частям речи/. - Nyelvef. Dok. 104-16 /1967/
71. Папп Ф., О некоторых количественных характеристиках словарного состава языка. - Slavica 7: 51 - 6 /1967/

72. G. Varga Györgyi, Alakváltozatok a budapesti köznyelven /Варианты форм в разговорном языке Будапешта/. Budapest 1968.
73. Papp Ferenc, Szófaj és végződés /Часть речи и концовка лексем/. - NyK. 92: 29-37 /1968/
74. Папп Ф., Kétfajta jellemzők egy magyar szókincs szótáránál. - Magyar Nyelv 1970
75. Deme László, Mondatszerkezeti sajátosságok gyakorisági vizsgálata /Изучение частотности некоторых синтаксических свойств/. Budapest 1971
76. Nagy Ferenc, A szóképzés valószínűsége /Вероятность словообразования/. - NyK. 73: 249-53 /1971/
77. Papp Ferenc, A lexémák alaktani szerkezete, szófaja és jelentésgazdagsága /Морфологическая структура, часть речи и количество значений лексем/. - MNy. 68: 23-33 /1972/
78. Papp Ferenc, A lexémák szófaja, alaktani szerkezete és eredete /Часть речи, морфологическая структура и происхождение лексем/. - MNy. 70: 55-68 /1974/

См. еще в следующих работах, помещенных в других разделах: /42/, /47/, /56/, /62/, /87/.

#### История языка и диалектология

79. Vermes Stefánia, A magyar szókészlet etimológiai statisztikája /Этимологическая статистика венгерского словарного состава/. - NyK. 51: 435 /1949/
80. Benkő Loránd, Adatok a magyar szókincs szerkezetének változásához /Данные к изменению структуры лексем в венгерском словаре/. - NyK. 64: 115-36 /1962/
81. Imre Samu, Nyelvjárási jelenségek statisztikai feldolgozása /Статистическая обработка диалектологических данных/. - MNy. 8: 11-26 /1962/
82. Farkas Vilmos, A type-token viszony vizsgálatának változatairól /О вариантах изучения отношения "входений-типов"/. - MNy. 59: 424-30 /1963/
83. Fodor István, A matematikai módszer nyelvtudományi alkalmazásának határai: a nyelvi változások lefolyásának tempója /О границах применения математических методов в языковедении: темпы языковых изменений/. - NyK. 65: 297-339 /1963/
84. Papp László, Nyelvjárástörténet és nyelvi statisztika /Историческая диалектология и лингвостатистика/. Budapest 1963
85. Gulya János, A fonémarendszerek változásának mennyiségi összetevői /Количественные составляющие изменения фонологических систем/. - NyK. 66: 303 - 14 /1964/

86. Fodor István, Az afrikai nyelvek osztályozásának problémái /Проблемы классификации африканских языков/. - NYK. 67: 41-111 /1965/
87. Дже Д., Статистическое исследование словообразования в языковых памятниках. - ALH. 16: 43-62 /1966/
88. Csúcs Sándor, A magyar szókészlet finnugor elemei a statisztika tükrében /Угро-финские элементы венгерского словарного состава в зеркале статистики/. - NYK. 70: 392-6 /1968/
89. Papp Ferenc, Adatok tőszókincsünk etimológiai rétegeinek végződés szerinti eloszlásához /Относительно распределения этимологических пластов венгерских корневых слов по окончаниям в словарной форме/. - NYK. 70: 200-9 /1968/
90. Vértés Edit, Rekonstrukció és valószínűségszámítás /Реконструкция и теория вероятности/. - NYK. 72: 297-322 /1970/
91. Papp Ferenc, Tőszókincsünk etimológiai rétegenkénti hangstatisztikája /Звуковая статистика разных этимологических пластов корневых слов венгерского языка/. - NYK. 75: 3-40 /1973/
92. Janurik Tamás, Le système de correspondance phonétique dans le dictionnaire étymologique "Les éléments finno-ougriens dans le vocabulaire du hongrois". - NYK. 77: 137-46 /1975/
93. Papp Ferenc, Adalékok a magyar nyelvtörténet két szinkron metszetéhez /К двум синхронным срезам в истории венгерского языка/. - NYK. 70: 404 - 8 /1976/
94. Papp Ferenc, Etimológia és statisztika /Этимология и статистика/. - NyTÉrt. 89: 234-9 /1976/
95. Szabó József, A mellérendelt és alárendelt mondatok gyakoriságának vizsgálata Nagykönyvi nyelvjárásában /К изучению частот сочинительных и подчинительных придаточных предложений в диалекте с. Надькони/. - NyT. 100: 201-33 /1976/
- См. еще в следующих работах, помещенных в других разделах:  
/ 29 /, / 31 /, / 39 /, / 53 /, / 57 /, / 59 /, / 68 /.

С о к р а щ е н и я , п р и н я т ы е  
в б и б л и о г р а ф и и

- ALH. - Acta Linguistica Academiae Scientiarum Hungaricae, Budapest
- ÁNyT. - Általános nyelvészeti tanulmányok. Budapest, I - 1963 (непериодическая серия)
- MNyJ. - Magyar nyelvjárások. Debrecen, I - 1951 (ежегодник)
- NyelvfelDok. - Nyelvfeldolgozás és dokumentáció. Szerk.: Szépe György (собрник)

- NyK. - Nyelvtudományi közlemények, Budapest  
Nyr. - Magyar Nyelvőr, Budapest  
NyÉrt. - Nyelvtudományi értekezések. Budapest (неперическая серия)  
Slavica - Annales..., Debrecen, I - 1961 (ежегодник)  
I. OK. - A Magyar tudományos akadémia nyelv- és irodalomtudományi osztályának közleményei, Budapest  
НТИ - Научно-техническая информация. Серия 2. Информационные процессы и системы. Москва.

## LINGUISTIC STATISTICS AND THE HUNGARIAN LANGUAGE

Ferenc Papp

### S u m m a r y

In the first part of the paper, the major statistical indexes of the Hungarian language are given in terms of the classification presented in Fig. 1. 1.-3.: phoneme and letter statistics from a diachronic and a synchronic viewpoint /cf. Fig. 2.: the most frequent Hungarian phonemes/; 4.-6.: the percentage of running words consisting of *i* phonemes/; 7.: word statistics of language records; 8.-10.: word statistics, word frequency counts and dictionaries /cf. Figs. 4. and 5.: the distribution of running words resp. lexemes occurring with an *i* frequency/; 11.-13.: morphological statistics, 14.: syntactic statistics. The second part contains a bibliography of statistical studies relating to the Hungarian language /without stylistic statistics/.

## СТАТИСТИЧЕСКОЕ ИССЛЕДОВАНИЕ МОРФОЛОГИЧЕСКОЙ СТРУКТУРЫ ПОЭЗИИ

---

---

Тибор Жилка

I. Характерным для начала статистического исследования в лингвистике являлось нагромождение чрезмерного количества квантитативных данных. В интересах строгой объективности математические результаты приобретались на основе большой совокупности элементов. Конечная цель исследования заключалась в подготовке частотного словаря и в выяснении частотности фонем, морфем, слов и т.д. Исходным положением исследований были принципы теории вероятности и статистики, обеспечивавшие более объективные и достоверные результаты, так как они приобретались вычислением из как можно большего количества элементов. Как показывают более поздние исследования, эти результаты можно достигнуть постепенно, шаг за шагом, т.е. детальным исследованием меньших целых, или же исследованием цельных и однотипных структур (Rákos, 1966; Tuldava, 1970; Tuldava, 1974). Таким образом можно легче проверить приобретенные математические данные и результаты, лучше "подхватить" на фоне интуиции и эмпирии некоторые отклонения, исследовать их смысл, воздействие, функцию. Особенно большое значение для стилистики имеет исследование меньших целых. Доказательством этого являются и результаты, достигнутые в этом направлении в последние годы.

Если статистическая ориентация стилистики у нас (в ЧССР) обыкновенное явление, то в Венгрии математические методы в этой области применяются реже (Dömölki-Fónagy-Szende, 1954; Fónagy, 1959; Szende, 1968; Tarnóczy, 1961; Török, 1968). Именно поэтому мы попытались провести статистический анализ морфологической структуры избранных стихотворений четырех венгерских поэтов XIX века.

В сущности исследуются стихотворения, а не цикл или целый сборник. При их выборе мы стремились к тому, чтобы тексты были разнообразными и типичными (характерными) для данного автора. Мы взяли по четыре стихотворения АТ, АИ, ЛК и СВ.\* Каждый из этих поэтов представляет особое литературное

---

\* АТ - Árpád Tóth  
АИ - Attila József

ЛК - Lajos Kassák  
СВ - Sándor Weöres

направление, или же другой период в венгерской литературе. АТ является типичным представителем венгерского импрессионизма, в творчестве АЙ переплетаются элементы сюрреализма и экспрессионизма, ЛК является представителем особого венгерского литературного направления, активизма, и о СВ не случайно говорят как об авторе, поэтическое произведение которого содержит знаки постимпрессионизма. Шестнадцать стихотворений упомянутых поэтов служат основным материалом нашего исследования. В статье будем пользоваться инициалами авторов и индексами I - 4.

<u>Название стихотворения</u>	<u>Обозначение стихотворения</u>
Новый Бог /Az új isten/	АТ /1/
Вечерний венок лучей /Esti sugárkoszorú/	АТ /2/
Маскируя /Álarcosan/	АТ /3/
В конце мая старое вино /Május végén a régi bor/	АТ /4/
Около Дуная /A Dunánál/	АЙ /1/
Ода /Óda/	АЙ /2/
Моя Родина /Hazám/	АЙ /3/
Ночь на окраине города /Külvárosi éj/	АЙ /4/
Мастера /Mesteremberek/	ЛК /1/
Марты /Márciusok/	ЛК /2/
Глаза /Szemek/	ЛК /3/
Одноглазый Брюггелъ /Brügel, a félszemű/	ЛК /4/
Valse triste	СВ /1/
Анналы природы /A természet évkönyvei/	СВ /2/
Поэзия будущего /A jövődo költészete/	СВ /3/
Влюбленные в большом городе /Szeretok a nagyvárosban/	СВ /4/

Исходным пунктом для анализа результатов статистического исследования будет расположение частей речи во всех стихотворениях, т.е. во всем исследуемом материале. Эта основа послужит для нас ориентиром для идентификации функциональной сущности "непредполагаемой" структуризации словных единиц у отдельных поэтов. И, наконец, попытаемся объяснить функциональную каузальность специальной дистрибуции семантически и стилистически наиболее релевантных частей речи в стихотворениях. Нужно, конечно, заметить, что результаты являются пока не достаточным показателем структурно-морфологического строя

венгерской поэзии XIX века. Но с другой стороны, некоторые результаты указывают и на типологические особенности венгерского языка в сравнении с другими языками. Таким образом, мы попытаемся начертить определенную проблематику, решение которой потребует дальнейших статистических исследований стиля или языка в целом.

2. Результаты, вычисленные на основе всех цитруемых стихотворений, мы представили в таблице (табл. I). Она является в какой-то мере показателем расположения частей речи в венгерской поэзии XIX века.

Таблица I

	N	p(%)	$(p-\bar{p})^2$	$s^2$	s	V	$V_r$
Существительные	1047	31,89	33,4677	8,367	2,89	0,0900	7,12
Формальные слова	709	21,60	6,2907	1,573	1,25	0,0593	2,92
Глаголы	548	16,69	29,9643	7,491	2,74	0,1787	6,89
Прилагательные	366	11,15	29,2273	7,307	2,70	0,2212	7,37
Наречия	218	6,64	4,8255	1,206	1,10	0,1537	2,90
Местоимения	206	6,27	7,3797	1,845	1,36	0,2438	3,20
Неопределенные глагольные формы	137	4,17	4,9399	1,235	1,11	0,2562	2,74
Числительные	38	1,16	0,7505	0,188	0,43	0,4759	1,12
Междометия	14	0,43	0,2756	0,069	0,26	0,5044	0,66
Всего	3283	100,00	-	-	-	-	-

Как показывает таблица, 16 стихотворений заключают в себе всего 3283 словоупотребления. Из общей частотности отдельных частей речи (N) мы вычислили процентную нагрузку частей речи (p), вторую степень отклонения от среднего  $(p - \bar{p})^2$ , дисперсию ( $s^2$ ), стандартное отклонение (s), вариационный коэффициент (V) и вариационный диапазон ( $V_r$ ). При вычислениях мы пользовались следующими формулами:

$$s^2 = \frac{\sum /p-\bar{p}/^2}{k}; \quad s = \sqrt{\frac{\sum /p-\bar{p}/^2}{k}};$$

$$V = \frac{g}{p};$$

$$V_r = P_{\max} - P_{\min}.$$

Приобретенные статистические данные можно интерпретировать, с одной стороны, на основе уже достигнутых методологических познаний, с другой стороны, на фоне предыдущих результатов, достигнутых в этой области (сравни: Ewgen, 1967; Honecký, 1969; Mistrík, 1969).

Давая общую оценку морфологических свойств исследуемых стихотворений четырех авторов, уже с первого взгляда заметен большой разброс частей речи в процентном отношении. Существительные составляют почти одну треть из общего объема текста. Прилагательные представляют приблизительно одну десятую текста. Процент числительных очень низкий. Если повысятся их процент, "отклонение" от нормы будет признаком определенной направленности текста. Глаголы после существительных и формальных слов (артикли, союзы и постпозиции) несут наименьшую процентную нагрузку, но все-таки их частотность достигает лишь примерно половины частотности существительных. Вероятно, это связано с номинальностью, большей частотностью именных сказуемых в венгерской поэзии, одновременно и в венгерском языке.

При определении части речи вместо венгерского названия *igenevek* мы употребили термин неопределенные глагольные формы. Это название для причастий, деепричастий и инфинитива составляет в финно-угорских языках общую группу. Их частотность небольшая. Встречаемость местоимений и наречий приблизительно одинакова. Между этими частями речи в венгерском языке все же существует много точек соприкосновения, вследствие чего иногда стираются границы между ними. Отдельно нужно отметить формальные слова, имеющие высокую частотность, хотя они в качестве лексических единиц составляют малый процент текста. Это значит, что их повторяемость большая. Междометия в тексте встречаются редко, это подтверждают и наши вычисления, причем их встречаемость распространяется на поэзию и разговорную речь; но их процент в поэзии все же небольшой.

С помощью вычисления дисперсии, стандартного отклонения, вариационного коэффициента и вариационного диапазона можно точно верифицировать определенные гипотезы о языке поэзии, или же о языке в целом. Числа абсолютных величин разброса

(рассеивания), т.е. величины дисперсии и стандартного отклонения значительны. Уже с первого взгляда заметно, что наивысшего (абсолютного) разброса достигают семантически и стилистически самые четкие части речи (существительные, прилагательные и глаголы). Значительно низкий разброс имеют числительные и особенно междометия. Важным показателем является вариационный коэффициент, который указывает на относительную величину разброса; интересным результатом являются, например, относительная стабильность существительных и сравнительно большая вариация процента прилагательных в тексте. Наивысшую степень относительной вариации имеют части речи с самой низкой частотностью, т.е. междометия и числительные. И, наоборот, самую низкую степень относительной вариации — части речи с наибольшей частотностью, т.е. формальные слова и существительные. Самый высокий вариационный диапазон имеют прилагательные, существительные и глаголы. Эти части речи можно на основе результатов о вариационном диапазоне противопоставить остальным частям речи. Больше всего к ним приближается результат о вариационном диапазоне местоимений.

3. От общих результатов распределения частей речи перейдем к общим результатам о поэтах (табл. 2). Здесь также приводим не отдельные результаты из каждого стихотворения, а вместе из четырех стихотворений каждого поэта.

Таблица 2

	АТ	АЙ	ЛК	СВ
Существительные	29,02	31,17	36,14	35,12
Формальные слова	19,82	22,74	21,97	20,00
Прилагательные	15,67	8,30	11,50	13,42
Глаголы	13,34	19,83	12,94	15,12
Наречия	8,29	5,39	7,80	7,07
Местоимения	6,48	7,31	4,11	4,39
Неопределенные глагольные формы	6,09	3,35	4,52	3,41
Числительные	0,91	1,61	0,62	0,49
Междометия	0,39	0,31	0,41	0,97

3.1 В таблице можно заметить, что в творчестве АТ имеют высокую частотность те части речи, которые в венгерских грамматиках называют именными формами; это касается существ-

вительных, прилагательных, местоимений и числительных, но в сущности также и глагольных форм. Заметна прежде всего высокая встречаемость прилагательных, что полностью соответствует "характеру" импрессионистской поэзии. Этим, наконец, повышается стилистическая нагрузка определений; число определений еще дополняется и большей встречаемостью причастий. С другой стороны, частотность глаголов, или же глагольных сказуемых значительно снижается, даже настолько, что их процент встречаемости не достигает такой "степени", как у прилагательных. Такое отклонение от "нормального" распределения частей речи нужно считать характерным признаком творчества АТ и в конечном счете и импрессионизма как литературного направления.

3.2 У АИ можно обнаружить значительное понижение частотности прилагательных по сравнению с АТ, и одновременное повышение процента глаголов. Это значит, что эволюция поэзии идет в направлении от "номинальности" к большей "вербальности". При этом у названного поэта повышается и частотность существительных. Вероятно, это связано и с важностью и социальной ангажированностью поэзии АИ. Дело в том, что такая тематика требует больше слов, изображающих реальные предметы и явления. Думается, что это одна из характерных черт экспрессионистически настроенной поэзии. (Наше мнение подтверждают и результаты, полученные статистическим анализом стихотворений Кассака).

3.3 ЛК собственно всем своим творчеством представлял крайнее крыло модернизма в венгерской литературе. Его поэзия проникнута обновляющими элементами, которые проявляются на уровне формы и темы. В сущности он никогда не избавился от влияния экспрессионизма и в значительной степени все свое творчество строил на традициях этого литературного направления. Экспрессионизм отличался как раз сильным процентом существительных в тексте, которые были признаком грубой, иногда и деформированной действительности. О номинальном характере творчества ЛК свидетельствует и то, что в тексте незначительно представлены глаголы; процент встречаемости прилагательных почти такой же. Сниженную частотность глаголов можно объяснить большой встречаемостью эллиптических предложений, в которых обычно опускаются именно глагольные формы. Эти эллиптические предложения характерны для венгерского экспрессионизма; это, между прочим, иногда напоминает публи-

цистический стиль.

3.4 Стихотворения СВ, как мы уже указывали, обладают постимпрессионистскими свойствами. Это отражается в проценте отдельных частей речи, особенно тех, которые характерны для данного литературного направления. Результаты вычислений действительно подтверждают, что стихотворения СВ имеют постимпрессионистский характер. Процент прилагательных настолько высок, что опережает его лишь АТ. При этом встречаемость существительных тоже очень высокая. В отношении частотности глаголов СВ следует сразу после АЙ.

4. В процессе статистических исследований словарного запаса, или же частей речи, можно вычислить и некоторые стилистически важные коэффициенты (табл. 3). Из этих коэффициентов в научной литературе самым известным является коэффициент Буземана. Его можно вычислить из пропорции прилагательных и глаголов (Antosch, 1969; Fischer, 1969; Tešitelová, 1968, Muller, 1972). Кроме коэффициента Буземана, нам важно выяснить и пропорции прилагательных и существительных, а также соотношение существительных и глаголов. При этом пользуемся терминами, которые ввела М. Тешителова (Tešitelová, 1968, 98-99). В первом случае мы будем говорить о коэффициенте развертывания, в другом - о коэффициенте номинальности. Коэффициенты развертывания  $K_r$  и номинальности  $K_n$ , а также коэффициент Буземана  $K_b$ , мы вычислили во всех стихотворениях, так как они независимы от числа элементов, т.е. от объема статистического комплекса (выборки). Мы считаем, что у меньшего числа элементов их величина скорее может быть очевидным доказательством объяснения стилистических особенностей стихотворного текста, чем у большего комплекса, или текста.

4.1 Вычисленные коэффициенты помогают открыть стилистические особенности поэзии АТ. Коэффициент развертывания свидетельствует о сравнительно высокой степени "распространения" поэтического языка при помощи прилагательных. "Эксплицитность", т.е. чрезмерной распространенностью выделяются стихотворения АТ /1/ и АТ /3/; меньше всего "распространено" стихотворение АТ /2/.

Коэффициент номинальности имеет значение прежде всего при идентификации соотношения номинальных и вербальных форм. Как видно по результатам, у АТ преобладает номинальность;

больше всего это видно в стихотворениях АТ /1/ и /4/. Кроме того, у АТ высоки и значения коэффициента Буземана, что связано со статичностью текста. В двух случаях (АТ /1/ и АТ /3/) прилагательные преобладают над глаголами, что является действительно редким явлением.

4.2 В стихотворениях АИ заметна редкая уравновешенность результатов развертывания. Для развертывания именных форм, прежде всего существительных, поэт использует прилагательные сравнительно редко. Конечно, существительные преобладают над глаголами, но уже не в такой мере, как у АТ. Чрезмерная номинальность в стихотворении АИ /4/ вытекает из того, что произведение содержит больше элементов экспрессионизма, чем остальные стихотворения. Результаты вычисления соотношения прилагательных и вербальных форм доказывают малую частотность прилагательных по сравнению с глаголами. Исключение составляет стихотворение АИ /4/, в котором часто встречаются под влиянием экспрессионизма эллиптические предложения с элиминлируемыми глаголами.

4.3. В стихотворениях ЛК коэффициент развертывания очень меняется. Самое большое отклонение мы можем наблюдать в стихотворении ЛК /2/, что вызвано тем, что оно является типичным произведением экспрессионистской поэзии. Дело в том, что экспрессионизм заявлял и в своих манифестах о намерении не пользоваться прилагательными. Эти причины вызваны, таким образом, чрезмерным отклонением от "нормы".

Таблица 3

	$K_r$	$K_n$	$K_b$
АТ /1/	0,602	2,572	1,547
АТ /2/	0,390	1,952	0,762
АТ /3/	0,690	1,526	1,053
АТ /4/	0,435	2,191	0,952
АИ /1/	0,245	1,056	0,258
АИ /2/	0,297	1,609	0,478
АИ /3/	0,262	1,494	0,391
АИ /4/	0,252	2,519	0,635
ЛК /1/	0,474	2,591	1,227
ЛК /2/	0,095	3,818	0,364
ЛК /3/	0,345	3,625	1,250
ЛК /4/	0,313	2,182	0,682

Продолжение таблицы 3

	$K_r$	$K_n$	$K_b$
СВ /1/	0,333	2,333	0,778
СВ /2/	0,310	3,000	0,929
СВ /3/	0,414	4,143	1,711
СВ /4/	0,516	1,348	0,696

С этим свойством тесно связан и коэффициент номинальности или же номинальное свойство экспрессионистской поэзии. Кроме приведенного произведения ЛК /2/, это ярче всего проявляется в стихотворении ЛК /3/, номинальность которого заложена в общем построении текста. Коэффициент Буземана имеет у ЛК также большие отклонения. Самой низкой величины он достигает в стихотворении ЛК /2/, самой высокой – в стихотворениях ЛК /1/ и ЛК /3/. Следует отметить, что в этих двух стихотворениях нет даже такой большой частотности глаголов; сказанное относится и к стихотворению ЛК /2/, в котором одновременно наблюдается очень низкая частотность прилагательных.

4.4 Результаты, относящиеся к стихотворениям СВ, в большинстве случаев нерегулярны, между ними существуют большие различия. Частичным исключением можно считать коэффициент развертывания, так как у него достаточно заметна регулярность. Также можно отметить, что этот коэффициент у СВ достигает всегда высокой степени (доказательство постимпрессионистского характера поэтического творчества). Коэффициент номинальности значительно "расшатанный", неслаженный, что также имеет свои причины. Например, в стихотворении СВ /4/ соотношение между существительными и глаголами 1:1,35, но в стихотворении СВ /3/ это соотношение намного выше – 1:4,14. В стихотворении СВ /3/ прилагательные широко представлены за счет глаголов (идентификация на основе коэффициента Буземана). В остальных стихотворениях хотя и высокая частотность прилагательных, но всегда ниже, чем у глаголов.

4.5 О стиле отдельных поэтов можно сделать некоторые выводы на основе общих результатов (табл. 4).

4.5.1. Соотношение прилагательных и существительных самое высокое у АТ, далее у СВ, затем у ЛК и, наконец, у АЙ. Результаты доказывают, что импрессионистски настроенная поэ-

Таблица 4

	$K_r$	$K_n$	$K_d$
АТ	0,540	2,175	1,175
АИ	0,267	1,572	0,419
ЛК	0,318	2,794	0,889
СВ	0,382	2,323	0,887

зия требует форм прилагательных, распространявших в предложении существительные в функции подлежащего, дополнения или же обстоятельства. Экспрессионизм, наоборот, ограничивает употребление прилагательных (сравним подсчеты в стихотворениях АИ и ЛК).

4.5.2 Коэффициент номинальности имеет большой разброс в творчестве отдельных авторов. Самой низкой степени номинальности достигает творчество АИ (большое количество глаголов). В какой-то мере удивляет, что после него следует АТ. Причиной этого является факт, что у АТ так называемый "номинальный стиль" строится на большом количестве прилагательных. Высокий коэффициент номинальности у следующих двух авторов вызван, с одной стороны, высокой частотностью эллиптических предложений (ЛК), с другой стороны, частой встречаемостью именных сказуемых (СВ).

4.5.3 Коэффициент Буземана, вычисляемый из соотношения прилагательных и глаголов, в какой-то мере является математическим измерением сюжетности или же субъективности текста. Если преобладают прилагательные, текст скорее субъективен, даже эмоционален; если преобладают глаголы, текст выделяется сюжетностью. Эмпирически бесспорно, что стихотворения АТ эмоциональны. То же самое показывают и математические вычисления, так как коэффициент Буземана в его стихотворениях самый высокий (1,175). Противоположностью является АИ: его стихотворения имеют высокую сюжетность, поэтому у него коэффициент самый низкий. В стихотворениях ЛК и СВ нет более заметных различий в коэффициенте Буземана; по результатам к каждому прилагательному относится в среднем один глагол. В данном случае это показывает, что коэффициент Буземана сам по себе недостаточен для дифференциации стилей.

5. Данные, полученные нашим статистическим исследованием морфологической структуры поэзии, можно подытожить следующим

образом.

Установлением частотности частей речи и их распределения получены результаты, которые можно считать определенным, пока лишь гипотетическим видом математической модели. Хотя ее нельзя считать абсолютно действующей, все-таки можно на нее опираться, если хотим определить стилистические различия в конкретных текстах стихотворений. Именно тогда, когда стиль определяем как отклонение от предполагаемого, "ожидаемого" распределения элементов, т.е. как перегруппировку дистрибуции конституэнтов текста. Одновременно это значит, что значительные отклонения от модели закономерны должны сигнализировать о функциональном характере стиля автора или же стихотворения. Например, высокий процент прилагательных может быть доказательством того, что не только стихотворения А1 имеют импрессионистский характер, но и стихотворения СВ. И, наконец, отдельные коэффициенты подтверждают предыдущие статистические результаты, или же результаты, которые были получены исследованием другого уровня текста (максимального, синтагматического и т.д.). В связи с этим нужно еще сказать, что статистика не может входить в текст "вслепую" и ее нельзя считать самостоятельным методом (Miko, 1970; Ермоленко, 1977). Если упустить из виду качественный и функциональный "фон" необычной дистрибуции элементов в тексте, то и данные, полученные с помощью статистики, останутся совершенно неполными. Статистическим исследованием часто остается лишь верифицировать качественные результаты. Желательно, чтобы статистический анализ текста был эвристическим и доказывающим инструментом функционального анализа текста. Но исходным пунктом должна стать математическая модель литературного направления, стиля автора и т.п., с которой можно сравнивать результаты, вычисленные из конкретного текста. Считаем, что именно в этом заключается значение и практический смысл статистики для стилистики.

#### Л и т е р а т у р а

- Antosch, F. The Diagnosis of Literary with the Verb-Adjectiv Ratio. In: Statistics and Style. - New York: American Elsevier Publishing Company, 1969, pp. 57-67.
- Dömölki, B., Fónagy, I., Szende, T. Könyvevi hangstatisztikai vizsgálatok. - Általános nyelvészeti tanulmányok. 2 Szerk. J. Kalmár - Zs. Telegdi. - Budapest: Akadémiai Kiadó, 1964, 1. 117-132.

- Essen, O. von. Statistische Rechnungen in der Phonetik. - Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung, 20, 1967, S. 3-59.
- Fischer, H. Entwicklung und Beurteilung des Stils. - In: Mathematik und Dichtung, 3. Aufl. Red. H. Kreuzer. - München: Nymphenburger Verlagshandlung, 1969, S. 171-183.
- Fónagy, I. A költői nyelv hangtanából. Irodalomtörténeti Füzetek 23. 1. kiad. - Budapest: Akademia Kiadó, 1959.
- Horecký, J. Úvod do matematické jazykovedy. (Skriptum) 1. vyd. - Bratislava: Rektorát Univerzity Komenského, 1969.
- Miko, F. Text a štýl. 1. vyd. - Bratislava: Vydavateľstvo Smena, 1970.
- Mistrík, J. Frekvencia slov v slovenčine. 1. vyd. - Bratislava: Vydavateľstvo SAV, 1969, l. 13-128.
- Muller, Ch. Einführung in die Sprachstatistik. - Berlin: Akademie-Verlag, 1972.
- Rákos, P. Rhythm and metre in Hungarian verse. - Acta Universitatis Carolinae. Philologica Monographia 11. 1. ed. - Praha: Universita Karlova, 1966.
- Szende, T. Szóstatisztikai vizsgálatok Tóth Árpád, Juhász Gyula és Szabó Lőrinc költeményei alapján. - Magyar Nyelvőr, 92, 1968, l. 287-300.
- Tarnóczy, T. A jelözlés és a hírtartalom nyelveket meghatározó tulajdonságairól. Nyelveket meghatározó tulajdonságairól. Nyelvtudományi Közlemények, 63, 1961, l. 161-178.
- Těšitelová, M. K typologii slovanského slovníku z hlediska kvantitativního (na českém materiálu). - Československé přednášky pro VI. mezinárodní sjezd v Praze. Red. M. Brozda - E. Pauliny. - Praha: Nakladatelství SAV, 1968, l. 95-99.
- Török, G. A Líra logika. I. kiad. - Budapest: Magvető-Tiszatáj, 1968, l. 33-51.
- Tuldava, J. Információnteória ja keeleteadus. - Keel ja Kirjandus, 13, 1970, nr. 6, lk. 329-339.
- Ермоленко Г.В. Роль речевых параллелей при атрибуции художественных текстов. - В кн.: Языковая норма и статистика. - Москва: Наука, 1977.
- Tuldava Ю.А. Об измерении лексической связи на уровне словаря. - В кн.: Вопросы статистической стилистики. - Киев: Наукова думка, 1974.

STATISTICAL INVESTIGATION OF THE MORPHOLOGICAL  
STRUCTURE OF POETRY

Tibor Zsilka

S u m m a r y

The article is devoted to the statistical investigation of style on the basis of contemporary Hungarian poetry. The distribution of parts of speech and various indexes showing the mutual proportions of parts of speech have been used for the construction of a mathematical model and for the comparison of the individual style of various poets. The results of statistical investigation have undergone qualitative-stylistic interpretation.

Juhan Tuldava

Sõnavara statistilise vaatluse alla kuulub traditsiooniliselt ka sõnade välise kuju (koostis, pikkus) analüüs. Sõnu vaadeldakse sel juhul kui tähtede (grafeemide), häälikute või foneemide jadasid ja kombinatsioone, mille kvantitatiivne uurimine võimaldab kindlaks teha üksikute elementide produktiivsust ja funktsionaalset koormust ning kokkuvõttes luua sõnastiku ja teksti kvantitatiiv-struktuurne tüpoloogiline foneetilis-grafeemilisel või fonoloogilisel tasandil. Jätakuks varem avaldatud töödele eesti keele sõnavara statistiliste omaduste kohta (vt. näit. Kaasik jt., 1976 ja 1977; Tuldava, 1977 ja 1978) vaatleme käesolevas artiklis eesti keele sõnavara foneetilis-grafeemilisi (häälikulisi ja tähelisi) mõõteid nii tekstis kui sõnastikus, kusjuures vaatlusmaterjal on võetud viiest eesti keele allkeelest (ilukirjandusproosa autorikõne ja tegelaskõne, ajaleht, teaduskirjandus, luule). Võrdluseks tuuakse näiteid teistest keeltest, esmajoones sugulas- ja naaberkeeltest.

Sissejuhatavad märkused. Foneetilis-grafeemiliste mõõdetega all mõtleme laiemas mõttes nii häälikute või tähtede kui ka hääliku- või tähekombinatsioonide sagedusi teksti ja sõnastiku tasandil. Häälikute või foneemide kombinatsioonid ja nende põhjal saadud struktuurid kuuluvad kitsamas mõttes fonotaktika (tähtede puhul vastavalt grafemotaktika) valdkonda, mida antud juhul ei vaadelda (eesti keele kohta vt. Kaasik jt., 1975; Tuldava, 1978). Käesolevas töös käsitletakse vaid eesti keele tähtede ja häälikute, sealhulgas algus- ja lõpptähtede (-häälikute) sagedusi tekstis ja sõnastikus.

Nimetatud kvantitatiivsed vaatlused on vajalikud mitmesuguste rakenduslike ülesannete lahendamisel ja seepärast on vastavaid uurimusi tehtud väga paljude keelte kohta. Tähtede sagedusi on uuritud stenograafiliste süsteemide optimeerimiseks, kirjutusmasinate klaviatuuri koostamiseks, krüptograafias, logopeedias, tootenimede konstrueerimisel

jne. Olulist osa etendavad tähelised mõõted (sealhulgas tähe-kombinatsioonide sagedused) informatsiooni edastamisel sidakanalite kaudu (kusjuures tähtede ja nende kombinatsioonide sagedusi vaadeldakse informatsiooniteooria valguses), informatsiooni töötlemisel elektronarvuti abil jms. Täheliste mõõdete vajalikkust näitab ka see, et nende abil on suudetud lahendada mõningad huvitavad lingvistilised probleemid. Tuntud on nõukogude teadlase B. Suhhotini katse klassifitseerida automaatselt tähti nende distributsiooni alusel (Сухотин, 1962); selle katse edasiarendamine võib viia ka foneemide automaatselt identifitseerimisele (Karlgren, 1968, 135). Võib nimetada ka soomlase S. Mustoneni (1965) katset määrata sõnade kuuluvust ühesse või teise keelde tähtede esinemissageduse alusel.

Tähelisi mõõteid on kasutatud ka keelte tüpoloogilisel võrdlemisel, kuigi sel juhul on õigem lähtuda häälikutest või foneemidest. Tuleb aga arvestada suuri raskusi keelte fonoloogilise-tüpoloogilisel uurimisel, mis on tingitud sellest, et sageli puudub ühtne teoreetiline alus eri keelte fonoloogiliseks kirjeldamiseks. Teatavasti puudub üksus isegi ühe keele fonoloogilise süsteemi (foneemivaru) esitamisel. Kujuka näite toob A. Isengeldina (Исеньгельдина, 1972), kes on kindlaks teinud, et erinevates uurimustes kõigub inglise keele vokaalfoneemide arv 8 - 22 vahel. Ka eesti keele fonoloogilise süsteemi suhtes leidub vastandlikke vaateid (vt. Вийтсо, 1979), mistõttu käesolevas töös on loobutud eesti keele sõnade kvantitatiivsest uurimisest foneemide tasandil. Omaette alapestükis vaadeldakse aga eesti keele häälikute ja nende rühmade sagedusi, lähtudes eesti keele traditsioonilisest häälikusüsteemist. See võimaldab teha tüpoloogilisi võrdlusi mõningate teiste keeltega. Kõrvutatavad tüpoloogilised uurimused nii foneetilise-grafeemiliste kui ka fonoloogiliste mõõdete alusel etendavad tähtsat osa Leningradi teadlase N. Andrejevi keelestatistika koolkonna töödes (Андреев, 1967), kusjuures tähtede (ja foneemide) sagedusi sõnade eri positsioonides kasutatakse sõnade automaatsel morfoloogilisel segmenteerimisel. Huvitavate tulemusteni keelte kõrvutataval uurimisel foneetilise-grafeemiliste või fonoloogiliste mõõdete alusel on jõutud ka paljudes teistes töödes (näit. Krámský, 1959; Kučera, Francis, 1968).

Милевский, 1963; Перебийнос, 1967; Вецовол, 1967).

Mitmed uurimused on käsitletud tähe- ja häälikusageduste muutumist keelte ajaloolises arengus (Бекраен, Лобин, 1969; Зиндер, Строева, 1972). Eesti keele suhtes on seni kindlaks tehtud, et tähtede sagedused pole viimase sajandi jooksul oluliselt muutunud (Kaasik, Laugaste, 1969). Eesti sõnade kõige varasemad kirjapanekud lubavad aga järeldada, et vanema eesti keele häälikute süsteem oli küllaltki erinev tänapäevasest (Kask, 1972, 15). Ka vene keele suhtes on täheldatud olulisi muutusi eri ajajärgude häälikute ja foneemide süsteemides ning vastavates kvantitatiivsetes näitajates (Зуравлев, 1974, 87). Huvi pakub rootsi teadlase B. Sigurdi (1963) uurimistulemus, mille kohaselt võib näetida foneemide üldarvu tõusu paljudes euroopa keeltes (näit. rootsi keeles on alates 13. sajandist foneemide arv tõusunud 25-lt 40-ni): B. Sigurd seletab seda kui "koodi nihet", nimelt liikumist sellise koodi poole, mis võimaldab edastada rohkem informatsiooni teatava ajaühiku vältel eeldusel, et foneemi keskmine vältus jääb samaks.

Paljusid uurijaid on huvitanud kardinaalne küsimus, miks üldse esinevad tähed (häälikud, foneemid) eri sagedusega. Informatsiooniteoreetilisest seisukohast võib seda seletada keele kui "koodi" optimaalse organisatsiooni tendentsiga, mis on täheldatav ka sõnasageduste erinevuse puhul (näit. Zipfi seadus). Miks aga esinevad just teatavad häälikud sagedamini kui teised, sellele pole suudetud anda lõplikku seletust. G. Zipf (1935) arvas, et sageduste erinevusi põhjustavad artikulatsiooni tingimused, nimelt olevat artikulatsiooni keerukus pöördvõrdeline hääliku sagedusega. Sellele on vastu vaieldud ja väidetud, et häälduse "keerukust" või "kergust" on võimatu mõõta (Trubetzkoy, 1939; Martinet, 1955). V. Nikonov arwab, et hääliku või foneemi sageduse "võti" ei peitu üldse foneetikas või fonoloogias, vaid sõnavara ja sõnatuletuse ajaloolises arengus (Никонов, 1963). A. Isengeldina väidab siiski, toetudes konkreetsetele uurimustele, et peab olema mingi seos kõnefüsioloogia ja häälikusageduse vahel: nii näiteks võib täheldada peaaegu kõigis keeltes apikaalide suurt sagedust, mis seletub keeletipu liikuvuse ja aktiivsusega (Исеньельдина, 1972).

Kui toetuda analoogiale sõnade ja grammatiliste nähtuste sageduse ja "raskuse" seose kohta, siis peaks väitma, et antud keele kandjale on sagedamini esinev häälik "lihtsam" kui harva esinev häälik (nii käsitleb küsimust ka P. Guiraud, 1954, 97). Igal juhul märgib hääliku suur esinemissagedus ühtlasi suurt funktsionaalset koormust ja suuri kombinatoorseid võimalusi, mis kahtlemata on teatud määral seotud häälduse lihtsusega (ökonoomiaprintsiip). Teisest küljest tuleb aga arvesse võtta häälikute tajumist ja eristamist kuulaja poolt, seda on viimasel ajal hakatud uurima ka psühholingvistiliste katsetega (näit. Меликишвили, 1970). Küsimuse lahendamisel võivad kasulikuks osutada katsed subjektiivsete sageduste määramisel (Фрумкина и др., 1971). Nn. "foneetilise sümbolismi" pooldajad seostavad mõningatel juhtudel häälikute sagedusi antud keele kandja hinnanguga ("hea", "halb", "ilus" jne.) häälikule või häälikut sisaldavale sõnale (vt. Журавлев, 1974, 86 jj.). Peab nimetama, et foneetilise sümbolismi uurimine on viimastel aastatel saanud uut hoogu nii Nõukogude Liidus kui välismaal, kusjuures on kasutusele võetud uusi täiustatud (sealhulgas statistilisi ja psühholingvistilisi) meetodeid (Левяцкий, 1969; Гурджиева, 1973; Журавлев, 1974; Marchand, 1959; Arndorfer, 1967; Ertel, 1969 jt.). Teatavasti uuris eesti keele häälikute omapära ja "tähendust", lähtudes vanakreeka filosoofide seisukohtadest, juba K.J. Peterson (vt. Ariste, 1968, 17 jj.). Uuemal ajal on tehtud katsed seostada luuleteksti foneetilist struktuuri semantikaga (Mäger, 1971).

Tähtede sagedused tekstis. Paljude keelte kohta tehtud uurimused on näidanud, et tähtede esinemissagedus on küllaltki stabiilne antud keele või allkeele tekstides. Kõnelemisel või kirjutamisel oleneb tähtede või häälikute valik nendest sõnadest, mida me kõnes kasutame. Keeles kehtivad aga kindlad seaduspärasused sõnade häälikulise struktuuri suhtes ja seepärast ei saa me reeglina vabalt valida häälikuid, välja arvatud kunstlikult esilekutsutud juhtudel (kurioosumina võib nimetada inglase E. Wright'i katset kirjutada terve romaan ilma ühegi e-täheta või 18. sajandi saksa luuletaja G. Burmanni 130 poemi ilma r-täheta). Tähtede sagedus on keelele omane nähtus, mis püsib konstantsena pi-

kema perioodi vältel, kuid täpsem analüüs näitab siiski mõningaid erinevusi allkeelte vahel. Sellest tingituna on uue-  
mates uurimustes püstitatud nõue, et tähtede (samuti hääli-  
kute, foneemide) sagedusi vaadeldaks allkeelte kaupa. Juhul,  
kui soovitakse esitada andmed kogu keele kohta, tuleb ära  
näidata allkeelte doseering üldises valimis. Seejuures peat-  
takse optimaalseks nelja-viie allkeele kaasatõmbamist "üld-  
keele" andmete kindlakstegemisel, näiteks järgmistes pro-  
portsioonides (Плотровский, 1968, 44): ilukirjandusproosa -  
30 %, kõnekeel (mida tavaliselt asendab tegelaskõne või  
draama) - 30 %, ajalehe- ja teaduskeel - 30 % ning luule-  
keel - 10 %. Umbes samasuguseid proportsioone kasutavad ka  
mõned välismaised autorid (näit. Kučera, Monroe, 1968: ilu-  
kirjanduskeel - 60 %, ajalehekeel - 20 %, teaduskeel - 10 %,  
luulekeel - 10 %).

Käesolev uurimus eesti keele tähtede sageduse kohta  
põhineb autori varasemal tööil (valimi maht N = 50.000 täht-  
te; vt. Tuldava, 1970a), millele lisati ligi 30.000 tähte  
erinevates allkeeltest.<sup>+</sup> Kokkuvõttes jaotuvad osavalimid  
järgmiselt:

ilukirjandusproosa	25420 tähte	(32 %)
kõnekeel (tegelaskõne)	20914 "	(27 %)
ajaleht	16208 "	(20 %)
teadusalane tekst	11358 "	(15 %)
luule	5000 "	(6 %)
	<hr/>	
	kokku 78900 "	(100 %)

Tekstid pärinevad tänapäeva eesti keelest (ilmumisaeg  
pärast 1960.a.). Ilukirjandusproosa ühendab antud juhul nii  
autorikõne kui ka tegelaskõne (vahekorras umbes 80:20), kus-  
juures osavalimid on võetud 30 eri autori teostest. On uu-  
ritud ka autorikõnet ja tegelaskõnet omaette (vt. allpool).  
Kõnekeelt esindab antud katses ilukirjandusteoste tegelas-  
kõne (peamiselt dialoog). Teadusalased tekstid on võetud  
enam-vähem võrdsetes proportsioonides keelteaduslikest,  
ajaloolistest, loodusteaduslikest, matemaatilistest ja teh-  
nika-alastest artiklitest (nii erialastest kui populaartea-

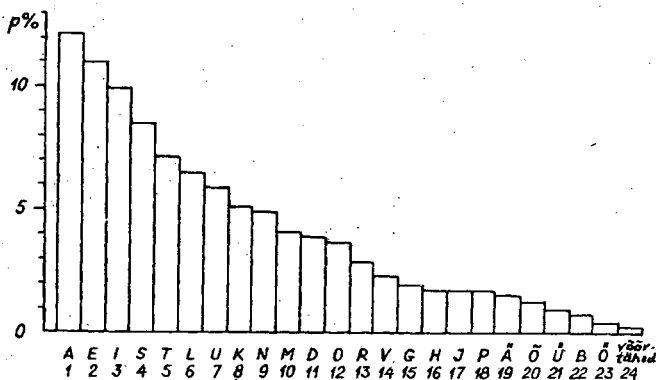
<sup>+</sup> Käesolevas töös avaldatavad tähtede sagedused saadi  
valdavalt osas tekstide automaattöötlisel TRÜ arvutuskes-  
kuses.

duslikest ajakirjadest). Luulekeele materjal (5·1000) pärineb erinevatest luulekogudest.

Koondvalimi proportsioonid on üldistavalt kokkuvõetuna järgmised:

ilukirjandus (proosa, tegelaskõne, luule) 65 %;  
 ajalehed ja teadusalsed tekstid 35 %.

Uurimise tulemused esitatakse tabelis 1. Tabelis on toodud 23 eesti tähe ja võõrtähtede rühma (koos) esinemisagedused viie allkeele lõikes ja koondtekstis. Tähed järjestati sageduste kahanevas reas (koondteksti alusel). Joonis 1 illustreerib tähesageduste kahanevat rida histogrammi näol.



Joqn. 1. Tähtede sagedused kahanevas reas (histogramm).

Koondteksti ("üldkeele") andmete järgi on kümme kõige sagedamat tähte ja vastavad suhtelised sagedused tekstis järgmised: a (12,2 %), e (10,9), i (9,8), s (8,6), t (7,1), l (6,5), u (5,8), k (5,1), n (4,9) ja m (4,1). Kõige harvemini esinevad tähed ü (0,9 %), b (0,8), š (0,3) ja võõrtähed (kokku ligikaudu 0,2 %). Vaadeldes võõrtähtede jaotumast eri allkeeltes (tabel 1)<sup>†</sup> näeme, et neid esineb kõige

<sup>†</sup> Hajuvushinnangud on antud tabelis 2 (arvutuste kohta vt. Tuldava, 1969).

rohkem ajalehekeeles (0,9%), kuna aga kõigis teistes allkeeltes on võõrtähtede protsent 0 - 0,1%. Huvi võib pakkuda võõrtähtede sageduste jaotumus meie koondvalimis (N = 78.900 tähte, sellest võõrtähti 203):

f	74	0,09 %	y	16	0,02 %
š	36	0,05	z	12	0,015
c	24	0,03	x	4	0,005
ž	18	0,02	q	2	0,0025
w	17	0,02			

Kokku 203 0,25 %

Üllatab, et väike hulk sageli esinevaid tähti katab suure osa tekstist. Nii näiteks moodustavad 5 sagedamat tähte (a, e, i, s, t) koondtekstist 48,6%, s.o. iga teine täht tekstis on üks nimetatud viiest. 10 sagedamat tähte katavad 75% tekstist. Samasugune olukord valitseb ka teistes keeltes. Soome keeles katavad 10 sagedamat tähte isegi 80% (Setälä, 1972, 11), hispaania keeles 76,2%, inglise keeles 75,2% (Pierce, 1962), saksa keeles 72,2% (Meier, 1964, 334) tekstist.

Meie katse andmetel on eestikeelses tekstis vokaaltähti keskmiselt 46,4% ja konsonanttähti 53,6%. Kui arvestada, et eesti tähestikus on 9 vokaalimärki ja 14 konsonandimärki, s.o. suhtes 39% : 61%, siis ilmneb, et vokaaltähtede "tekstikoormus" on tunduvalt suurem kui konsonantidel. Soomekeelses tekstis esineb vokaaltähti 49% ja konsonanttähti 51% (Setälä, 1972, 11), ungari keeles vastavalt 39% ja 61% (Статистико-комбинаторное моделирование, 1965, 206).

Sageduste hajuvuse analüüs usalduspiiride järgi näitab, et eestikeelsest tekstist võib kindlasti lugeda kõige sagedamateks tähti a, e, i, s, kusjuures viimane neist paikneb alati neljandal kohal. Viiendal kohal võib olla t või l. Hajuvust eri allkeelte vahel väljendab kõige paremini suhteline näitaja - suhteline vigade (vt. tabel 2).

Erinevused tähtede sagedustes peegeldavad erinevusi sõnavaras, eriti sageli esinevates sõnavormides. Nii näiteks esinevad sõnad (sõnavormid) ja, ta, ka ilukirjandusproosas ja kõnekeeles eriti sageli ja seetõttu suureneb nendes allkeeltes ka a-tähe osatähtsus. Kõnekeeles (tegelaskõnes) ka-

sutatakse palju l. isiku asesõna vorme (mina, ma, minu,  
meie, meid jne.), mis loomulikult suurendab m-tähe sagedust.  
 Tabelist 1 on näha, et m-tähte esineb kõnekeeles 4,9 %, s.t.:  
 oluliselt üle keskmise. Selgub samuti, et t-tähte esineb  
 palju ajalehe- ja teaduskeeles, kuna aga luuletekstides on  
t-tähe sagedus oluliselt alla keskmise. Seevastu esinevad  
 luules tähed n ja l sagedamini kui teistes allkeeltes. Ar-  
 vutused näitavad, et a-täht on kõige sagedam neljas allkee-  
 les viiest, kuid teadusalastes tekstides paikneb esikohal a.  
 Seda kinnitab ka H. Holmi (Холм, 1965a) uurimus tähtede  
 sageduste kohta eesti tehnika-alastes tekstides. Meie kat-  
 sematerjali lähemal analüüsimisel selgub, et teadusalastes  
 tekstides on a-täht keskmiselt kaheksal juhul kümnest sa-  
 geduselt esikohal. See seletub teaduskeele eripäraga, ni-  
 melt esineb selle allkeele tekstides palju omastavat käänat  
 lõpphäälikuga a- (näit.: ... uue nõukogude kirjanduse tek-  
kimine ..., tugevate mõjude), alaleütlevat käänat lõpuga -le  
(tavaliaale olukorrale, kesksale kohale), sõnu ja sõnade  
 ühendeid nagu selline, selleks et, sellega seoses, seetõttu,  
sel teel jne. Teadusalasele tekstile on iseloomulikud ka  
 adjektiivid liitega -ne (konkreetne, negatiivne, deskrip-  
tiivne), substantiivid liitega -mine (arenemine, tootmine),  
 rahvusvahelised verbid liitega -eeri- (defineerima, demonst-  
reerima, reprodutseerima) ja muud laen- ning võõrsõnad  
 (probleem, protsess, meetod jt.).

Allkeeltevahelisi erinevusi saab kindlaks teha ka tä-  
 hestike otsese kõrvutamiseega. Eespool vaatlesime ilukirjan-  
 dusproosat terviklikult, s.o. autori- ja tegelaskõnet koos.  
 Üksikute tähtede sagedused erinevad aga oluliselt a u-  
 t o r i- ja t e g e l a s k õ n e s. Seda tõendab meie  
 eriuurimus (vt. tabel 3), millest selgub, et näit. tegelas-  
 kõnes võib sagedamini kohata tähti a, m ja n, autorikõnes  
 aga tähti u ja p (vahe olulisust on kontrollitud z-testi  
 abil; vt. Tuldava, 1970b, 137-138). Täheliste möödete eri-  
 nevus ilmneb ka selles, et vokaaltähti on autorikõnes 45,4,  
 tegelaskõnes aga 47,4 %. Samuti võib täheldada olulist  
 erinevust sagedate tähtede kontsentratsioonil: 5 sagedamat  
 tähte katavad autorikõne tekstist 47,8 %, tegelaskõne teks-  
 tist aga 49,6 %.

Allkeeli võib võrrelda ka teisest seisukohast - läh-

tudes vastavate tähestikkude omavahelisest korrelatsioonist. Korrelatsioonikordaja (r) rakendamine võib esile suure läheduse kõigi allkeelte tähestike kui tervikute vahel:

Allkeel	II	III	IV	V	VI
I Autorikõne	0,986	0,986	0,968	0,987	0,994
II Tegelasõne	-	0,984	0,976	0,979	0,996
III Ajaleht		-	0,985	0,975	0,994
IV Teadus			-	0,950	0,984
V Luule				-	0,985
VI Üldkeel					-

Antud korrelatsioonimaatriksit on võimalik kasutada allkeelte rühmitamiseks dots. L. Vöhandu meetodi järgi (vt: Lepamaa, 1968, 101 jj.). Tulemused esitame skeemi kujul:

Luule - Autorikõne - Üldkeel - Ajaleht - Teadustekst  
 Tegelasõne

Näeme, et moodustub kolm rühma, mis grupeeruvad üldkeele<sup>+</sup> ümber: 1) ilukirjandusproosa autorikõne ja luule, 2) tegelasõne, 3) ajaleht ja teadus- ning tehnika-alased tekstid. Kvantitatiivse meetodi abil saadud hinnangud vastavad intuiitvisele kujutlusele ja kvalitatiivse analüüsi tulemusena kujunenud rühmitustele; näiteks eristab E. Tissenko (Тиссенко, 1971, 353) järgmisi grupeeringuid: 1) ajaleht ja teadustekstid, 2) dramaturgia, 3) ilukirjandus. Käesoleva katse andmetel moodustavad tuumiku ilukirjandusproosa autori- ja tegelasõne ning ajalehekeel, mis on kõige lähemal koondtekstile ja ka omavahel lähedalt seotud. Perifeerias ja teineteisest kaugel asuvad luulekeel ja teaduskeel. Esimene neist on kõige tihedamini seotud ilukirjandusproosa autorikõnega, teine seostub kogusüsteemiga ajalehekeele vahendusel.

Käesoleva eksperimendi tulemused näitavad, et kuigi eri allkeelte tähelised mõõted on antud keele piirides omavahelises korrelatsioonis, mis lubab kõnelda teatud keelele kui tervikule omasest "graafilis-tähestikulisest struktuu-

<sup>+</sup> Üldkeele all mõeldakse antud juhul allkeelte summat, mis puhul saadakse keele keskmised kvantitatiivsed näitajad (koondteksti andmed).

rist" (Трещко, 1971), võib siiski vastavate statistiliste meetodite abil kindlaks teha süsteemisesiseid erinevusi allkeelte vahel. Küllalt suurte valimite korral peaks olema võimalik teksti tähtsageduste alusel allkeeli a u t o m a a t s e l t i d e n t i f i t s e e r i d a. Allkeelte ja stiilide vahelisi erinevusi on suudetud tähtede sageduste abil kindlaks määrata ka mõningates teistes keelestatistilistes uurimustes (näit. ukraina keele suhtes on olulised erinevused allkeelte vahel kindlaks tehtud foneemide tasandil, vt. Статистичні параметри, 1967, 44 jj.): Katsete tulemused kummutavad seega laialt levinud seisukohta, mille järgi tähtede sagedused on antud keele kõigi allkeelte ulatuses täiesti stabiilsed (statistiliselt homogeensed).

Lõpuks mõni sõna varasematest uurimustest tähtede sageduste kohta eestikeelses tekstis (uurimusi häälikute sageduste kohta vaatleme hiljem). Esimene teadaolev uurimus pärineb H. Hansenilt (1961), kes vaatles eestikeelses tekstis esinevate väikeste trükitähtede sagedusi trükikoja ladumismasinamatriitside ratsionaalsema komplekteerimise huvides. Järgnevad H. Holmi (Холми, 1965a, 1965b) uurimused tähtede sageduste kohta eestikeelsetes raadioelektronika ja ajalehetekstides. 1969.a. avaldati "Keeles ja Kirjanduses" Ü. Kaasiku ja E. Laugaste elektronarvuti abiga teostatud uurimuse tulemused tähtede esinemissageduse kohta vanemas ja uemas ajalehekeeles, ilukirjanduses ja rahvalaulus. Seda uurimust täiendati uute andmetega 1975.a. (Kaasik, Laugaste, Ääremaa, 1975). Käesoleva töö autor avaldas 1970.a. andmed viie allkeele tähtede sageduste kohta artiklis "Informatsiooniteooria ja keeleteadus" (Tuldava, 1970a). Kõigi nimetatud uurimuste tulemused, kaasa arvatud ka käesoleva töö andmed, on üksteisele küllaltki lähedased, ja erinevusi võib seletada sellega, et valimite koostis (allkeelte proportsioonid) on olnud eri uurimuste puhul erinev. Tabelis 4 esitatakse H. Hanseni, Ü. Kaasiku, E. Laugaste ja K. Ääremaa uurimuste tulemused koos käesoleva katse andmetega. Tabeli viimases veerus võetakse kokku Ü. Kaasiku, E. Laugaste ja K. Ääremaa andmed (tänapäeva eesti keele kohta) ja käesoleva töö autori uurimuse tulemused. See võimaldab esitada eestikeelse teksti tähtede

sagedused suurema valimi (üldmaht N = 163.340 tähte) põhjal ja järgmistes proportsioonides:

I: ilukirjandusproosa	65970 tähte	-	40 %
kõnekeel (tegelaaskõne)	20914	"	- 13 %
luule	5000	"	- 3 %
<hr/>			
ilukirjandus kokku	91884	"	- 56 %
<hr/>			
II: ajaleht	60098	"	- 37 %
teadus- ja tehnikakeel	11358	"	- 7 %
<hr/>			
"tarbekirjandus" kokku	71456	"	- 44 %
<hr/>			
K o o n d t e k s t	163340 tähte	-	100 %

Sellises "ühendatud" valimis on 10 sagedamat tähte järgmised: a (12,6 %), e (11,0), i (9,6), s (8,7), t (7,1), l (6,3), ü (5,9) k (5,0), n (4,8), d (4,0). Järgneb m (3,9), mis meie koondvalimi andmetel (vt. 5 veerg tabelis 5) on 10. kohal enne d-tähte. Ühendatud valimis katavad 5 sagedamat tähte 49,0 % ja 10 sagedamat tähte 75,0 % tekstist; vokaaltähti on 46,6 %, konsonanttähti 53,4 %. Need andmed on lähedased meie koondvalimi vastavatele näitajatele (48,6 - 75,0 - 46,4 - 53,6; vt. eespool).

Ü. Kaasiku, E. Laugaste ja K. Ääremaa uurimustes (1971 ja 1975) on esitatud huvitavad andmed tähtede sageduse kohta vanemas eesti kirjakeeles ja rahvalaulus. Vanemat kirja-keelt esindavad 5 kirjutist 1860-ndate aastate "Eesti Postimehest" vanas kirjaviisis (valimi maht N = 47700 tähte); rahvalaulu tekstid on võetud J. Hurda "Vana kandle" II köitest ja antoloogiast "Eesti rahvalaulud" (Tallinn, 1969) üldmahuga N = 43250 tähte. Ka vanemas ajalehekeeles ja rahvalauludes on 10 sagedamat tähte samad mis tänapäevakeeles (10. ja 11. kohal vahelduvad m ja d), kusjuures 10 tähte tekstikatvus on vanemas ajalehekeeles 74,5 ja rahvalauludes 76,2 %. Vokaaltähti on vastavalt 47,4 ja 48,3 %. Üldjoontes on tähtede sagedused vanemas kirjakeeles ja rahvalaulus küllaltki lähedased tänapäevakeelele, kusjuures vanem ajalehekeel läheneb enam tänapäeva ilukirjandus- ja kõnekeelele kui tänapäeva ajalehekeelele. Nii vanemas ajalehekeeles kui ka rahvalauludes torkab silma a- ja ä-tähte suur sagedus võrreldes tänapäevakeelega, kuna aga o-tähte on oluliselt vähem; rahvalauludes esineb suhteliselt palju l-tähte, sub-

taliselt vähe aga t-tähte (selle poolest on rahvalaul lähedane tänapäeva luulele). Üeldut illustreerib väljavõtte tabelist 4:

	Vanem aja- lehekeel	Rahvalaul	Tänapäeva- keel
a	13,8 %	13,7 %	12,6 %
ä	2,3 %	2,7 %	1,4 %
o	2,8 %	2,8 %	3,6 %
t	6,9 %	5,7 %	7,1 %
l	5,6 %	7,1 %	6,3 %

.x.x.x.x.

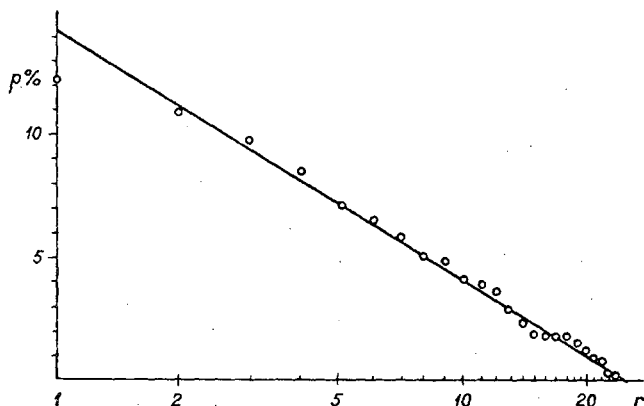
Respool nägime, et tähtede sagedused moodustavad kõigis allkeeltes enam-vähem ühtlaselt langeva rea. See nähtus on omase üldise kõigile teadaolevatele keeltele: Urijad on tundnud huvi sobiva funktsiooni vastu, mis väljendaks seost tähesageduse ja astaku vahel (samuti nagu Zipfi seaduse puhul sõnade sageduse ja astaku vahekorra uurimisel). Prantslased R. Morpeau (1963) ja A. Verglas (1963) näitasid esimestena, et seos tähesageduse ja astaku vahel on antud keele puhul konstantse iseloomuga. Ameeriklane A. Mackay (1965) formuleeris seaduspärasuse logaritmilise regressioonvõrrandi näol:  $p_r = a + b \lg r$ , s.t. tähe suhteline sagedus ( $p_r$ ) on lineaarses seoses astaku logaritmiga ( $\lg r$ ), kusjuures  $a$  ja  $b$  on konstandid. Kontrollides funktsiooni eesti keele andmete põhjal (võttes aluseks tänapäeva keele "suure" koondteksti, vt. 6. veerg tabelis 4), saame tulemuse

$$p_r = 13,7 - 9,6 \lg r.$$

Joonisel 2 on kujutatud logaritmilise horisontaalskallaga diagramm, mis näitab küllaltki head lineaarset seost tähtede sageduse ja astaku logaritmi vahel, kuigi peab noteerima, et kõige sagedam täht hälbib mõnevõrra üldisest tendentsist. Soome keele kohta (Setälä, 1972, 11) on saadud sellele lähedane tulemus:

$p_r = 14,4 - 10,3 \lg r$ . Vene keele andmeil (Kypšamon, 1968, 161) on vastav funktsioon:

$p_r = 11,4 - 7,5 \lg r$ .<sup>\*</sup>



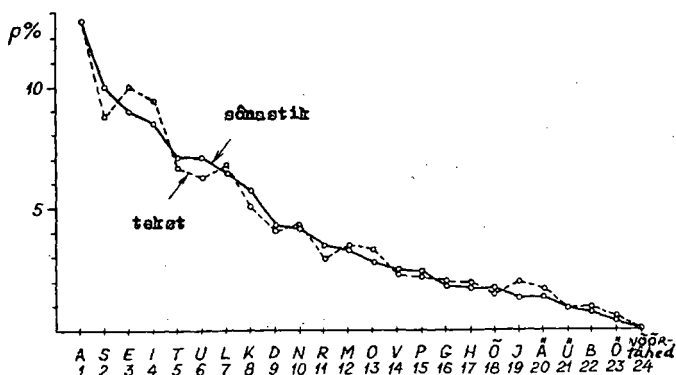
Joon. 2. Seos tähtede sageduse ja astaku vahel (poollogaritmdiagramm).

Tähtede sagedused sõnastikus. Tähtelisi mõtteid võib uurida nii tekstis kui ka sõnastikus. Sõnastikusageduste uurimisel kerkib aga alati küsimus, missugune sõnastik võtta aluseks foneetiliste üksuste vaatlemisel. Kui lähtuda tavalisest lekseemide sõnaraamatust, milles on antud ainult sõnade põhivormid (verbi infinitiiv, substantiivi nimetav käänne jne.), siis peab arvestama, et põhivorm on suvaliselt (kokkuleppeliselt) valitud sõnavorm. Foneetiliste üksuste sageduspilt esineb lekseemide sõnastikus mõnel määral moonutatud kujul, võrreldes tavalise kõnega või tekstiga, näit.

<sup>\*</sup> Kui võtta arvesse ainult 15-20 sagedamat tähte, siis saab tähtede sageduse ja astaku vahelist seost kõigi nimetatud keelte puhul veelgi täpsemalt väljendada eksponentfunktsiooni  $p_r = Ae^{-Br}$  abil, kus A ja B on konstandid. Eesti keele suhtes (20 sagedama tähe põhjal) kehtib  $p_r = 14,2 e^{-0,13r}$ , soome keele puhul  $p_r = 15,7 e^{-0,13r}$  ja vene keele puhul  $p_r = 11,2 e^{-0,1r}$ .

infinitivivormide suure esinemissageduse tõttu tõuseb ülemäära tähtede m ja a sagedus (tunnuse -ma arvel), nimetava käände ühekülgne domineerimine toob endaga kaasa konsonantide sageduse suurenemise (lõpptähe arvel) jne. Järelikult ei ole päris õige lähtuda täheliste või foneetiliste (fonoloogiliste) üksuste statistilisel uurimisel lekseemide sõnastikust (veel enam kehtib see väide sõnapikkuste ja fonotaktiliste nähtuste uurimise puhul). Objektivsema pildi saame, kui vaatleme tähelisi mõtteid sõnavormide sõnastikus. Kuid igal juhul tuleb arvesse võtta sõnastiku mahtu. Peab silmas pidama, et suurtes sõnastikkudes esineb tavalisest rohkem perifeerset sõnavara, näit. eriala- ja võõrsõnu, mille foneetiline koostis erineb igapäevaste sõnade omast. Uurimisel peab seepärast eelnevalt kokku leppima sõnastiku liigi ja mahu suhtes, eriti kui on tegemist võrdleva tüpoloogilise käsitlusega.

Käesolevas töös vaadeldakse paralleelselt kitsapiirilise allkeele teksti ja sõnastikku. On valitud tänapäeva eesti ilukirjandusproosa autorikõne, millest on tehtud tekstivalim 10420 tähe ulatuses ja sõnastikuvalim sõnavormide tasandil üldmahuga 19770 tähte (valimid on võetud 10 eri autori teostest). Tabelis 5 esitame tähtede sagedused tekstis ja sõnastikus ja sageduste vahed. Joonisel 3 on kujutatud graafiliselt teksti- ja sõnastikusagedused viimaste kahaneva rea järgi.



Joon. 3. Tähtede sagedused sõnastikus ja tekstis.

Autorikõne sõnavormide sõnastikus on sagedamad tähed järgmised: a (12,8 %), ä (10,0), e (9,0), i (8,5), t (7,1), u (7,1), l (6,5), k (5,7), d (4,3), n (4,2). Olulisim erinevus, võrreldes tekstisagedustega, on nähtavasti g-tähe nihkumine teisele kohale sagedusjärjestuses; sageduste vahe on 1,2 % sõnastiku kasuks (vt. tabel 5), mis on statistiliselt oluline. Sisuliselt tähendab see seda, et g-täht esineb "laisalipaisatult" väga paljudes erinevates sõnades (sõnavormides) ja ei ole kontsentreerunud teatava väiksema rühma sõnadesse. Sama võib öelda u, k, r, õ kohta, kuigi vähemal määral (teksti- ja sõnastikusageduse vahe on vähemalt 0,3 %). Nende tähtede tekstikoormus (funktsionaalne koormus teksti tasandil) on suhteliselt väike, kuid "informatiivsus" suur, s.t. nende esinemine tekstis signaliseerib keskmiselt rohkem erinevate sõnade või vormide ilmumisest kui teised tähed. Suhteliselt suure tekstikoormusega on aga tähed e, i, l, o, j, ä, mille tekstisagedused ületavad sõnastikusagedusi. Kui mõõta tekstikoormust tekstisageduse (T) ja sõnastikusageduse (S) suhtega, siis näeme, et suurim T/S väärtus on i-tähel ( $2,0/1,4 = 1,43$ ). See seletub peamiselt sõna ja suure sagedusega tekstis (sõnastikus üksainus vorm). Ka e ja i tekstikoormust mõjustab suurel määral sõnavormi ei suur sagedus. Tähe ä suhtes tuleb nentida, et sõnavarasse kuulub võrdlemisi vähe ä-lisi sõnu, kuid mõni neist esineb tekstis sageli, näit. ära, pärast, välja, jälle, läks.

Vaadeldes vokaalide ja konsonantide sageduste suhet, näeme, et tekstis on suhe 46,4:53,6 ja sõnastikus 44,6:55,4. Järelikult on vokaalide osatähtsus tekstis mõnel määral suurem kui sõnastikus:

Viis sagedamat tähte katavad autorikõne tekstist 47,8 % ja sõnastikust 47,3 %, kümme sagedamat tähte aga vastavalt 74,4 % ja 75,1 %. Need arvud on küllaltki lähedased. Ka tähestikusageduste üldine struktuur (vt. joon. 3) on samalaadne. Teksti- ja sõnastikusageduste vaheline korrelatsioon on erakordselt kõrge ( $r = 0,988$ ). Võib teha üldise järelduse, et teksti ja sellele tekstile vastava sõnastiku sagedused on tõepoolest väga lähedased. Üksikute tähtede puhul ilmnevad mõningad erinevused, mis seletuvad sellega, et näit. rida sõnu esineb tekstis suure sagedusega, kuna aga sõnastikus on neil ainult üks koht. Selle ebaproportsionaal-

suse arvel toimub ka mõningate tähtede sageduse suurenemine tekstis, võrreldes vastava sõnastikuga (ja ümberpöörduvalt).

Häälikute sagedused tekstis. Eesti keeles on traditsioonilise arvestuse järgi 20 häälikut: 9 täishäälikut ja 11 kaashäälikut (tugev ja nõrk sulghäälik loetakse üheks häälikuks, seega  $k+g$ ,  $p+B$ ,  $t+D$  annavad kolm häälikut; kokku arvestatakse ka  $n+\eta$ ).<sup>+</sup> Peale selle võivad esineda vöörhäälikud ( $f$ ,  $\check{z}$  ja helilised  $\underline{z}$ ,  $\underline{\check{z}}$ ,  $g$ ,  $\underline{b}$ ,  $\underline{d}$ ), murdeti ka rida teisi häälikuid (vt. Ariste, 1968, 90-91). Ühendeid  $ts$ ,  $tš$ ,  $dž$ ,  $dž$  käsitletakse vahel afrikaatidena, kuid silmas pidades seika, et eesti keeles paikneb silbipiir keset afrikaati (täpsemalt keset afrikaadi sulghäälikulist ossa) võime nimetatud ühendites arvestada kahte häälikut.

Häälikute sagedusi eestikeelses tekstis uuris esimesena A. Saareste (1932), kuid kahjuks on tema andmed trükkis avaldatud ainult osaliselt ja pole ka teada, kui suure valimi ja mis laadi tekstide alusel sooritati häälikute loendus. Põhjalikuma uurimuse teostas arstiteadlane V. Särgava (1967), kes vaatles häälikute sagedusi tänapäeva ajalehe- ja ajakirjatekstides üldmahuga umbes 30.000 häälikut (uurimuse tulemusi kasutati kõneaudiomeetriliste katsete läbiviimisel).

Käesolev töö on mõeldud varasemate uurimuste täienduseks, kusjuures valimi materjali on mitmekesistatud ja analüüsitud lingvistilisest seisukohast. Valim on koostatud 4 allkeele põhjal (ilukirjandusproosa autorikõne, tegelaskõne, ajalehe- ja populaarteaduslik tekst - kõik võrdsetes osades), valimi üldmaht, s.o. teksti pikkus  $N = 23.560$  häälikut. Tabelis 6 on toodud nii meie katse tulemused kui ka A. Saareste ja V. Särgava andmed häälikute sageduse kohta eestikeelses tekstis ("kirjalikus kõnes"). Kõigi kolme katse tulemused langevad üldjoontes kokku. Väikesed erinevused üksikute häälikute sagedustes (näit. on käesoleva töö autori läbiviidud katse andmeil häälikut  $t/D$  rohkem ja  $k/G$  vähem kui V. Särgava uurimuse tulemuste järgi) tulenevad ilmselt valimite erinevast koostisest allkeelte lõikes. Ka

---

<sup>+</sup> Leenisklusiliid  $G$ ,  $B$  ja  $D$  esinevad kirjas kui  $g$ ,  $b$ ,  $d$ . Kombinatorset nasaali  $\eta$  eesti keele kirjas eri märgiga ei tähistata.

häälikute puhul peab arvestama mõningaid erinevusi allkeelte vahel, nagu seda võisime nentida tähtede sageduste uurimisel.

Eesti keele häälikud ja tähed on teatavasti tihedas korrelatsioonis, kusjuures peamine erinevus seisneb selles, et pikki häälikuid väljendatakse vastavate tähtede kahekordse kirjutamise teel. Sellest tulenevalt on ühe ja sama teksti ulatuses alati rohkem tähti kui häälikuid. Antud katse puhul oli valimis 24 796 tähte ja 23 560 häälikut, seega on suhe 105:100. Võrdluseks toome andmed mõnede teiste keelte kohta (Meier, 1964, 321): itaalia keel - 104:100, saksa keel 112:100, hollandi keel - 114:100, taani keel - 124:100, prantsuse keel - 148:100. Nagu näha, on eesti keel lähedane itaalia keelele ortograafia ökonoomsuse poolest. Pikkade häälikute kahekordne kirjutamine eesti keeles ei koorma seega kuigivõrd meie õigekirjutussüsteemi.

Tabelis 7 esitame uuesti meie katse andmed häälikute sageduse kohta, märkides eraldi ühe ja kahe tähega kirjutatud häälikute sagedused. Kokkuvõttes on eestikeelses tekstis lühikesi, s.o. ühe tähega kirjutatud vokaale 92 % ja pikki, s.o. kahe tähega kirjutatud vokaale - 8 %. Soome keeles on lühikeste ja pikkade vokaalide suhe samuti 92:8 (arvestused V. Setälä uurimuse andmete põhjal). Ühe tähega kirjutatud konsonanthäälikuid on eestikeelses tekstis 97 % ja kahe tähega kirjutatud konsonante - 3 %. Soome keeles on vastav suhe - 90:10. Siinjuures tuleb aga arvestada, et eesti keeles on tegelikult pikad ka II- ja III-väitelised geminaatklusiilid, mis kirjas märgitakse ühe tähega (näit. lepib [leppiB], seepi [séppi]), samuti sõnalõpulisel ühe tähega kirjutatud fortisklusiilid (näit. kompevk, taburet). Nimetatud juhtude kaasahaaramisega suureneks "pikkade" konsonanthäälikute osakaal ligikaudu 13 %-ni (kõigist konsonanthäälikutest tekstis).

Kõige sagedamad häälikud eestikeelses tekstis on käesoleva uurimuse andmeil a (12,2 %), t/D (11,9), e (11,0), i (9,5), s (9,0). Järgnevad k/G, l, u, n/ŋ, m, o, r, p/B, v. Väikese sagedusega (igauks alla 2%) on j, h, ä, õ, ü, ö ja võõrhäälikud. Võrdluseks toome viis sagedamat häälikut soome keeles (Setälä, 1972, 38):

a (11,6 %), n (10,2), i (10,2), e (9,4), t (8,9). Viis sagedamat häälikut katavad eestikeelsest tekstist 53,6 %, soomekeelsest 50,3 %. Seega mõjustavad sagedamad häälikud küllaltki oluliselt häälduse üldpilti. Võrreldes mõlemat sugulaskeelt, näeme, et kuigi viie sagedama hääliku hulgas on neli ühist (a, t, e, i), seisneb erinevus selles, et soomekeelses kõnes on n-häälik suure esinemissagedusega (eesti keeles on n 9. kohal sagedusega 4,3 % soome 10,2 % vastu), kuna aga a- ja t-häälikut esineb soome keeles oluliselt harvemini (a-hääliku sagedus on soome keeles 7,0 %).

Viis sagedamat häälikut ungari keeles on L. Hakuline ni (1941) andmeil: a (14 %), ä (13 %), t (8 %), i (6 %), n (5,5 %). Mansi keeles reastab B. Kálmán (1963) viis sagedamat häälikut järgmiselt: a (10,1 %), t (9,0 %), i (7,0 %), l (5,8 %), m (5,3 %). Kõigile ülalnimetatud soome-ugri keeltele on ühised suure sagedusega häälikud a ja t (viie sagedama hääliku hulgas).

Huvipakkuv on vaadelda v o k a a l i d e ja k o n s o n a n t i d e suhet eesti keeles, võrreldes teiste keeltega. Meie koondvalimi (4 allkeelt) andmeil esineb eestikeelses tekstis vokaalhäälikuid 45,5 % ja konsonanthäälikuid 54,5 %, seega 100 vokaali kohta 120 konsonanti (A. Saareste andmeil 117). Soome keeles on täis- ja kaashäälikute suhe 48,2:51,8, s.o. 100 vokaali kohta 108 konsonanti.

Täis- ja kaashäälikute sageduste suhet on alati peetud tähtsaks karakteristikuks keelte tüpoloogilisel uurimisel. Juba B. Bourdon (1892) jaotas keeled vokaalseteks, konsonantseteks ja "segatüüpi" keelteks. J. Krámský (1948) mõõtis keelte "vokaalsust" erilise koefitsiendi abil:  $v = \frac{p_1}{p_t}$ , kus  $p_1$  tähistab kaashäälikute protsenti süsteemis (häälikuvaryus) ja  $p_t$  - kaashäälikute protsenti tekstis. Kui arvestada ainult eesti oma häälikuid (11 konsonanti, 20 hääliku kohta), saame  $p_1 = 55\%$  ja vokaalsuskoefitsiendi  $v = 55:54,5 = 1,01$ . Soome keeles on 13 konsonanti 21 hääliku kohta, seega  $p_1 = 62\%$  ja  $v = 62:51,8 = 1,20$ . Võrdluseks võib tuua J. Krámský andmed mõne teise keele kohta: saksa keel - 0,85, inglise keel - 0,91, hispaania keel - 1,46, itaalia keel - 1,58. Antud juhul on nähtavasti õigem arvestada eesti keeles ka suhteliselt kodunenud võõrhääli-

kuid (näit. f ja š). Konsonante oleks siis kokku 13 ehk 59 % häälikute üldarvust (22). Vokaalsuskoefitsiendi väärtuseks saame  $v = 59:54,5 = 1,07$ . Veelgi kõrgema vokaalsuse hinnanagu saaksime, kui arvestaksime kõiki eesti keeles esinevaid võõrhäälikuid.

Keelte "vokaalsusastet" võib objektiivselt määrata ka tekstisageduste põhjal, võttes aluseks suhte  $p_v/p_k$ , kus  $p_v$  tähistab vokaalide ja  $p_k$  konsonantide sagedust tekstis, arvutame eri keelte vokaalsusastme  $v_t = p_v/p_k$  (sulgudes närgime konsonantide arvu 100 vokaali kohta tekstis), ja jaotame keeled vokaalsusastme järgi kolme rühma, näiteks:

$v_t > 1$ : tahiiti 2,40 (41), hawaii 1,54 (65), jaapani 1,22 (82);

$0,70 \leq v_t \leq 1$ : portugali 0,98 (102), ruumeenia 0,94 (106), itaalia 0,93 (108), soome 0,93 (108), esperanto 0,87 (115), prantsuse 0,87 (116), hispaania 0,85 (118), eesti 0,83 (120), leedu 0,78 (129), läti 0,72 (138), ungari 0,71 (141);

$v_t < 0,70$ : mansi 0,67 (149), poola 0,67 (148), vene 0,67 (150), slovaki 0,67 (150), inglise 0,63 (158); rootsi 0,63 (158), saksa 0,61 (164).

Andmed häälikute sageduste kohta on võetud eri uurimustest (Weiss, 1961; Meier, 1964; Zsilka, 1971; Setälä, 1972; Саеневрвдс, 1966; Бенсона, 1973). Antud juhul on tegemist peamiselt ilukirjandusproosa või ajalehetekstidega. Ei tohi aga unustada, et vokaalide ja konsonantide sagedused võivad kõikuda allkeelte kaupa. Nii näiteks esineb M. Weissi (1961) andmeil rootsi ajalehetekstis 160 konsonanti 100 vokaali kohta (vokaalsusaste  $v_t = 0,63$ ), kuna aga telefonikõnedes on vastav sagedussuhe 138:100 ( $v_t = 0,72$ ). Vokaalide osatähtsus tõuseb seega rootsi kõnekeeles märgatavalt, võrreldes kirjutatud tekstiga. Sama kehtib ka eesti keele kohta. Meie uurimistöös andmeil esineb tegelaskõne osavalimis täishäälikuid 46,8 % ja kaashäälikuid 53,2% (koondvalimis oli vahekord 45,5:54,5). See tähendab, et 100 vokaali kohta tuleb kõnekeelt imiteerivas ilukirjandusproosa tegelaskõnes 114 konsonanti (vokaalsusaste  $v_t = 0,88$ )<sup>3</sup>.

Analüüsimise lähemalt eesti keele täis- ja kaashäälikute sagedusi tekstis, võrreldes neid andmetega soome, ungari jt. keeltest.

V o k a a l i d järjestuvad oma sageduste põhjal eestikeelses tekstis järgmiselt (siin ja järgnevalt on võetud aluseks koondvälimi, s.o. 4 allkeele ühendteksti andmed; protsendid on arvatud täishäälikute koguarvust tekstis, sulgudes on antud lühikeste/pikkade täishäälikute protsent):

<u>a</u>	26,8	(25,7/1,1)
<u>e</u>	24,2	(22,2/2,0)
<u>i</u>	20,8	(19,6/1,2)
<u>u</u>	13,2	(12,2/1,0)
<u>o</u>	6,8	(5,4/1,4)
<u>ä</u>	2,9	(2,4/0,5)
<u>õ</u>	2,9	(2,8/0,1)
<u>ü</u>	2,0	(1,7/0,3)
<u>ö</u>	0,4	(0,0/0,4)

kokku 100,0 % (92,0/8,0)

Soome keeles on 8 täishäälikut (ö puudub). Sagedamad täishäälikud on a (24,1 % täishäälikute koguarvust tekstis), i (21,2 %), e (19,5 %), ä (12,2 %) ja o (10,4 %). Järgnevad u (9,5 %), ü (2,5 %, ortograafias y) ja õ (0,3 %). Kahe keele võrdlemisel paistab silma, et eesti keeles esineb rohkem e-, a- ja u-häälikut, kuna aga soome keeles on oluliselt rohkem o- ja eriti ä-häälikut (viimast esineb soomekeelses tekstis 12,2 % eesti 2,9 % vastu).

Võrdluseks toome andmed ka ungari keele täishäälikute sageduse kohta tekstis (Kálmán, 1963): a (á) - 26,0 %, a - 23,6 %, o+ó - 12,5 %, i+í - 10,7 %, ä - 8,4 %, é - 8,4 %, ö+ő - 4,9 %, u+ú - 3,6 %, ü+ű - 1,9 % (kokku 100,0 %).

Pikkudest täishäälikutest on eesti keeles esikohal pikk e (kirjutatud ee), mis moodustab 2,0 % vokaalide kogusisenumusest tekstis (näit. sagedates sõnades see, veel, mees, ees, sees, teeb). Järgnevad oo (1,4 %), ii (1,2 %), aa (1,1 %) ja uu (1,0 %). Pikka varianti kohtab kõige rohkem ä-hääliku juures (tää, öö, sööma, mööda jne.), kuna aga ö esineb peamiselt lühikesena (õber, lõbus, õnn, yöi

jt.). Soome keele sagedamad pikad täishäälikud on aa (2,0 % kõigist vokaalidest tekstis), ii (1,5 %), ee (1,3 %), ää (1,2 %), uu (0,9 %).

Rühmitades eesti keele täishäälikud moodustuskoha järgi ees-, kesk- ja tagavokaalideks, saame statistilise uurimise alusel järgmised tulemused.

**E e s v o k a a l i d :**

labialiseerimata	<u>i</u> , <u>e</u> , <u>ä</u>	47,9 %
labialiseeritud	<u>ü</u> , <u>ö</u>	2,4 %
		<hr/>
kokku		50,3 %

**T a g a - j a k e s k v o k a a l i d :**

velaarsed	<u>u</u> , <u>o</u> , <u>a</u>	46,8 %
velaarpalataalne	<u>õ</u>	2,9 % <sup>+</sup>
		<hr/>
kokku		49,7 %

Taga- (ja kesk-) ning eesvokaalide suhe on eesti keeles ligikaudu 50:50. Soome keeles on taga- ja eesvokaalide suhe V. Setälä järgi 44:56 (L. Hakulineni andmetel 43:57). Ungari keeles on suhe 48:52 (Kálmán, 1963). Kui eesti keeles jätta arvestusest välja õ-häälik (keskvokaal), siis on ka eesti keeles taga- ja eesvokaalide suhe 48:52. Nimetatud kolmes keeles on eesvokaalide osatähtsus küllaltki suur. Võrdluseks võib tuua ungari naaberkeele - slovaki keele (Zsilka, 1971, 111), kus tagavokaale on 57 % eesvokaalide 43 % vastu. Leidub keeli, kus tagavokaalid on suures ülekaalus (näit. sanskritis on suhe 80:20).

K o n s o n a n t i d e sagedused on meie katse andmeil eesti keeles järgmised (protsendid on arvestatud kaashäälikute koguarvust tekstis):

<sup>+</sup> Täpsemalt võttes on sinult lühike õ suhteliselt kesk- vokaal; pikk õ moodustatakse u keeleasendiga ja on seega tagapoolsem kui o ja a (T.-R. Viitso märkus). Pikk õ esineb tekstis aga suhteliselt harva (alla 0,1 %).

t }	13,4 }	21,8
D }	8,4 }	
s		16,5
k }	9,7 }	13,4
G }	3,7 }	
l		11,4
n }	7,9 }	8,5
ŋ }	0,6 }	
m		7,3
r		5,3
p }	3,5 }	4,8
B }	1,3 }	
v		4,2
j		3,5
h		3,1
võõr-		0,2
konsonandid		

---

Kokku 100,0 %

Soome keeles on kaashäälikuid rohkem kui eesti keeles, nimelt lisanduvad heliline d ja geminaat n (mida märgitakse ng-ga). Peale selle võivad samuti nagu eesti keeleski esineda mõned võõrkonsonandid (f, h, g jt.). Soome kaashäälikute sagedused tekstis on V. Setälä (1972, 38) andmete alusel järgmised: n 19,7 %, t 17,2 %, s 13,5 %, k 10,0 %, l 9,3 %, m 6,4 %, j 5,4 %, h 5,4 %, v 4,6 %, r 3,1 %, r 2,9 %, d 1,7 % ja ŋ 0,8 % (konsonantide üldarvust). Eesti keelega võrreldes esineb soome keeles märgatavalt rohkem n-häälikut (peamiselt sõnavormi lõpus), samuti h-häälikut (eeskätt algushäälikuna).

Pikkadest kaashäälikutest, mida kirjutatakse kahe tähega, on eesti keeles erikohal ll sagedusega 1,0 % kõigest konsonantidest tekstis (näit. sagedates sõnavormides selle, mille, talle, mulle, jälle, olla, kõll, all). Järgnevad nn (0,5 %), tt (0,4 %), kk (0,4 %), ss (0,3 %), mm (0,2 %), rr (0,1 %), pp (0,1 %). Eesti keeles võivad esineda ka vy ja hh, kuid nende esinemus kõnes on tühine. Võrdluseks toome sagedamad kahe tähega kirjutatud konsonandid soome keeles: ll (1,5 %), tt (1,5 %), nn (0,5 %), ss (0,5 %). Nagu varem juba nimetatud, kuuluvad pikkade kaashäälikute hulka eesti keeles ka mõned sõnasisesed ja -lõ-

pulised klusiilid, mida märgitakse ühe tähega (näit. k sõnades pika, kompvek).

M o o d u s t u s v i i s i järgi saame järgmised eesti kaashäälikute rühmad koos esinemissagedusega tekstis:

klusiilid	<u>k/G</u> , <u>p/B</u> , <u>t/D</u>	40,0 %
spirandid	<u>i</u> , <u>e</u> , <u>l</u> , <u>r</u> , <u>v</u> , <u>h</u>	
(kaasa arvatud vöörhäälikud		
	<u>f</u> ja <u>š</u> )	44,2 %
nasaalid	<u>n/ŋ</u> , <u>m</u>	15,8 %
		<hr/>
kokku		100,0 %

Soomekeelses tekstis esineb klusiile 31,7 %, spirante 39,5 % ja nasaale 26,8 %. Järelikult on klusiilide ja spirantide osatähtsus eestikeelses kõnes oluliselt suurem kui soome keeles, kuna aga nasaalide poolest on rikkam soome keel (tegelikult n-hääliku arvel, sest m-häälik esineb eesti keeles pisut sagedamini) Spirantidest esinevad eesti keeles oluliselt sagedamini e, l ja r, kuna aga soome keeles on sagedamad h, v ja i.

M o o d u s t u s k o h a järgi liigitatakse eesti konsonante järgmiselt (Ariste, 1968, 90-91):

labiaalid	<u>p/B</u> , <u>m</u> , <u>v</u> , ( <u>f</u> )	16,5 %
alveodentaalid	<u>t/D</u> , <u>n</u>	29,7 %
velaarpalataalid	<u>k/G</u> , <u>ŋ</u> , <u>s</u> ,	
	<u>i</u> , <u>l</u> , <u>r</u> , ( <u>š</u> ), ( <u>z</u> ), ( <u>ž</u> )	50,7 %
larüngaal	<u>h</u>	3,1 %

---

kokku 100,0 %

Arvestades keeletipu (apex) ja keeleselja (dorsum) aktiivsust alveodentaalide ja velaarpalataalide moodustamisel, võime nimetatud konsonandid omakorda jagada järgmisteks rühmadeks:

predorsaal-apikaalid	<u>t/D</u> , <u>n</u> , <u>s</u> , ( <u>z</u> ),	62,9 %
	<u>l</u> , <u>r</u>	
mediodorsaalid	<u>i</u> , ( <u>š</u> ), ( <u>ž</u> )	3,5 %
postdorsaalid	<u>k/G</u> , <u>ŋ</u>	14,0 %
		<hr/>
kokku		80,4 %

(Ülejäänud 19,6 % on labiaalid ja larüngaalid.)

Paistab silma keeletipu- ja eeskeelshäälikute (predor-saal-apikaalide) suur osatähtsus eestikeelses kõnes (62,9% kõigist konsonantidest ja 34,3% kõigist häälikutest). Nähtus on omene väga paljudele keeltele, nii näiteks on J. Krámský (1959) poolt uuritud 23 eri keelkondadesse kuulvas keeles predorsaal-apikaalide esinemissagedus üle 50% (kõigist konsonantidest).

Soome keeles on labiaale 14,1%, predorsaal-apikaale 64,3%, medio-dorsaale 5,4%, post-dorsaale 10,8% ja larüngaale 5,4%. Erinevus eesti keelest ei ole kuigi suur, eriti kui ühendada tagapoolsed kaashäälikud (post-dorsaalid ja larüngaalid) ühte rühma. Eesti keeles on neid 17,1%, soome keeles 16,2%.

H e l l i s e d konsonandid on eesti keeles l, m, n, r, v (välja arvatud sõna lõpus, kui neile eelneb h või s, näit. lehm, mahl, rasv; vt. Ariste, 1968, 42). Peale selle hääldub heliliselt j ja teatavatel juhtudel ka h (Ariste, 1968, 71 ja 72). Võttes arvesse ülalmainitud, võime meie materjali põhjal konstateerida, et eestikeelses tekstis (kõnes) esineb helilisi kaashäälikuid ligikaudu 41% ja helituid 59%. Kõigi häälikute ulatuses (kaasa arvatud vokaalid, mis on helilised), on heliliste ja helitute häälikute suhe 68:32.

Soome keeles (tekstis) on umbes 53% helilisi kaashäälikuid (suur osakaal on n-häälikul) ja 48% helituid. Kõigi häälikute ulatuses on heliliste ja helitute suhe 75:25. Võrdluseks võib tuua, et näit. saksa keeles moodustavad helilised konsonandid 63% kõigist kaashäälikutest ja üldsuhe on 78:22 (Meier, 1964, 252), rootsi keeles on vastavad arvud 66% ja 80:20 (Weiss, 1961, 51).

Eelõeldust ilmneb, et eesti keeles on mõnevõrra vähem helilisust kui mõnedes teistes keeltes. Kuid seda korvab asjaolu, et vokaalide osatähtsus on eesti keeles küllaltki suur ja kõnes esineb sageli kõlv l-häälik (eesti keeles 6,2%, soome keeles 4,8%, rootsi keeles 4,1%, saksa keeles 4,0%, vene keeles 3,5%, läti keeles 2,5%).

Esitame lõpuks ülevaate viiest sagedamast kaashäälikust kümnes eri keeles:

eesti keel	t/D	s	k/G	l	n
soome keel	n	t	s	k	l
ungari keel	t	l	n	m	k
mansi keel	t	l	m	ʃ	j
läti keel	s	r	t	k	n
leedu keel	s	k	t	r	n
vene keel	n	t	s	r	v
inglise keel	n	t	s	θ	d
saksa keel	n	r	t	d	s
rootsi keel	n	t	r	d	s

Torkab silma suur lähedus selliste sugulaskeelte vahel nagu läti ja leedu keel, saksa ja rootsi keel. Ka eesti ja soome keeles on kõik viis sagedamat kaashäälikut ühised (ka järjekord on sama, välja arvatud n-häälik). Testavat sarnasust võib märgata ungari ja mansi keele konsonantide jaotumuses. Kõigis vaadeldavates keeltes, välja arvatud ungari ja mansi keel, on viie sagedama konsonandi seas häälikud n, s, t.

Algus- ja lõpptähtede (resp. -häälikute) sagedused tekstis ja sõnastikus. Tähestikuliste mõõdete seas on oma kindel koht tähtede (häälikute, foneemide) sagedustel p o s i t s i o o n i j ä r g i sõnas või sõnavormis. Eri- list tähtsust omavad algus- ja lõpptähtede (häälikute, foneemide) sagedused, mida kasutatakse mitmesuguste lingvistilis-tüpoloogiliste ülesannete lahendamiseks (näit. Амреес, 1967; Maneca, 1967). Olulisi tulemusi on saavutatud vene ja ukraina keele morfoloogilisel automaatanalüüsimisel sõnulõpu tähtede ja tähekombinatsioonide statistika abil (Бело-короб, 1971; Савченко, 1970). Huvipakkuv on G. Herdani (1963) katse kindlaks määrata inglise keele sõnavara etümo- loogiliste komponentide muutumist eri perioodidel sõnade algustähtede sageduse uurimise teel. Tähtede positsioonili- sed sagedused tekstis ja sõnastikus on olulise tähtsusega ka mõningate informaatika-alaste probleemide uurimisel, näit. sõnade komprimeerimisel (vt. Купцов, 1968, 163 jj.).

Eesti keele tähtede positsioonilisi sagedusi on varem uurinud H. Holm (Хольм, 1965a ja 1965b) ning Ü. Kaesik ja E. Laugaste (1969). Esitame täienduseks ja sageduste sta- biilsuse kontrollimiseks käesoleva uurimistöõ andmed, kus- juures toome võrdlusmaterjali soome jt. keeltest ning vaat-

lame algus- ja lõpptähtede (-häälikute) sagedusi mõnest uuest aspektist. Lisame ka andmed sõnastikusageduste kohta, mida seni pole eesti keeles uuritud. Meie koondvalim koosneb 4 allkeelse osavalimiteest (ilukirjandusproosa autorikõne, tegelaskõne, populaarteaduslik tekst ja ajaleht) üldmahuga 4097 sõnet. Uurimuse tulemusena saadud algus- ja lõpptähtede sagedusjärjestused on eesti keeles võrdsed vastavate häälikute sagedusjärjestusega (häälikute puhul tuleb ainult ühendada k ja g, t ja d, p ja b sagedused).

Tabelis 8 esitatakse algustähtede sagedused tekstis. Kõik viis sagedamat algustähte on konsonandid: k (14,1 %), t (9,5 %), g (8,8 %), m (8,3 %), p (7,4 %). Järgnevad o, v, a, i, ä, ü, l, h, r, f, ü, ö, u, ä, d, f, b, ö. Viis sagedamat algustähte katavad tekstist 48,1 % ja kümme sagedamat 78,9 %. Kokku on algustähtede hulgas 74,8 % konsonante ja 25,2 % vokaale. Ü. Kaasiku ja E. Laugaste uurimuse andmeil (ilukirjandus- ja ajalehetekstid) on viis sagedamat algustähte: k (13,2 %), t (10,7 %), g (9,2 %), p (7,8 %), v (7,5 %), mis kokku moodustavad 48,4 %. H. Holmi andmeil (publitsistlik tekst) on algustähtede järjestus: k, t, p, g, v. Kõigile nimetatud uurimustele on seega ühised k, t, g, p, kusjuures tähtedele k ja t kuuluvad kindlalt kaks esimest kohta. Meie uurimistöö andmeil on m-täht võrdlemisi suure sagedusega; see tuleneb tegelaskõne kaasatõmbamisest koondvalimisse. Tabelist 8 nähtub, et peale ühiste joonte on allkeeltele ka mõned olulised erinevused. Näiteks on tegelaskõnes esikohal m-täht, mis seletub l. isiku asesõna ohtra kasutamisega (mina, mind, meie jt.). Tegelaskõnes esinevad keskmisest oluliselt sagedamini ka algustähed g ja a (sõnavormide sina, sind, sa, ta, see, seda, aga jt. sageda kasutamine), kuna aga näiteks y-täht esineb märksa harvemini, võrreldes teiste allkeeltega. Erinevused mõnede algustähtede sageduste vahel eri allkeeltes on küllaltki suured, näit. k-tähte esineb tegelaskõnes 11,7 %, ajalehes aga 17,3 %. Osa tähti esineb allkeeltele lõikes stabiilselt, näit. t (sagedused: 8,6 - 9,6 - 9,7 - 10,0 %).

Soome keeles on V. Setälä (1972) järgi sagedamad algustähed j (13,0 %), g (10,9 %), k (9,9 %), h (9,5 %), t

(9,3%). Konsonante on algustähtede hulgas 80% ja vokaale 20%. Viis sagedamat algustähte katavad tekstist 52,6%. Võrreldes eesti keelega seisneb peamine erinevus selles, et sageli esinevad algustähtedena j ja h. Kui aga vaadelda algusvokaalide sagedusi, siis ilmneb, et sagedusjärjestus on mõlemas keeles sama: o, e, a, i.

Ungari keele publitsistlikus tekstis on kuus sagedamat algustähte k, e, s, v, m, t (Статистико-комбинаторное моделирование, 1965, 207). Ühised sagedamini esinevad algustähed (resp. -häälikud) on kolmes sugulaskeeles k, t ja s.

Võrdluseks toome andmed uurimustest teiste keelte kohta (Статистико-комбинаторное моделиров., 1965; Петрова, 1968; Meier, 1964). Läti keel: p, s, a, i; vene keel: n, c, o, r, к; inglise keele: s, t, w, f, c; saksa keel: d, s, w, f, u; prantsuse keele: a, d, l, s, e.

Lõpptähtede sagedused tekstis on arvatud samade allkeelte ja koondteksti alusel, mida kasutati algustähtede uurimisel.<sup>+</sup> Tulemused esitatakse tabelis 9. Koondtekstis on kõige sagedamad lõpptähed a (19,1%), e (17,7%), s (13,0%), d (11,6%), i (11,6%). Järgnevad t, l, n, u, b, k, m, g, r, h, v, o, p, ö. Tähelepanuväärne on seik, et viis sagedamat lõpptähte moodustavad 73,0% kõigist lõpptähtedest tekstis, 10 sagedamat lõpptähte aga 96,2%. Vokaale on lõpptähtede hulgas 52,7% ja konsonante 47,3%. Seega võib nentida olulisi erinevusi, võrreldes algustähtede sagedustega. Eestikeelses kõnes on ülekaalus vokaallõpud, mida peetakse parema kuuldavuse tingimuseks (Maneca, 1967).

Varasemate uurimuste järgi on eestikeelses tekstis sagedamad lõpptähed a, e, s, i, d (Kaasik, Laugaste, 1969; Kaasik, Tuldava, 1979); e, a, d, s, i (Холы, 1965). Seega langevad kõigis uurimustes viis sagedamat lõpptähte kokku, kuigi järjestus pole täpselt sama.

Tabelist 9 nähtub, et eri allkeelte osas võib kohati täheldada olulisi erinevusi, näit. on teaduslikes tekstides esikohal e sagedusega 23,9%, kuna aga sama lõpptähte

<sup>+</sup> Vt. ka eesti keele sõnavormide pöörsagedussõnastiku materjalide põhjal saadud andmeid autorikõne kohta (Kaasik, Tuldava, 1980).

sagedus teistes allkeeltes kõigub 15-18 % vahel (vrd. lk. 77). Tegelasõnes esinevad a ja n lõpptähena palju sagedamini kui teistes allkeeltes.

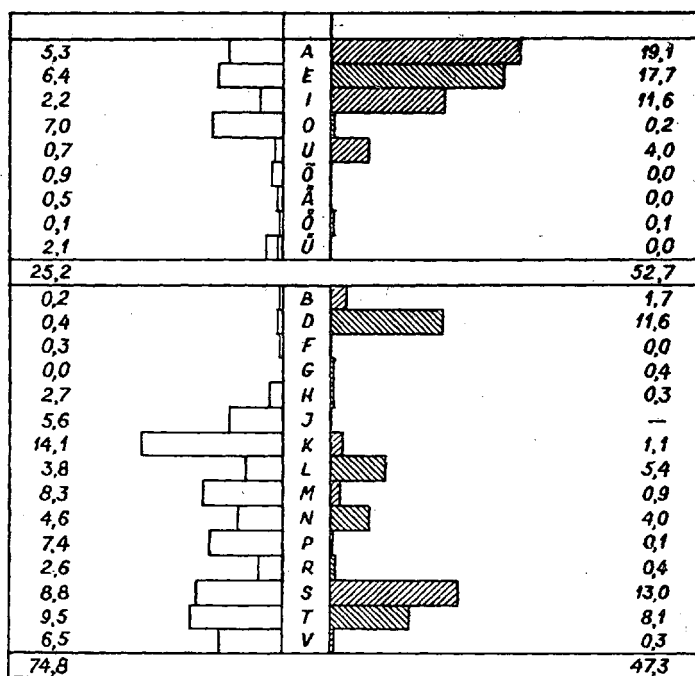
Soome keeles (Setälä, 1972, 26) on lõpptähtede sagedusjärjestus niisugune: n (28,1 %), a (23,7 %), ä (13,1 %), i (11,0 %), e (10,1 %). Viis sagedamat lõpptähte moodustavad kokku 86,0 %, seega veelgi rohkem kui eestikeeles (73,0 %). Järgnevad t, s, o, u, ö, y, r. 10 sagedamat tähte katavad 99,9 % lõpptähtede kogusinemusest. Vokaale on 60,7 % ja konsonante 39,3 %.

Nii eesti kui soome keeles ilmneb lõpptähtede kontsentratsioon, s.t. vähesed tähed (resp. häälikud) katavad suure osa kõigist lõpptähtedest (-häälikutest). Eesti keeles esinevad sõna lõpus harva niisugused tähed nagu h, v, o, p, ö ja väga harva tähed ü, ä, õ ning võõrtähed. Eesti täht i võib esineda lõpptähena ainult võõrnimeses (näit. Raj Ka-poor). Lõpptähed peegeldavad grammatilisi suhteid tekstis, nii näiteks on eesti keeles d-täht mitmuse tunnuseks nii substantiivides kui verbides. Soome sagedaim lõpptäht n esineb käändelõppudes (näit. genitiivis, illatiivis), t on mitmuse tunnus.

Ungari keeles on sagedamad lõpptähed k, t, n, s. Seega domineerivad lõppasendis konsonandid. Läti keeles esineb sõna lõpus kõige sagedamini s (ülejäanud tähtede kohta puuduvad andmed). Vene keeles seisavad lõpptähtede sagedusjärjestuse alguses vokaalid: u, e, o, a. Inglise keeles on viis sagedamat lõpptähte e, s, t, d, n.

Joonisel 4 esitatakse meie koondvalimi põhjal eesti algus- ja lõpptähtede sagedused võrdlevalt. Võrdlus näitab kujukalt, missugust osa täidab üks või teine täht eelistavalt - kas leksikaalset (sõna algul) või grammatilist osa (sõna lõpul). Paistab silma, et niisugused tähed nagu k, p, j, v, r, h, o eelistavad sõnaalgulist positsiooni, kuma aga d, u, a, e, i esinevad rohkem sõna (sõnavormi) lõpus. Tähti s ja t kasutatakse enam-vähem võrdselt nii algus- kui ka lõppasendis.

Algus- ja lõpptähtede sagedusi võib täiendada andmetega tähtede sõnasisese esinemuse kohta. Kui arvestada iga tähe üldsageduseks 100 %, siis jaotuvad näit. tähe a puhul sagedused algus-, sise- ja lõpptähena järgmiselt: 7 % -



Joon. 4. Algus- ja lõpptähed tekstis.

- 67 % - 26 %. Järelikult kasutatakse a-tähte kõige sagedasini sõnasiseses positsioonis. Alljärgnevalt esitame koordinaalimisi andmed kõigi tähtede kohta esinemuse kohta tekstis (arvud tähe järel tähistavad esinemissagedust protsentides vastavalt algus-, sise- ja lõpptähena):

a 7-67-26	h 27-70-3	n 16-70-14	u 2-87-11
b 5-52-43	i 4-76-20	o 33-66-1	v 48-50-2
d 1-55-44	j 52-48-0	p 67-32-1	õ 11-89-0
e 9-65-26	k 44-52-4	r 15-83-2	ä 5-95-0
f 58-42-0	l 10-76-14	s 17-59-24	ö 9-87-4
g 0-96-4	m 36-60-4	t 22-59-19	ü 35-65-0

Tekstisageduste kõrval pakuvad huvi algus- ja lõpptähtede sagedused s õ n a s t i k u s. Et saaks paremini võrrelda teksti- ja sõnastikusagedusi, võtame aluseks üheainsa allkeele, nimelt ilukirjandusproosa autorikõne andmed

kõllaltki suure valimi põhjal (kokku 15.000 sõnet). Materjal on võetud võrdsetes osades eri autorite teostest. Viis sagedamat algustähte on k, t, p, g, v/j nii tekstis kui sõnastikus, kusjuures tekstis on esimesed kaks alati k ja t, sõnastikus k ja p. Autorikõne sõnavormide sõnastikus on algustähtede sagedused järgmised: k (16,6 %), p (10,1 %), g (9,8 %), t (p,5 %), v (8,5 %). Vokaalid moodustavad sõnastiku algustähtede kogusesinemusest 16,0 % (tekstist 20,0 %). Algustähtede sagedused tekstis ja sõnastikus on kokkuvõttes tugevas korrelatsioonis (seda on täheldatud ka teiste keelte puhul, vt. Karlgren, 1962). Mõnevõrra suurem erinevus teksti ja sõnavormide sõnastiku vahel ilmneb lõpptähtede osas. Sagedamad lõpptähed autorikõne sõnastikus on g, a, g, d, t (tekstis a, a, a, i, d), kusjuures vokaalide osakaal on 46,7 % (tekstis 53,8 %). Kokkuvõtlikud andmed teksti- ja sõnastikusageduste kohta koos tekstikoormuse (T/S-suhte) arvutamiseks esitame tabelis 10.

Tähtede korrelatiivne funktsioon. Tähtede funktsionaalset koormust eri asendites väljendatakse keelestatistikas nn. korrelatiivse funktsiooni mõel, mida määratletakse kui "keeleelemendi tingtõenäosuse suhet sõltumatasse tõenäosusse" (Андреев, 1967, 22). Näiteks: p-tähe "sõltumatu tõenäosus" ehk teiste sõnadega suhteline sagedus tekstis üldse on 1,8 %, algustähena aga 7,4%. Viimane väljendab "tingtõenäosust" antud olukorras, s. o. algpositsioonis. Korrelatiivne funktsioon on sel juhul  $7,4 : 1,8 = 4,1$ . See tähendab, et p-täht esineb algustähena umbes neli korda sagedamini kui keskmiselt tekstis üldse, seega on p-tähe funktsionaalne koormus tekstis olulisel määral koondatud sõna algusse. Samas aga on p-tähe sagedus lõpptähena ainult 0,1 % (kõigist lõpptähtedest), s. t. korrelatiivne funktsioon  $KF = 0,1 : 1,8 = 0,06$ . Tähe funktsionaalne koormus sõna lõpul on seega väga väike.

Korrelatiivse funktsiooni väärtused kõigi tähtede kohta alg- ja lõpp-positsioonis esitatakse tabelis 11. Ilmneb, et sõnaalgulises positsioonis on eesti keeles kõige suuremad korrelatiivse funktsiooni väärtused järgmistel tähtedel: p (4,1), j (3,1), v (3,0), k (2,7), m (2,2), ü (2,1), o (2,0). Väikesed KF väärtused on tähtedel ä ja ö (mõlemal 0,3), i (0,2) ja u (0,1). Peale

selle on KF väärtused väikesed tähtedel b, g, d, mis teatavasti võivad eesti keeles esineda sõna algul ainult vöör-sõnades. Seevastu on täht f spetsiaifiline algustäht (KF=3,0).

Soome keeles on suurte KF väärtustega järgmised algustähed: j (4,9), p (3,5), h (3,5), y (2,4), m (2,4), k (2,0). Võrreldes eesti keelaga on sarnasus suur, erineb ainult h-tähe koormus, mis soome keeles on algustähe osas väga suur. Kui võrrelda korrelatiivse funktsiooni väärtusi kahes keeles vokaalide ja konsonantide osas, saame järgmised tulemused: eesti keeles on vokaalide korrelatiivne funktsioon 0,6, konsonantide KF aga 1,4; soome keeles on vastavad väärtused 0,4 ja 1,5. Mõlemas keeles on konsonantide funktsionaalne koormus algustähena oluliselt suurem kui vastav vokaalide koormus.

Võrdluseks toome andmed mõne teise keele kohta (Андреев, 1967). Suuremad KF väärtused on järgmistel algustähtedel: läti keeles - p, n, s; vene keeles - п, с, б; tšehhi keeles - p, y, s; saksa keeles - p, w, b, prantsuse keeles p, c, d, hispaania keeles - p, c, e, itaalia keeles - c, p, g. Kõigis nimetatud keeltes on samuti nagu eesti ja soome keeleski sõnaalgulises positsioonis suure funktsionaalse koormusega labiaal p. Labiaalid ja labiodentaalid on aktiivsed ka inglise keeles: w, b, f. Ungari keeles on suurima KF väärtusega algustähed y (6,0), k (3,1), m (2,0), s (1,7), e (1,4) ja t (0,9).

Sõnalõpulisel positsioonis on eesti keeles eriti koormatud järgmised tähed (vt. tabel 11): d (KF = 2,6), b (2,4), a (1,6), e (1,6), s (1,5), i (1,2) ja t (1,1). Ülejäänud tähtede KF väärtus on alla 1. Soome keeles saame järgmise pingerea: n (2,8), h (2,2), a (2,0), e (1,1) ja i (1,1), t (1,0). Põhiline erinevus seisneb selles, et eesti keeles on aktiivne lõpptäht -b (esineb lõpptähena peamiselt verbi oleviku ainsuse 3. pöördes, näit. tuleb, võtab; soome keeles on vastav vorm vokaallõpuline), soome keeles aga esineb lõppasendis kõige sagedamini -n (näit. genitiivi- ja illatiivivormi lõpus; jalan, jalkaan jne.; mõningates asesõnavormides, näit. kukin, kunkin, mikin, min-kin, kukaan jt.; umbisikulises tegumoes, näit. otetaan; eesti keeles on need juhud vokaallõpulisel positsioonis on eesti kee-

les 1,1 ja soome keeles 1,3; konsonantide KF eesti keeles 0,9 ja soome keeles 0,8.

Kui sõnaalgulises positsioonis võis täheldada küllaltki suurt sarnasust paljude keelte vahel tähtede korrelatiivse funktsiooni seisukohast, siis sõnalõpulisel KF väärtused diferentseerivad keeli oluliselt. Näiteks on suure sõnalõpulisel koormusega vene keeles tähed Н, Б, Я, М, И, inglise keeles у, д, з, prantsuse keeles з, а, т, itaalia keeles о, е, і, ungari keeles к, з, т. Pole kahtlust, et lõpetähete sagedused ja vastavad korrelatiivse funktsiooni väärtused võimaldavad kokkuvõttes küllaltki täpselt diferentseerida ja identifitseerida keeli tekstide automaattöötlemisel.

Tähtede korrelatiivse funktsiooni võib arvutada ka kõigi ülejäänud positsioonide kohta. Tabelis 12 esitame andmed sõna algusest teiselt kohalt asuvate tähtede sageduste ja vastavate KF väärtuste kohta. Tulemused on küllaltki huvitavad puhtlingvistilisest seisukohast. Nimelet selgub, et KF väärtused eristavad selgepiirilisel vokaale konsonantidest. Kui KF on suurem kui 1, siis on tegemist vokaaliga. Kõigil konsonantidel on KF väärtus alla 1, kusjuures sonoorsed р, п ja л seisavad vokaalidele kõige lähemal. Siin tutvustatud katse tulemused on kujukaks näiteks selle kohta, kuidas formaalne kvantitatiivne analüüs suudab esile tuua tähtsad kvalitatiivsed suhted keeleelementide vahel.

T a b e l 1

Tähtede esinemissagedus eestikeelses tekstis  
(5 allkeelt ja koondtekst; valimi üldmaht N = 78900 tähte)

Täht	Iluk. proosa	Kõne- keel	Aja- leht	Tead.- tehn.	Laulu	KOKKU	
						arv	%
a	13,1	13,0	11,4	10,2	12,4	9660	12,2
e	10,0	11,6	10,5	12,7	9,2	8593	10,9
i	9,6	10,0	10,3	9,8	8,7	7757	9,8
s	8,6	8,2	8,9	9,0	8,1	6785	8,6
t	6,8	6,8	7,9	7,4	6,0	5567	7,1
l	6,7	6,5	6,4	6,1	7,0	5124	6,5
u	5,9	5,3	6,0	6,0	6,7	4603	5,8
k	5,2	4,9	5,0	4,8	5,9	3990	5,1
n	4,7	5,2	4,9	4,5	5,8	3888	4,9
m	3,7	4,9	3,7	4,1	4,0	3216	4,1
d	3,8	3,6	4,2	4,2	4,2	3081	3,9
o	3,3	3,4	4,1	4,2	4,1	2913	3,7
r	2,9	2,4	2,7	3,3	2,5	2170	2,8
v	2,3	2,0	2,3	2,3	2,4	1771	2,3
g	1,9	2,0	1,7	2,0	1,7	1507	1,9
h	2,0	1,9	1,4	1,7	2,2	1439	1,8
j	2,1	1,7	1,7	1,4	2,0	1420	1,8
p	2,0	1,6	1,7	1,7	1,8	1402	1,8
ä	1,7	1,7	1,2	1,2	1,6	1181	1,5
õ	1,5	1,1	1,1	1,4	1,2	990	1,3
ü	0,9	1,0	0,8	1,0	1,0	720	0,9
b	0,8	0,8	0,8	0,7	1,1	666	0,8
ö	0,4	0,3	0,4	0,2	0,4	254	0,3
võõr- tähed	0,1	0,1	0,9	0,1	0,0	203	0,2
Kokku (%)	100,0	100,0	100,0	100,0	100,0	-	100,0
Tähti	25420	20914	16208	11358	5000	78900	-
Allk. osa	32%	27%	20%	15%	6%	100%	-

T a b e l 2

Tähtede esinemissagedus ja hajuvushinnangud koondteksti andmete põhjal:  $\bar{p}$  (%) - keskmine suhteline sagedus,  $s$  - standardhälve,  $\varepsilon_{\bar{p}}$  - keskvärtuse piirviga ja  $\delta_{\bar{p}}$  - suhteline viga 95%-lisel usaldusnivool

Täht	$\bar{p}$ %	$s$	$\varepsilon_{\bar{p}}$	Usaldus- piirid (%)	$\delta_{\bar{p}}$ %
a	12,2	1,07	1,3	10,9 ... 13,5	10,7
e	10,9	1,04	1,3	9,6 ... 12,1	11,9
i	9,8	0,39	0,5	9,3 ... 10,3	5,1
s	8,6	0,32	0,4	8,2 ... 9,0	4,7
t	7,1	0,47	0,5	6,6 ... 7,6	7,0
l	6,5	0,23	0,3	6,2 ... 6,8	4,6
u	5,8	0,37	0,5	5,3 ... 6,3	8,6
k	5,1	0,23	0,3	4,8 ... 5,4	5,9
n	4,9	0,37	0,5	4,4 ... 5,4	10,2
m	4,1	0,50	0,6	3,5 ... 4,7	14,6
d	3,9	0,25	0,3	3,6 ... 4,2	7,7
o	3,7	0,39	0,5	3,2 ... 4,2	13,5
r	2,8	0,30	0,4	2,4 ... 3,2	14,3
v	2,3	0,15	0,2	2,1 ... 2,4	8,7
g	1,9	0,12	0,2	1,7 ... 2,1	10,5
h	1,8	0,25	0,3	1,5 ... 2,1	16,7
j	1,8	0,25	0,3	1,5 ... 2,1	16,7
p	1,8	0,16	0,2	1,6 ... 2,0	11,1
ä	1,5	0,24	0,3	1,2 ... 1,8	20,0
õ	1,3	0,18	0,2	1,1 ... 1,5	15,4
ü	0,9	0,08	0,1	0,8 ... 1,0	11,1
b	0,8	0,04	0,1	0,7 ... 0,9	12,5
ö	0,3	0,03	0,1	0,2 ... 0,4	33,3

Tabel 3

Tähtede esinemissagedus ilukirjandusproosa autorikõnes ja tegelaskõnes (tärnikesega on märgitud statistiliselt olulised vahed olulisusnivool  $\alpha < 0,001$ )

Täht	Autorikõne		Tegelaskõne		Vahe
	$P_a$ %	astak	$P_t$ %	astak	$P_a - P_t$
a	12,8	(1)	13,0	(1)	-0,2
e	10,0	(2)	11,6	(2)	-1,6 <sup>##</sup>
i	9,4	(3)	10,0	(3)	-0,6
s	8,8	(4)	8,2	(4)	+0,6
t	6,7	(6)	6,8	(5)	-0,1
l	6,8	(5)	6,5	(6)	+0,3
u	6,3	(7)	5,3	(7)	+1,0 <sup>##</sup>
k	5,2	(8)	4,9	(9,5)	+0,3
n	4,3	(9)	5,2	(8)	-0,9 <sup>##</sup>
m	3,4	(11)	4,9	(9,5)	-1,5 <sup>##</sup>
d	4,1	(10)	3,6	(11)	+0,5
o	3,3	(12)	3,4	(12)	-0,1
r	3,0	(13)	2,4	(13)	-0,6
v	2,3	(14)	2,0	(14,5)	+0,3
g	1,9	(17,5)	2,0	(14,5)	-0,1
h	1,9	(17,5)	1,9	(16)	0
j	2,0	(16)	1,7	(17,5)	+0,3
p	2,2	(15)	1,6	(19)	+0,6 <sup>##</sup>
ä	1,7	(19)	1,7	(17,5)	0
õ	1,5	(20)	1,1	(20)	+0,4
ü	0,9	(21,5)	1,0	(21)	-0,1
b	0,9	(21,5)	0,8	(22)	+0,1
ö	0,5	(23)	0,3	(23)	+0,2
Võõr- tähed	0,1	(24)	0,1	(24)	0
Kokku (%)	100,0	-	100,0	-	-
Tähti	10420	-	20914	-	-

T a b e l 4

Tähtede esinemissagedus eestikeelsetes tekstides: I - "Eesti Postimees" 19.saj., vana kirjaviis (Kaasik, Laugaste, 1969); II - rahvalaul (Kaasik, Laugaste, Ääremaa, 1975), III - segatekst (Hansen, 1961); IV - ajaleht 52%, ilukirjandus 48% (Kaasik, Laugaste, Ääremaa, 1975); V - ilukirjandusproosa 32%, tegelaskõne 27%, ajaleht 20%, teaduslik-tehniline tekst 15%, luule 6%; VI - IV+V: ilukirjandus 56%, ajaleht, teaduslik-tehniline tekst - 44%

Täht	I	II	III	IV	V	VI
a	13,8	13,7	13,0	12,9	12,2	12,6
b	0,9	0,7	0,5	0,9	0,8	0,8
d	4,3	3,8	4,3	4,0	3,9	4,0
e	11,3	12,1	11,0	11,1	10,9	11,0
g	1,7	1,2	1,6	2,2	1,9	2,1
h	2,1	1,8	1,6	1,9	1,8	1,9
i	8,6	9,0	10,0	9,4	9,8	9,6
j	2,3	1,2	1,6	1,9	1,8	1,8
k	5,0	5,5	5,7	4,9	5,1	5,0
l	5,6	7,1	6,2	6,1	6,5	6,3
m	4,3	3,9	4,1	3,8	4,1	3,9
n	4,3	5,1	4,9	4,7	4,9	4,8
o	2,8	2,8	3,1	3,5	3,7	3,6
p	1,7	2,2	1,7	1,9	1,8	1,9
r	2,3	2,6	2,7	2,7	2,8	2,7
s	8,7	8,7	8,7	8,7	8,6	8,7
t	6,9	5,7	7,0	7,1	7,1	7,1
u	6,0	5,4	5,4	6,0	5,8	5,9
v	2,5	2,4	2,4	2,2	2,3	2,2
õ	1,4	1,3	1,4	1,3	1,3	1,3
ä	2,3	2,7	1,6	1,4	1,5	1,4
ö	0,3	0,3	0,4	0,5	0,3	0,4
ü	0,9	1,0	0,9	0,8	0,9	0,8
Võõr- tähed	-	-	0,2	0,1	0,2	0,2
Kokku (%)	99,9	100,2	100,0	100,0	100,0	100,0
Tähti	43250	47700	?	84440	78900	163340

T a b e l 5

Tähtede suhtelised sagedused tekstis ja sõnastikus  
(sõnavormid) eesti ilukirjandusproosa autorikõne  
põhjal

Täht	Suhteline sagedus (%)		Vahe T-S
	Tekstis (T)	Sõnastikus (S)	
a	12,8	12,7	+0,1
e	10,0	9,0	+1,0
i	9,4	8,5	+0,9
s	8,8	10,0	-1,2
l	6,8	6,5	+0,3
t	6,7	7,1	-0,4
u	6,3	7,1	-0,8
k	5,2	5,7	-0,5
n	4,3	4,2	+0,1
d	4,1	4,3	-0,2
m	3,4	3,3	+0,1
o	3,3	2,8	+0,5
r	3,0	3,5	-0,5
v	2,3	2,4	-0,1
p	2,2	2,4	-0,2
j	2,0	1,4	+0,6
g	1,9	1,9	0
h	1,9	1,8	+0,1
ä	1,7	1,4	+0,3
ö	1,5	1,8	-0,3
b	0,9	0,8	+0,1
ü	0,9	0,9	0
õ	0,5	0,4	+0,1
Võõr- tähed	0,1	0,1	0
Kokku	100,0	100,0	-
Tähti	10420	19770	-

T a b e l 6

Eesti keele häälikute sagedused tekstis: I - Saareste, 1932; II - Särgava, 1967 (N = 30.000 häälikut, ajalehed ja ajakirjad); III - Tuldava (N = 23.560 häälikut, 4 allkeelt võrdsetes osades: ilukirjandusproosa autorikõne, tegelaskõne, ajalehetekst, teadusalane tekst)

Häälik	Sagedused protsentides		
	I	II	III
<b>Vokaalid:</b>			
a	14,0	12-14	12,2
e	10,5	11-12	11,0
i	8,5	9	9,5
o		4-5	3,1
u		5	6,0
õ	1,5	1,5	1,3
ä		1,5-2	1,3
ö	0,2	0,2-0,5	0,2
ü	0,8	0,8-1	0,9
<b>Konsonandid:</b>			
p+B		3	2,6 (1,9+0,7)
t+D	10,5	10,5-11	11,9 (7,3+4,6)
k+G		8,9	7,3 (5,3+2,0)
h	1,0	1-1,5	1,7
j	2,0	1,5-2	1,9
l	7,0	7-7,5	6,2
m		3-4	4,0
n+ŋ		4-5	4,6 (4,3+0,3)
r		2-3	2,9
s	10,0	10	9,0
v		2-3	2,4
-	-	100	100,0

T a b e l 7

Häälikute sagedus eestikeelses tekstis (4 allkeelt võrdsetes osades: ilukirjandusproosa autorikõne, tegeelaakõne, ajalehetekst, teadusalane tekst), valimi üldmaht  $N = 23560$  häälikut

Häälik	Suhteline sagedus (%)			Ühe ja kahe tähega kirjutatud häälikute suhe (%)
	Kokku	Ühe tähega kirjutatud	Kahe tähega kirjutatud	
a	12,2	11,7	0,5	96:4
t	11,9	7,3	7,1	97:3
D		4,6	4,6	-
e	11,0	10,1	0,9	92:8
i	9,5	9,0	0,5	94:6
s	9,0	8,8	0,2	98:2
k	7,3	5,3	5,1	96:4
G		2,0	2,0	-
l	6,2	5,7	0,5	91:9
u	6,0	5,5	0,5	93:7
n	4,6	4,3	4,0	93:7
η		0,3	0,3	-
m	4,0	3,9	0,1	98:2
o	3,1	2,5	0,6	80:20
r	2,9	2,8	0,1	97:3
p	2,6	1,9	1,8	96:4
B		0,7	0,7	-
v	2,3	2,3	0,0	100:0
j	1,9	1,9	0,0	100:0
h	1,7	1,7	0,0	100:0
ä	1,3	1,1	0,2	83:17
õ	1,3	1,3	0,0	100:0
ü	0,9	0,7	0,2	85:15
ö	0,2	0,0	0,2	0:100
Võõr- hää- likud	0,1	-	-	-
Kokku	100,0	94,6	5,3	95:5

T a b e l 8

## Algustähtede sagedused tekstis

Algus- täht	Autori- kõne	Tegelas- kõne	Tead. tekst	Aja- leht	Koondtekst	
					arv	%
a	4,3	6,7	5,0	4,9	218	5,3
b	0,4	0,1	0,1	0,2	8	0,2
c	0,0	0,0	0,0	0,0	0	0,0
d	0,0	0,2	1,1	0,3	15	0,4
e	5,4	7,2	6,9	5,9	262	6,4
f	0,1	0,1	0,6	0,5	11	0,3
g	0,0	0,0	0,0	0,0	0	0,0
h	4,3	2,4	2,2	1,5	110	2,7
i	1,4	2,3	2,6	2,9	91	2,2
j	6,5	5,3	4,2	6,3	229	5,6
k	13,7	11,7	15,1	17,3	580	14,1
l	4,7	3,3	3,9	3,1	154	3,8
m	6,7	12,5	7,3	5,6	342	8,3
n	5,6	4,1	4,4	4,2	189	4,6
o	4,9	7,4	7,6	8,6	287	7,0
p	8,5	6,7	6,7	7,9	305	7,4
r	2,3	2,3	3,7	2,3	106	2,6
s	8,2	10,3	8,1	8,2	361	8,8
t	10,0	9,7	8,6	9,6	390	9,5
u	0,9	0,3	0,8	0,7	27	0,7
v	8,1	4,7	7,3	6,2	267	6,5
õ	0,8	0,6	1,2	0,9	35	0,9
ä	0,7	0,8	0,2	0,0	20	0,5
ö	0,3	0,1	0,0	0,2	6	0,1
ü	2,2	1,2	2,4	2,7	84	2,1
Kokku (%)	100,0	100,0	100,0	100,0	-	100,0
Algus- tähti	1106	1230	900	861	4097	-

Tabel 9

## Lõpptähtede sagedused tekstis

Lõpptäht	Autorikõne	Tegelas-kõne	Tead.tekst	Aja-leht	Koondtekst	
					arv	%
a	18,2	23,8	14,6	18,1	805	19,1
b	1,6	1,7	2,1	1,4	72	1,7
d	12,6	9,6	12,1	12,7	489	11,6
e	14,2	15,5	23,9	18,4	745	17,7
g	0,6	0,2	0,6	0,5	19	0,4
h	0,1	0,9	0,0	0,2	14	0,3
i	12,1	12,1	10,3	11,5	488	11,6
k	1,3	0,9	1,3	0,9	46	1,1
l	5,4	4,9	5,5	6,1	227	5,4
m	1,2	1,1	0,7	0,5	38	0,9
n	4,1	5,9	3,2	2,2	171	4,0
o	0,1	0,3	0,1	0,5	10	0,2
p	0,0	0,3	0,0	0,0	4	0,1
r	0,5	0,0	0,3	0,7	15	0,4
s	14,5	12,7	11,0	13,8	550	13,0
t	8,5	6,2	9,5	8,9	342	8,1
u	4,6	3,4	4,5	3,4	168	4,0
v	0,2	0,3	0,3	0,2	11	0,3
ä	0,0	0,1	0,0	0,0	1	0,0
ö	0,2	0,1	0,0	0,0	3	0,1
Kokku (%)	100,0	100,0	100,0	100,0	-	100,0
Tähti	1115	1276	953	874	4218	-

Tabel 10

Algus- ja lõpptähtede sagedused ilukirjandusproosa autorikõne tekstis ja sõnavormide sõnastikus koondvalimi põhjal

Täht	Algustähena			Lõpptähena		
	teks- tis (T)	sõnasti- kus (S)	T/S	teks- tis (T)	sõnasti- kus (S)	T/S
a	4,2	4,2	1,0	22,4	16,1	1,4
b	0,7	0,2	3,5	0,8	1,1	0,7
d	0,2	0,2	1,0	11,0	13,8	0,8
e	4,8	3,0	1,6	13,2	15,6	0,9
g	0,1	0,1	1,0	0,9	0,4	2,3
h	3,3	4,6	0,7	0,1	0,1	1,0
i	1,9	1,8	1,2	12,8	9,8	1,3
j	7,5	3,5	2,1	-	-	-
k	14,7	16,6	0,9	0,9	1,2	0,8
l	4,7	6,7	0,7	5,1	5,3	0,9
m	5,7	6,1	0,9	0,7	0,8	0,9
n	5,2	4,2	1,2	0,9	0,5	1,8
o	5,1	2,0	2,6	0,2	0,2	1,0
p	8,7	10,1	0,9	0,1	0,1	1,0
r	2,8	3,8	0,7	0,8	0,7	1,1
s	8,9	9,8	0,9	15,8	17,4	0,9
t	10,2	9,5	1,1	9,5	11,3	0,8
u	0,8	1,1	0,7	4,4	4,9	0,9
v	7,2	8,5	0,9	0,3	0,6	0,5
õ	0,9	1,4	0,6	0,0	0,0	0,0
ä	0,5	0,6	0,8	0,0	0,0	0,0
ö	0,2	0,2	1,0	0,1	0,1	1,0
ü	1,6	1,7	0,9	0,0	0,0	0,0
Võõr- tähed	0,1	0,1	1,0	0,0	0,0	0,0
Kokku (%)	100,0	100,0	-	100,0	100,0	-
Sõnasid, sõnavor- me	14981	8192	-	14981	8192	-

Tabel 11

Algus- ja lõpptähtede suhtelised sagedused tekstis ja korrelatiivne funktsioon (KF) 4 allkeele koondteksti põhjal

Täht	Tekstis uldse (T)	Algus- tähe (A)	KF = A/T	Lõpp- tähe (L)	KF = L/T
a	12,0	5,3	0,44	19,1	1,59
b	0,7	0,2	0,29	1,7	2,43
d	4,4	0,4	0,09	11,6	2,64
e	11,3	6,4	0,57	17,7	1,56
f	0,1	0,3	3,00	0,0	-
g	1,9	0,0	-	0,4	0,21
h	1,7	2,7	1,59	0,3	0,18
i	9,5	2,2	0,23	11,6	1,22
j	1,8	5,6	3,11	-	-
k	5,3	14,1	2,66	1,1	0,21
l	6,4	3,8	0,59	5,4	0,84
m	3,8	8,3	2,18	0,9	0,24
n	4,7	4,6	0,96	4,0	0,85
o	3,5	7,0	2,00	0,2	0,06
p	1,8	7,4	4,11	0,1	0,06
r	2,8	2,6	0,93	0,4	0,14
s	8,8	8,8	1,00	13,0	1,49
t	7,1	9,5	1,34	8,1	1,14
u	6,1	0,7	0,11	4,0	0,65
v	2,2	6,5	2,95	0,3	0,14
õ	1,3	0,9	0,69	0,0	-
ä	1,5	0,5	0,33	0,0	-
ö	0,3	0,1	0,33	0,1	0,33
ü	1,0	2,1	2,10	0,1	-
Kokku (%)	100,0	100,0	-	100,0	-
Tähti	24796	4097	-	4097	-

Algustähele järgneva tähe suhteline sagedus tekstis  
ja korrelatiivne funktsioon (KF) 4 allkeele koond-  
valimi põhjal

Täht	Tekstis üldse (T)	Teisel kohal ( $T_2$ )	KF = $T_2/T$
õ	1,3	6,2	4,8
ä	1,5	6,0	4,0
ü	1,0	2,3	2,3
ö	0,3	0,6	2,0
o	3,5	6,4	1,8
a	12,0	18,8	1,6
u	6,1	9,6	1,6
e	11,3	15,4	1,4
i	9,5	11,9	1,3
r	2,8	2,3	0,9
n	4,7	4,1	0,9
l	6,4	5,2	0,8
g	1,9	1,4	0,7
h	1,7	1,0	0,6
t	7,1	2,2	0,3
b	0,7	0,2	0,3
s	8,8	2,5	0,3
j	1,8	0,5	0,3
k	5,3	1,4	0,3
p	1,8	0,4	0,2
m	3,8	0,8	0,2
d	4,4	0,4	0,1
v	2,2	0,1	0,05
Kokku (%)	100,0	100,0	-
Tähti	24796	4097	-

## VIITEKIRJANDUS

- Ariste, P. Eesti keele foneetika. 3. trükk. - Tartu: TRU, 1968.
- Arndorfer, H. Das Phänomen der Lautbedeutbarkeit aus biophonetischer Sicht. - Proceedings of the Sixth International Congress of Phonetic Sciences, Prague, 1967, 125-127.
- Bourdon, B. L'expression des émotions et des tendances dans le langage. Paris, 1892.
- Ertel, S. Psychophonetik. Untersuchungen über Lautsybollok und Motivation. - Göttingen: Hogrefe, 1969.
- Guiraud, P. Les caractères statistiques du vocabulaire. Essai de méthodologie. Paris, 1954.
- Hakulinen, L. Suomen kielen rakenne ja kehitys. Helsinki, 1941.
- Hansen, H. Lõdumismasinamatriitside ratsionaalsem komplekterimine. - Polügrafist, 1961, nr. 2, 21-24.
- Herdan, G. The Calculus of Linguistic Observations. - The Hague: Mouton, 1962.
- Herdan, G. A Method for the Quantitative Analysis of Language Mixture. - Statistical Methods in Linguistics, 2. - Stockholm: Skriptor, 1963, 110-123.
- Kaasik, Ü., Laugaste, E. Tähtede sagedus eestikeelsetes tekstides. - Keel ja Kirjandus, 1969, nr. 10, 600-605.
- Kaasik, Ü., Laugaste, E., Ääremaa, K. Tähtede ja silpide sagedus eestikeelsetes tekstides. - Keel ja Kirjandus, 1975, nr. 1, 21-29.
- Kaasik, Ü., Tuldava, J., Villup, A., Ääremaa, K. Eesti ilukirjandusproosa autorikõne sõnavormide sagedussõnastik. - Tõid keelestatistika alalt, I. TRU Toimetised, vihik 377. Tartu, 1976, 107-153.
- Kaasik, Ü., Tuldava, J., Villup, A., Ääremaa, K. Eesti tänapäeva ilukirjandusproosa autorikõne lekseemide sagedussõnastik. - Tõid keelestatistika alalt, II. TRU Toimetised, vihik 413. Tartu, 1977, 5-140.
- Kaasik, Ü., Tuldava, J. Sõnalõpu ja sõnapikkuse vahakorrast eestikeelses tekstis. - Eesti keele sõnavarastatistika erikõsimusi. Tõid keelestatistika alalt, IV. TRU Toimetised, vihik 477. Tartu, 1980, 154-167.
- Kaasik, Ü., Tuldava, J., Ääremaa, K. Eesti keele sõnavormide pöördõsagedussõnastik. - Eesti keele sõnavarastatistika erikõsimusi. Tõid keelestatistika alalt, IV. TRU Toimetised, vihik 477. Tartu, 1980, 5-153.
- Kálmán, B. Häälikustatistilisi vaatlusi mansi keele alalt. - Nonaginta. J.V. Veski 90. sünnipäevaks 27. juunil 1963. Emakeele Seltsi Toimetised nr. 6. Tallinn, 1963, 123-127.
- Karlgren, H. Statistical methods in phonetics. - Manual of Phonetics. Edited by B. Malmberg. - The Hague: North-Holland Publishing Co., 1968, 129-154.
- Kask, A. Eesti keele ajalooline grammatika, I. Häälikulugu. 2. trükk. - Tartu: TRU, 1972.

- Krámský, J. Fonologické využití samohláskových foném. - *Linguistica Slovaca*, IV - VI. Bratislava, 1948.
- Krámský, J. A Quantitative Typology of Languages. - *Language and Speech*. Vol. 2, 1959.
- Kučera, H., Monroe, G.K. A Comparative Phonology of Russian, Czech and German. New York, 1968.
- Lepmaa, A. Tõenäosusteooria ja matemaatiline statistika. Tallinn, 1968.
- Mackay, A. On the Type-Found of the Phaistos Disc. - *Statistical Methods in Linguistics*, 4.-Stockholm; Skriptor, 1965, 15-25.
- Maneca, C. Considérations statistiques sur les finales vocalliques en roumain. - *Proceedings of the Sixth International Congress of Phonetic Sciences*. Prague, 1967, 609-611.
- Marchand, H. Phonetic symbolism in English Word-Formation. - *Indogermanische Forschungen*. Bd. 64, H.2. Berlin, 1959.
- Martinet, A. Economie des changements phonétiques. Berne, 1955.
- Meier, H. Deutsche Sprachstatistik, I - II. Hildesheim, 1964.
- Moreau, R. Quantité d'information et redondance aux différents niveaux du français écrit. Paris, 1963.
- Mustonen, S. Multiple Discriminant Analysis in Linguistic Problems. - *Statistical Methods in Linguistics*, 4. - Stockholm; Skriptor, 1965, 37-44.
- Mäger, M. Lauleteksti foneetilise struktuuri ja semantika vahekorrast. - *Keel ja Kirjandus*, 1971, nr. 6, 325-334.
- Pierce, J.R. Symbols, Signals and Noise: The Nature and Process of Communication. - London; Hutchinson, 1962.
- Saareste, A. Die estnische Sprache. Tartu, 1932.
- Setälä, V. Suomen kielen dynamiikkaa. - Helsinki: Suomalaisen Kirjallisuuden Seura, 1972.
- Sigurd, B. A Note on the Number of Phonemes. - *Statistical Methods in Linguistics*, 2.-Stockholm; Skriptor, 1963, 94-99.
- Särgava, V. Eestikeelse kõne kasutamine kuulmise uurimisel. - *Keel ja Kirjandus*, 1967, nr. 1, 27-30.
- Trubetzkoy, N.S. Grundzüge der Phonologie. - *Travaux du Cercle Linguistique de Prague*, VII. Praha, 1939.
- Tuldava, J. Statistiline väljavõttemeetod keeleteaduses. - *Linguistica*, I.-Tartu; TRÜ, 1969, 5-49.
- Tuldava, J. Informatsiooniteooria ja keeleteadus. - *Keel ja Kirjandus*, 1970a, nr. 6, 329-339.
- Tuldava, J. Statistilised testid keeleteaduses. - *Linguistica*, II.-Tartu; TRÜ, 1970b, 125-196.

- Tuldava, J. Sagedussõnastik leksikostatistilise uurimise objektina. - Tõid keelestatistika alalt, II. TRÜ Toimetised, vihik 413. Tartu, 1977, 141-171.
- Tuldava, J. Sõnavormide esinemus eestikeelses tekstis. - Eesti keele sõnavarastatistika probleeme. Tõid keelestatistika alalt, III. TRÜ Toimetised, vihik 446. Tartu, 1978, 107-126.
- Verglas, A. Remarques sur la relation entre rang et fréquence des lettres dans les mots français. - Bulletin d'information du laboratoire d'analyse lexicologique. Vol. 6. Besançon, 1963.
- Weiss, M. Über die relative Häufigkeit der Phoneme des Schwedischen. - Statistical Methods in Linguistics, I.-Stockholm; Skriptor, 1961, 41-55.
- Zipf, G.K. The Psycho-Biology of Language. An Introduction to Dynamic Philology. - Boston: Houghton Mifflin, 1935.
- Zsilka, T. Rekurentnost foném v počzii. - Metryka słowianska. - Wrocław: Ossolineum, 1971.
- Андреев Н. Д. Статистико-комбинаторные методы в теоретическом и прикладном языковедении. - Л.: Наука, 1967.
- Вектаев К. В., Лобин В. Г. Статистические приемы описания причин фонетических изменений в языке. - Известия Академии наук КавССР, серия общественная, 1969, № 1, с. 81-87.
- Бедоногов Г.Г. Определение грамматических признаков "новых" слов с помощью словаря. - В кн.: Иллюстрированная лингвистика. Л., 1971, с. 225-229. (Учен. зап. Ленинградского гос. ин-та им. А. М. Герцена. Том 458, ч. 2).
- Ведозол М. Фонемная структура слова в современном английском и латинском языках. - Proceedings of the Sixth International Congress of Phonetic Sciences. Prague, 1967, p. 985-988.
- Ведозол М. Количественное соотношение использования согласных и гласных фонем в слове в современном латинском и английском языках. - Балтийские языки и их взаимосвязи со славянскими, финно-угорскими и германскими языками. Тезисы докладов научной конференции, посвященной 100-летию со дня рождения акад. Я. Эндзелина. Рига, 1973, с. 83-86.
- Вийтсо Т.-Р. Проблемы количества в эстонском языке. - Советское финно-угроведение XV, 1979, № 1, с. 1-17.
- Гурдишева Е. А. Звуковой символем и факторы, влияющие на него. - Учен. зап. МПНИИ им. М. Тореза, 1973, т. 72.
- Курбаев А. П. Фонетическое значение. - Л.: Изд-во ЛГУ, 1974.
- Зиндер Л. Р., Строева Т. В. Статистика гласных немецкого языка в диакронии. - В кн.: Исследования по немецкой филологии. Тула, 1972, с. 5-20.
- Исенгалядина А. А. Некоторые вопросы фонологической статистики. - Вопросы языковедения, 1972, № 5, с. 119-129.
- Курбаков К. И. Кодирование и поиск информации в автоматическом словаре. - М.: Советское радио, 1968.

- Лёвицкий В. В. К проблеме звуко-символизма. - В кн.: Психологические и психолингвистические проблемы владения и овладения языком. М., 1969, с. 123-132.
- Меликишвили И. Г. Восприятие фонем и некоторые универсальные закономерности частотных соотношений признаков фонем. - Материалы 3-го Всесоюзного симпозиума по психолингвистике. М., 1970, с. 162-163.
- Милевский Т. Предпосылки типологического языкознания. - В кн.: Исследования по структурной типологии. - М.: Изд-во АН СССР, 1963, с. 3-31.
- Никонов В. А. Интерпретация фонетических частот. - Ученые записки Ин-та славяноведения, 1963, т. 17, с. 259-270.
- Перебийнис В. И. Использование статистических методов в типологических исследованиях (на материале славянских и германских языков). - В кн.: Проблемы языкознания. М. 1967.
- Петрова Н. В. Кодовые характеристики письменного текста. - В кн.: Статистика речи. / Отв. ред. Р. Г. Пиотровский. - Л.: Наука, 1968, с. 5-49.
- Пиотровский Р. Г. Информационные измерения языка. - Л.: Наука, 1968.
- Савченко И. Ф. Определение части речи словоформы украинского языка по конечной фонеме. - В кн.: Вопросы лингвостатистики и автоматизации лингвистических работ. М., 1970, с. 60-68. (Труды ЦНИИЛИ. Серия 3/70. Выпуск 3).
- Свечевичус Б. И. К вопросу о частоте встречаемости фонем в литовской письменной речи. - Материалы коллоквиума лаборатории экспериментальной фонетики и психологии речи Ч. 2. Вильнюс, 1966, 19-22.
- Статистико-комбинаторное моделирование языков. / Под ред. Н. Д. Андреева. - М.; Л.: Наука, 1965.
- Статистичні параметри стилів. / Ред. В. С. Перебийніс. - Київ: Наукова думка, 1967.
- Сухотин Б. В. Экспериментальное выделение классов букв с помощью электронной вычислительной машины. - В кн.: Проблемы структурной лингвистики. М., 1962, с. 198-206.
- Тисенко Э. В. Алфавитно-графическая структура словаря. - Научный симпозиум "Семиотические проблемы языков науки, терминологии и информатики". Ч. 1. - М.: Изд-во МГУ, 1971, с. 251-253.
- Тулдава Ю. А. Квантитативное исследование структуры односложного слова в эстонском языке. - В кн.: Проблемы общей и прикладной лингвистики. Linguistica, X. Тарту, 1978, с. 115-135. (Учен. зап. ТГУ, вып. 453).
- Фрумкина Р. М., Василевич А. П., Герганов Е. Н. Субъективные оценки частот элементов текста как прогнозирующий фактор. - В кн.: Вероятностное прогнозирование в речи. М. 1971, с. 70-93.
- Хольм Х. А. Выделение первого морфологического типа в эстонском языке на основе статистико-комбинаторного моделирования в подъявике радиоэлектроники. - В кн.: Статистико-комбинаторное моделирование языков. - М.; Л.: Наука, 1965а, с. 212-218.
- Хольм Х. А. Выделение первого морфологического типа в эстонском языке на материале публицистических текстов. - В кн.: Статистико-комбинаторное моделирование языков. - М.; Л.: Наука, 1965б, 219-224.

## ФОНЕТИКО-ГРАФЕМНЫЕ ИЗМЕРЕНИЯ ЭСТОНСКОЙ ЛЕКСИКИ

Юхан Тулдава

Р е з ю м е

В статье приводятся данные о частотности букв и звуков в эстонском тексте и словаре, а также данные о частотности начальных и конечных букв (звуков) в тексте. Констатируется наличие некоторых различий в распределении частот букв в текстах из разных подязыков (см. табл. I, стр. 83). На основе суммарной выборки объемом в 163 340 букв (56% художеств. и 44% нехудожеств. текстов) пять самых частых букв (а, е, i, v, t) покрывают 49% и десять самых частых букв - 75% текста (табл. 4, колонка VI, стр. 86). Связь между частотой и рангом букв хорошо описывается логарифмической функцией:  $p_r = a + b \lg r$  ( $p_r$  - частота,  $r$  - ранг,  $a$  и  $b$  - константы; для эстонского языка  $a = 13,7$  и  $b = 9,6$ ; см. рис. 2 на стр. 63). Устанавливается близость частотности букв в тексте и в соответствующем словаре (табл. 5 на стр. 87 и рис. 3 на стр. 64). Частотность звуков в тексте определяется на основе выборки из четырех подязыков общим объемом в 23 560 звуков (соотв. 24 796 букв). Пять самых частых звуков (а, t/D, е, i, v) покрывают 53,6% текста (табл. 6, колонка III, стр. 88). Соотношение гласных и согласных звуков в тексте - 45,5:54,5% (т.е. на 100 гласных приходится в среднем 120 согласных). Передние гласные составляют 50,3%, задние (и средние) - 49,7% в тексте. Среди согласных 40% смычных, 44,2% шелевых и 15,8% носовых звуков. Частотность начальных букв (табл. 8, стр. 90) и конечных букв (табл. 9, стр. 91) определяются на основе выборки из текстов четырех подязыков. Устанавливается функциональная нагрузка букв в начальной и конечной позиции слова (табл. 10, стр. 92 и рис. 4, стр. 79). Статистические данные эстонского языка сопоставляются с соответствующими данными из других языков, в частности из родственных финского и венгерского языков.

# PHONETIC-GAPHEMIC MEASURES OF ESTONIAN LEXICS

Juhan Tuldava

## S u m m a r y

The article presents statistical data on the frequency of letters (graphemes) and sounds in Estonian text and vocabulary as well as on the frequency of initial and final letters (sounds) of the words in a text. Some differences in the distribution of letters in the texts of various sublanguages are revealed (see Table 1, p. 83). On the basis of a summary sample of 163,340 letters (56 % fiction and 44 % non-fiction material) the five most frequent letters (a, e, i, s, t) cover 49 % and the ten most frequent letters - 75 % of the text (Table 4, Column VI, p. 86). The relation between the frequencies and the ranks of the letters can be expressed by the logarithmic function  $p_r = a + b \lg r$  ( $p_r$  - frequency,  $r$  - rank,  $a$  and  $b$  - constants; for the Estonian text  $a = 13.7$  and  $b = 9.6$ , see Fig. 2 on p. 63). Proximity between the frequencies of letters in a text and the corresponding vocabulary can be stated (Table 5 on p. 87 and Fig. 3 on p. 64). The frequency of the sounds in Estonian texts is determined on the basis of a summary sample from the texts of four sublanguages - the volume of the sample was 23,560 sounds (which corresponds to 24,796 letters). The five most frequent sounds (a, t/D, e, i, s) cover 53.6 % of the text (see Table 6, Column III, p. 88). The ratio between vowels and consonants is 45.5/54.5 % (i.e. 100 vowels to 120 consonants). The front vowels make up 50.3 %, the back (and medial) vowels - 49.7 % in the text. Of the consonants 40 % are occlusives, 44.2 % spirants, and 15.8 % nasals. The frequency of initial letters (Table 8 on p. 90) and final letters (Table 9 on p. 91) is determined on the basis of four sublanguages. The functional load of the letters in initial and final positions is calculated (Table 10 on p. 92 and Fig. 4 on p. 79). The statistical data on the Estonian language are compared with the corresponding data from other languages, in particular from the cognate Finnish and Hungarian languages.

## SÕNALÕPU GRAFEMOTAKTIKAST EESTIKEELSES TEKSTIS

Ülo Kaasik, Juhan Tuldava

Grafemotaktika uurib tähtede kombinatsioone ja nende põhjal saadud struktuure, mis iseloomustavad sõna sisemist ehitust sõnastikus või tekstis. Sõna sisemist ehitust häälikute või foneemide tasandil vaatleb vastavalt fonotaktika. Grafemotaktiline vaatlus on vajalik neil juhtudel, kus otseseks uurimisobjektiks on trükitud tekst, näit. teksti automaatanalüüsi küsimuste lahendamisel.

Eriolist tähtsust teksti leksikaal-grammatilisel uurimisel tuleb omistada sõnalõpu tähestikulisele struktuurile. Sõna lõpus esinevate tähtede kombinatsioonid väljendavad liiteid, tunnuseid ja muutelõppe, mis aitavad ära tunda sõnaklasse (sõnaliike) ja nende süntaktilisi seoseid teiste tekstis esinevate sõnadega. Sõnalõpu grafemotaktika andmeid koos sõnatüvede nimistuga ja reeglitega sõnalõpu ning sõnatüve omavahelise "sobivuse" kohta kasutatakse käesoleval ajal sõnade morfoloogilise analüüsi ja sünteesi läbiviimisel teksti automaattöötlemise raames (vt. näit. БЕЛОНОГОВ, НОВОСЕЛОВ, 1979, lk. 121 jj.). Seejuures etendab olulist osa sõnalõpu tähe kombinatsioonide statistiline esinemus tekstis, mille alusel on võimalik eristada tsentraalset perifeersest ja sel teel ökonoomsemalt lahendada automaatanalüüsi probleeme (eriti sünteetiliste keelte puhul). Statistilised andmed sõnalõpu tähe kombinatsioonide esinemuse kohta võivad olla abiks ka keeletüpoloogilises uurimistöös.

Käesolevas töös esitame ülevaatlikud tabelid sõnalõpu grafemotaktika kohta eestikeelses tekstis nelja lõpptähe ulatuses. Tabelite põhjal on võimalik kindlaks teha üksiktähtede ning kahe-, kolme- ja neljätäheliste ühendite esinemissagedust eestikeelse sõna lõpus.<sup>+</sup> Andmed on saadud eesti tänapäeva ilukirjandusteksti (autorikõne) töötlemise tulemusena arvuti abil, kusjuures vaadeldav tekst langeb kokku eesti keele sõnavormide pöörd-sagedussõnastiku koostamisel

<sup>+</sup> Lõpptähtede sagedusloendit vt. ka Kaasik, Tuldava, 1980. Sõnalõpu grafemotaktilisi omadusi eestikeelse tehnikaalase ja ajaleheteksti põhjal on vaadelnud H. Holm (Хольм, 1965).

kasutatud tekstiga (vt. Kaasik, Tuldava, Ääremaa, 1980). Sel-line kokkulangevus võimaldab edaspidises uurimistöös kerge-mini kindlaks määrata seosed sõnaalguse (resp. sõnatüve) ja sõnalõpu vahel. Tabelid võivad pakkuda huvi ka mitmesuguste muude lingvistiliste küsimuste lahendamisel.

Käesoleva tekstivalimi maht on ligikaudu 95 000 sõnet (27 000 eri sõnavormi). Vaatleme algul lühidalt lõpp-tähe-paare. Kõige rohkem esineb eestikeelses t e k s t i s st-lõpulisel sõnavorme (antud valimis 3886 juhul, s. o. 3,6 % kõigist sõnalõppudest tekstis); siia kuuluvad sagedamate sõnavormidena pärast, sest, just, vist, ennast, misugust, esimest, inimest, millest, neist, temast, käest, minust, endast jt. Sagedusnimistus järgnevad lõpp-tähepaarid le (3412), ud (3294), id (2893), is, as, ga, ma, da, se, ks, lt jt., kusjuures näit. le-lõpulistest on sagedamad sõnavormid sel-le, pole, üle, talle, poole, jälle, mulle, mille, peale, kel-le, neile, sellele; ud-lõpulistest polnud, olnud, saanud, jäänud, tulnud, läinud, teinud, teadnud, muud, kuud, antud, tehtud; id-lõpulistest - kuid, olid, neid, vaid, meid, tulid, läksid, oleksid, hakkasid, jäid, poisid, aastaid jt. (eraldil arvestame kahetähelisi sõnavorme, vt. allpool.) Võiks arvata, et sagedamad lõpp-paarid prognoosivad hästi mõningaid muute- või tuletusmorfeeme, näit. le, ga, lt, st, is, kuid tegelikult see nii pole. Näiteks ga-lõpulistest sõnavormidest on vaid umbes 50 % kaasaütleva käände vormid. Järelikult tuleb sõna morfoloogiliseks identifitseerimiseks kaasa haarata suuremal hulgal lõpptähti.

Vaatleme näiteks kolmetäheliste ühendite esinemust sõnalõpulisel asendis tekstis. Kaksikümmend sagedamat täheühendit on järgmised (arvud tähistavad esinemissagedust tekstis meie valimi andmeil):

nud	2213	ega	962	ele	763	alt	693
est	1454	sid	962	ist	748	tud	678
oli	1142	ast	952	ust	713	see	674
kui	1059	eks	790	eda	700	vad	668
aga	1039	lle	768	ale	695	aks	658

Antud loendist on välja jäetud kahetähelised ühendid, mis kolme lõpptähe ulatuses osutusid omaette sõnadeks (sõnavormideks), näit. -ja 3266 (kriips tähistab eelnevat sõnavahet), -ta 1722, -ei 1306, -on 1272, -et 1218.

Nelja lõpptähe kombinatsioonidest on eestikeelses tekstis sagedamad

anud	549	ning	429	kuid	324	rast	290
nagu	523	seda	374	dagi	321	avad	286
sest	477	selt	367	unud	302	eile	280
siis	474	leks	343	atud	302	veel	280
lnud	443	inud	331	tema	297	asid	278

Neljatäheliste lõppude ulatuses osutusid omaette kolmetähelisteks sõnavormideks: -oli 1103, -kui 993, -see 674, -aga 671, -oma 553 jt. Näeme, et kolmetäheliste lõppude loendi alusel oleks võinud kindlalt omaette sõnavormiks lugeda vaid täheühendi see (sagedus 674 jäi muutumatuks järgneval analüüsil neljatäheliste lõppude tasandil). Teistel juhtudel aga võisid kolmetähelised ühendid olla nii eri sõnavormid kui ka sõnalõpulisel komponendid, näit. aga (sagedus 1039 kolmetäheliste lõppude nimistus) tähistas eri sõnavormi vaid 671 juhul, ülejäänud juhtudel oli tegemist sõnalõpulisel komponendiga. Tabelist näeme, et kõige rohkem esines taga 78, seejärel maga 58, jaga 38, naga 37, vaga 26, laga 21 jt. Sõnavormide pöördsagedussõnastik aitab kindlaks teha, et näit. taga on omaette sõnavorm 76 juhul, ülejäänud kahel juhul aga esineb sõnavormides seljataga ja aastaga; jaga võib esineda nii omaette sõnavormina (verbina) kui ka sõnalõpulisel komponendina sõnavormides ajaga, majaga, juhatajaga, kuhjaga, naljaga, asjaga, saatjaga, sõjaga, müüjaga jne., s. t. kaasatlevas käändes oleva nimisõna lõpposana.

Esialgne analüüs näitab, et eestikeelses tekstis esinevate sõnavormide morfoloogiline identifitseerimine (sõnaliigi ja osalt ka muutevormi kindlaksmääramine) nelja lõpptähe alusel on võimalik vaid piiratud ulatuses. Viienda tähe lisamine (sõnalõpust lugedes) parandab aga järsult automaatanalüüsi võimalusi.<sup>+</sup> Antud juhul on võimalik sellist analüüsi teostada sõnavormide pöördsagedussõnastiku kaasabil, arvestades ka sõnapikkuse faktorit (vt. Kaasik, Tuldava, 1980).

<sup>+</sup> Võrdluseks võib tuua fakti, et samalaadse ülesande lahendamiseks lätikeelse teksti põhjal piisab tavaliselt sõnalõpu kanetähelisest ühendist, kusjuures vaid moningatel juhtudel on vaja kaasa haarata kolme- või neljatähelisi ühendeid (vt. Дризул, 1977).

Järgnevatel tabelites on ülemises reas antud neljas täht sõnalõpust lugedes ja iga rea algul on toodud sõnalõpu kolmetäheline ühend. Kriips tähistab eelnevat sõnavahet. Võtame näit. ühendi maa; tabeli ülemises reas esimeses veerus on kriips (-), s. t. sõnavahe; sel juhul saame omaette sõnavormi maa, mille sageduse leiame vaadeldava veeru ja rea ristumiskohalt (antud juhul on sagedus 58); teise veeru kohal on täht a, mis koos vaadeldava kolmetähelise lõpuga annab ühendi amaa (sagedus 11) jne.

#### VIITEKIRJANDUS

Kaasik, Ü., Tuldava, J. Sõnalõpu ja sõnapikkuse vahekorrad eestikeelses tekstis. - Eesti keele sõnavara statistika eriküsimusi. Tõid keelestatistika alalt, IV. TRÜ Toimetised, vihik 477. Tartu, 1980, lk. 154-167.

Kaasik, J., Tuldava, J., Ääremäe, K. Eesti keele sõnavormide põrdsagedussõnastik. - Eesti keele sõnavara statistika eriküsimusi. Tõid keelestatistika alalt, IV. TRÜ Toimetised, vihik 477. Tartu, 1980, lk. 5-153.

Белоногов Г.Г., Новоселов А.П. Автоматизация процессов накопления, поиска и обобщения информации. - М.: Наука, 1979.

Дрикул В.А. Алгоритм приближенного морфологического анализа текстов латышского языка. - В кн.: Автоматизация переработки текста методами прикладной лингвистики. Материалы 2-ой Всесоюзной конференции. Кишинев, 1977, с. 107-108.

Хольм Х.А. Выделение парадигмы первого морфологического типа на различных подъязыках при статистико-комбинаторном моделировании эстонской морфологии. - Учен. зап. ТГУ, вып. 172. Tartu, 1965.

Sõnalõpu tähtehendite osagedused eestikeelses tekstis  
(vasakul rea algul - sõnalõpu kolmetäheline ühend, vasaru poolises - neljas täht sõnalõpust loodes)

	A	B	D	E	G	H	I	J	K	L	M	N	O	P	R	S	T	U	V	õ	ä	ö	ü	x	SUM
AAA	0	1	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
JAA	7	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	9
HAA	58	11	0	7	7	0	0	6	0	0	0	0	0	0	0	5	2	1	0	0	0	0	0	0	90
NAA	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
HAA	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
SAA	45	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	45
AAA	0	0	0	0	0	0	0	0	0	2	0	3	0	0	2	5	2	0	16	0	0	0	0	0	30
AAA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	0	0	0	0	0	0	0	0	2
AAA	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	17
AAA	0	1	0	0	13	0	0	0	0	0	0	0	0	0	2	1	0	0	0	0	0	0	0	0	5
AAA	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5
AAA	0	10	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	2	0	0	0	0	13
AAA	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
AAA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
AAA	0	2	0	0	0	0	0	0	0	5	0	0	0	0	0	0	19	0	0	0	0	0	0	0	25
AAA	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
AAA	0	34	7	44	69	7	0	0	17	14	1	1	0	2	10	27	226	0	5	0	0	0	0	0	444
AAA	0	3	1	7	2	18	8	0	1	31	15	5	13	0	1	6	374	222	0	0	0	0	0	0	700
AAA	3	3	5	11	19	6	0	3	0	44	4	144	7	12	9	9	24	6	0	11	1	1	0	0	342
AAA	0	20	0	0	93	0	0	0	0	0	0	0	0	0	0	0	0	20	0	1	0	0	0	0	137
AAA	0	46	0	0	59	0	0	0	0	0	0	0	8	0	0	0	0	20	0	36	0	0	0	0	205
AAA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	5
AAA	0	7	0	0	5	0	0	0	0	0	0	0	73	0	0	0	0	4	0	0	0	0	0	0	89
AAA	0	22	1	14	0	1	1	0	9	37	10	4	4	0	1	2	15	47	12	1	9	0	0	0	190
AAA	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	25	0	0	0	0	0	0	0	0	25
AAA	0	0	0	0	0	0	12	0	0	0	0	1	0	0	0	0	0	0	0	0	19	0	0	0	32
AAA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	126	0	0	126
AAA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	0	0	0	0	0	0	7	0	21
AAA	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
AAA	79	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	79
AAA	7	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8
AAA	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
AAA	4	1	0	0	1	0	0	11	0	0	2	0	0	0	0	2	0	0	0	0	0	0	0	0	81
AAA	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
AAA	43	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	43
AAA	671	2	1	16	17	6	6	3	38	11	21	58	37	3	1	24	13	78	6	26	0	0	1	0	1039
AAA	151	79	1	185	10	5	14	8	2	2	77	22	4	9	1	10	204	148	1	5	2	21	1	0	962
AAA	47	5	10	22	0	7	2	1	0	3	20	9	22	1	8	23	10	15	0	24	0	2	0	0	302
AAA	0	17	0	1	22	0	0	1	0	0	0	0	0	5	0	0	0	0	9	0	3	0	0	0	57
AAA	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	10
AAA	0	0	0	0	3	0	0	0	0	1	0	2	6	1	0	0	2	0	1	0	16	0	0	0	32







































	I	A	B	D	E	G	H	I	J	K	L	M	N	O	P	R	S	T	U	V	W	X	Y	Z	SUM
URI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	43
VRI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10
ARI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12
ORI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14
URI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6
URI	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20
URI	4	21	0	0	3	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
URI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30
URI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16
URI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6
URI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	193
URI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	41
URI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
URI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
URI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14
URI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12
URI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5
URI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	74
URI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	59
URI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	99
URI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13
URI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20
URI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7
URI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
URI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
URI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	173
URI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
URI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10
URI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	90
URI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6
URI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	103
URI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19
URI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10
URI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	18
URI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23
URI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13
URI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
URI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	22
URI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	62
URI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	25
URI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	25







	T	A	B	D	E	G	H	I	J	K	L	M	N	O	P	R	S	T	U	V	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	SUM
VAL	175	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	26	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	219
UAL	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	21
VAL	0	35	0	0	71	0	0	3	5	0	0	0	0	0	0	58	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	373
VAL	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
VAL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7
VAL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23
VAL	2	59	0	0	11	0	0	24	0	0	0	0	24	3	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	146
VAL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	326
VAL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	46
VAL	0	143	0	0	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	205	
VAL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
VAL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
VAL	0	2	0	0	0	0	0	14	1	0	0	0	0	0	0	0	9	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	55
VAL	0	1	0	0	1	27	0	1	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	61
VAL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	48
VAL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14
VAL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	17
VAL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
VAL	0	7	0	0	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	377
VAL	0	25	0	0	35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	304
VAL	0	32	0	0	14	0	0	1	11	0	0	0	0	0	0	7	66	1	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	187
VAL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11
VAL	0	3	0	0	1	1	1	0	0	2	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	48
VAL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
VAL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13
VAL	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15
VAL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6
VAL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20
VAL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	77
VAL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
VAL	0	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	43
VAL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
VAL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10
VAL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	46
VAL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23
VAL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23
VAL	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8
VAL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23
VAL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	24

	I	A	B	C	E	G	H	I	J	K	L	M	N	O	P	R	S	T	U	V	W	X	SUM
YIL	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	2	0	1	0	0	0	21
YIL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
YIL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16
YIL	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
YIL	123	0	0	0	0	0	22	0	0	0	0	0	0	0	4	0	1	1	0	1	0	0	196
YIL	0	0	0	0	0	0	0	0	0	27	0	0	0	0	0	1	0	2	0	0	0	0	31
YIL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
YIL	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	3
YIL	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	3
YIL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	18
YIL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
YIL	2	0	0	0	0	0	0	0	188	0	0	0	0	0	0	0	0	0	0	0	0	0	190
YIL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
YIL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
YIL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
YIL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	85
YIL	14	0	0	0	0	0	0	0	0	0	1	0	0	2	0	0	1	0	0	0	0	0	17
YIL	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	11
YIL	0	0	0	0	0	0	0	0	0	0	7	0	0	0	1	0	0	0	0	0	0	0	9
YIL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
YIL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6
YIL	0	0	0	0	0	0	0	0	0	0	27	0	0	0	0	0	0	0	0	0	0	0	30
YIL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19	0	0	0	0	28
YIL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
YIL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
YIL	0	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	96
YIL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12
YIL	73	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	75
YIL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	21
YIL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12
YIL	12	2	0	0	0	0	0	0	0	23	0	0	0	0	0	0	0	0	0	0	0	0	12
YIL	10	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	39
YIL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	80
YIL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	44
YIL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
YIL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23
YIL	14	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20
YIL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12
YIL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10
YIL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
YIL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9



	SUN	MON	TUE	WED	THU	FRI	SAT	SUN
AMH	7	0	0	0	0	0	0	0
EMH	9	0	0	0	0	0	0	0
IMH	1	0	0	0	0	0	0	0
OMH	1	0	0	0	0	0	0	0
UMH	2	0	0	0	0	0	0	0
BMH	1	0	0	0	0	0	0	0
DMH	1	0	0	0	0	0	0	0
LOH	1	0	0	0	0	0	0	0
DOH	1	0	0	0	0	0	0	0
IRH	1	0	0	0	0	0	0	0
ORH	5	0	0	0	0	0	0	0
URH	0	0	0	0	0	0	0	0
BRH	0	0	0	0	0	0	0	0
ARR	1	0	0	0	0	0	0	0
ISM	1	0	0	0	0	0	0	0
DTM	1	0	0	0	0	0	0	0
AUH	1	0	0	0	0	0	0	0
IUH	1	0	0	0	0	0	0	0
KUH	1	0	0	0	0	0	0	0
LUH	1	0	0	0	0	0	0	0
NUH	4	0	0	0	0	0	0	0
SUH	3	0	0	0	0	0	0	0
TUH	2	0	0	0	0	0	0	0
UOH	0	0	0	0	0	0	0	0
SOH	0	0	0	0	0	0	0	0
AH	15	0	0	0	0	0	0	0
BAH	15	0	0	0	0	0	0	0
CAH	16	0	0	0	0	0	0	0
EAH	20	0	0	0	0	0	0	0
GAH	23	0	0	0	0	0	0	0
HAN	24	0	0	0	0	0	0	0
IAN	1	0	0	0	0	0	0	0
JAN	2	0	0	0	0	0	0	0
KAN	7	0	0	0	0	0	0	0
LAN	7	0	0	0	0	0	0	0
MAN	11	0	0	0	0	0	0	0
PAN	14	0	0	0	0	0	0	0
RAN	14	0	0	0	0	0	0	0
SAN	2	0	0	0	0	0	0	0
TAN	2	0	0	0	0	0	0	0





	I	A	B	D	E	G	H	F	Y	K	J	M	N	O	P	R	S	T	U	V	Y	Q	4	6	U	X	SUM
-10	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
010	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
E10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
J10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
L10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
J10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
U10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7
U10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
-10	0	4	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9
ANC	0	0	0	1	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9
INC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5
MNC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5
UNC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5
-10	1	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
T00	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
K00	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
L00	3	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11
K00	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
S00	4	4	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	12
T00	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
HPC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
ARC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
ERC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
PRC	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
RC	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
R00	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
R20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
OTC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
YTC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
UTC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PVC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
XVC	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
XVC	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
XVC	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	17
XVC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5
XVC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
OAL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
CEL	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
EEF	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1



	-	A	B	D	E	G	H	I	J	K	L	M	M	O	P	R	S	T	U	V	W	X	Y	Z	SUR
MER	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
NER	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16
PER	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7
SER	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	4
TER	0	0	0	0	0	0	2	1	0	0	0	0	0	0	0	0	21	0	0	0	0	0	0	0	23
VER	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
XER	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8
YOR	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
ZOR	0	0	0	0	0	0	0	0	0	2	0	0	15	0	0	0	0	0	0	0	0	0	0	0	10
SOR	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	2	0	0	0	0	0	0	0	4
TOR	0	0	0	0	0	0	0	1	0	16	0	0	0	1	0	0	5	0	0	0	0	0	0	0	11
ARR	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	3
OTR	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
AUR	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
DUR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
HUR	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
JUR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7
KUR	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	7
PUR	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	2
TUR	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
UUR	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	60
VUR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
WUR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
XUR	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
YUR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
ZUR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
ORR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
OUR	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7
AAS	0	0	0	0	0	0	1	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0
BAS	0	13	0	0	2	0	0	0	0	1	25	0	0	6	0	1	0	10	0	0	0	0	0	0	61
DAS	0	7	0	0	20	0	0	171	0	0	16	0	59	0	0	12	0	4	0	0	0	0	0	0	301
EAS	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10
GRS	0	0	0	0	0	0	0	0	0	0	42	0	0	0	0	0	0	1	0	0	0	0	0	0	30
HAS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	113
IAS	0	11	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	21
JAS	0	61	1	5	0	0	0	0	0	0	40	0	0	0	0	0	0	0	0	0	0	0	0	0	61
KAS	189	47	10	0	16	0	0	3	31	0	46	0	0	6	0	0	0	0	0	0	0	0	0	0	155
LAS	0	14	0	0	19	1	0	0	0	110	14	1	1	1	0	19	19	1	26	0	0	0	0	0	456
MAS	0	0	0	0	44	0	4	32	0	0	19	16	0	0	0	4	0	14	0	0	0	0	0	0	182
NAS	0	0	0	0	5	1	5	15	0	0	19	0	0	2	0	0	0	0	0	0	0	0	0	0	280
OAS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100







	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	SUM
UUS	25																									25
VUS	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DUS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
KAS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
008	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
005	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DAT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DAY	0	1	0	0	0	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	17
GAT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HAT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
IAT	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
JAT	0	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	17
KAT	0	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15
LAT	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11
NAT	0	1	0	0	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
PAT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RAT	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
SAT	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
TAT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
UAT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-REP. N18	0	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19
ABY	0	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12
BEF	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
DEF	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5
EFP	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GEF	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HET	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11
IEF	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
JEF	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
KEF	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LEF	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MEF	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PEF	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SEF	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TEF	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0













	SUM	X	O	B	A	G	V	U	T	S	R	P	O	N	X	L	K	J	I	H	G	E	D	B	A	I
-OU	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EU	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
JOU	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HOU	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
UOU	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
OUU	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AVV	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SAV	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DAV	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CAV	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TAV	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
JAV	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
KAV	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LAV	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SAV	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HAV	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RAV	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SAV	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TAV	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
VAV	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LEV	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
OEV	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GEV	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HEV	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
JEV	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LEV	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NEV	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
REV	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SEV	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TEV	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
UEV	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
WEV	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
OMV	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BMV	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DMV	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EMV	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
IMV	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LMV	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
KIV	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LIV	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0



## О ГРАФЕМОТАКТИКЕ КОНЦА СЛОВА В ЭСТОНСКОМ ТЕКСТЕ

Юло Каазик, Юхан Тулдава

### Р е з ю м е

В статье приводятся статистические данные о встречаемости двух-, трех- и четырехбуквенных комбинаций в конце слова в эстонском тексте. Объем текста ок. 95000 словоупотреблений (авторская речь художественной прозы), объем словаря ок. 27000 словоформ. Материал представляется в виде специальных таблиц, в которых в начале строки дается трехбуквенное сочетание конца слова, а сверху в начале колонки приводится четвертая буква с конца слова. Данные можно использовать для решения задач автоматического морфологического анализа слова в тексте.

## ON GRAPHEMOTACTICS OF WORD-ENDING IN ESTONIAN TEXTS

Ulo Kaasik, Juhan Tuldava

### S u m m a r y

The article presents statistical data on the occurrence of two-, three- and four-letter combinations at the end of words in a text. The sample consists of about 95,000 running words (belles-lettres text, non-conversational material only), and the dictionary consists of 27,000 word forms. The material is presented in special tables where the three last letters are given at the beginning of each line and the fourth letter from the end stands above the column. The data can be used for solving problems of the automatic morphological analysis of word forms in a text.

## SISUKORD - CONTENTS

P. Saukkonen, Statistical Linguistic Research in Finland. - Keelestatistiline uurimistöö Soomes (ingl. k.) . . . . .	5-14
F. Papp. Linguistic Statistics and the Hungarian Language (in Russian). - Keelestatistika ja ungari keel (vene k.) . . . . .	15-37
	Summary in English . . . . . 37
T. Zsilka, Statistical Investigation of the Morphological Structure of Poetry (in Russian). - Luulekeele morfoloogilise struktuuri statistilisest uurimisest (vene k.) . . . . .	38-49
	Summary in English . . . . . 50
J. Tuldava, Eesti keele sõnavara foneetilis-grafeemilised mõõted. - Phonetic-Graphemic Measures of the Estonian Lexics (in Estonian). . . . .	51-98
	Summary in English . . . . . 100
U. Kaasik, J. Tuldava, Sõnalõpu grafeemotaktikast eestikeelses tekstis. - On the Graphemotactics of Word-Ending in Estonian Texts (in Estonian)	101-149
	Summary in English . . . . . 150

## СОДЕРЖАНИЕ

П. Саукконен. Статистико-лингвистические исследования в Финляндии (на английском яз.) . . . . .	5-14
	Резюме на русском языке . . . . . 14
Ф. Папп. Лингвостатистика и венгерский язык (на русском яз.) . . . . .	15-37
	Резюме на английском языке . . . . . 37
Т. Жилка. Статистическое исследование морфологической структуры поэзии (на русском яз.) . . . . .	38-49
	Резюме на английском языке . . . . . 50
Ю. Тулдава. Фонетико-графемные измерения лексики эстонского языка (на эстонском яз.) . . . . .	51-98
	Резюме на русском и английском языках . . . . 99-100
Ю. Каасик, Ю. Тулдава. О графемотактике конца слова в эстонском тексте (на эстонском яз.) . . . .	101-149
	Резюме на русском и английском языках . . . . 150

Ученые записки Тартуского государственного университета.  
Выпуск 518. ЛИНГВОСТАТИСТИЧЕСКИЕ ИССЛЕДОВАНИЯ ПО ФИННО-  
УГОРСКИМ ЯЗЫКАМ. Труды по лингвостатистике У. На англий-  
ском, русском и эстонском языках. Резюме на английском,  
русском языках. Тартуский государственный университет,  
ЭССР, г. Тарту, ул. Юликооли, 18. Ответственный редактор  
Я. Соонтак. Сдано в печать 16.01.80. Бумага писчая 30x45  
1/4. Печ. листов 9,5. Учетно-издат. листов 8,96. Тираж  
500. МВ 01026. Типография ТТУ, ЭССР, г. Тарту, ул. Пялсо-  
на, 14. Зак. # 1712. Цена 1 руб. 30 коп.