A FRAMEWORK FOR AUTHENTICATION OF MEDICAL REPORTS BASED ON KEYSTROKE DYNAMICS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS
OF
THE MIDDLE EAST TECHNICAL UNIVERSITY

BY

MUSA KAZIM ÖZDEMİR

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
THE DEPARTMENT OF MEDICAL INFORMATICS

JUNE 2010

Approval of the Graduate School of Informatics

____

Prof. Dr. Nazife BAYKAL

Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

____

Prof. Dr. Nazife BAYKAL

Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

____

Asst. Prof. Dr. Didem GÖKÇAY

Supervisor

Examining Committee Members

Assoc. Prof. Dr. Erkan MUMCUOĞLU (METU)__________________

Asst. Prof. Dr. Didem GÖKÇAY (METU)__________________

Dr. Yeşim AYDIN SON (METU)__________________

Asst. Prof. Dr. Altan KOÇYİĞİT (METU)__________________

Asst. Prof. Dr. Sevgi ÖZKAN (METU)__________________

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name:Musa Kazım ÖZDEMİR

Signature : __________________

ABSTRACT

A FRAMEWORK FOR AUTHENTICATION OF MEDICAL REPORTS BASED ON KEYSTROKE DYNAMICS

Özdemir, Musa Kazım
M.S., Informatics Institute, Department of Medical Informatics
Supervisor: Asst. Prof. Dr. Didem GÖKÇAY

June 2010, 75 pages

Privacy of personal health records is of ultimate importance. Unfortunately, it is easy to obtain illegal access to electronic health records under insufficient security precautions. Access control based on token or username/password is not adequate for applications in health domain which require heightened security. Currently, electronic signature mechanisms are being employed as a strong alternative to classic methods. In addition, biometrics provide more precise results in comparison to electronic signature methods. However, applicability of biometrics in this field has been prohibited by factors such as the need for special hardware, increased implementation costs, and invasiveness of the biometry sensors (eg. iris topology, fingerprint). Behavioral biometrics such as speech, and keystroke dynamics are easier to implement, and do not suffer from the disadvantages mentioned for the static biometrics. Especially, using keystroke dynamics for user authentication is more advantageous than other advanced biometrics because the implementation is inexpensive and continuous identity control is plausible. The aim of this study is to show the feasibility of merging a biometry-based advanced identity verification method together with an initial access control procedure such as password check. In this study, we provide an authentication framework based on measuring similarity of the typing characteristics of medical reporters, while they are typing medical reports. We have made a prototype of the system and provided classification of keystroke timings for each operator. We have generated a testbed and measured similarity of typing patterns of 5 medical reporters upon typing 4 different kinds of medical reports. Our system performs with hundred percent accuracy in identifying the authorized operators from the reports they type. In current practice, electronic signatures are indispensable for health information systems, but our study shows that keystroke dynamics can easily be included in this chain for increased security.

Keywords: Medical Report; Keystroke Dynamics; Electronic Signature; Biometrics; Authentication

ÖZ

MEDİKAL RAPORLARDA KİMLİK DOĞRULAMASI İÇİN TUŞ VURUŞU DİNAMİKLERİ ÜZERİNE TASARLANMIŞ TEMEL BİR ALTYAPI

Özdemir, Musa Kazım
Yüksek Lisans, Enformatik Enstitüsü
Tez Yöneticisi: Yrd. Doç. Dr. Didem Gökçay

Haziran 2010, 75 sayfa

Kişisel sağlık kayıtlarında gizlilik ve güvenlik hayati önem taşır. Buna rağmen yetersiz güvenlik önlemleri nedeniyle elektronik sağlık kayıtlarına kural dışı erişim kolayca sağlanabilir. Sağlık alanındaki yüksek güvenlik gerektiren uygulamalar için belirteç veya kullanıcı adı/parola ile yapılan erişim denetimi yeterli değildir. Günümüzde elektronik imza mekanizmaları klasik yöntemlere güçlü bir alternatif olarak kurgulanmaktadır. Bunun yanında, elektronik imza yöntemleriyle karşılaştırıldığında Biyometri daha kesin sonuçlar sağlar. Ancak, özel donanım gerekliliği, artan uygulama maliyeti ve cihazlara alışma zorluğu (örn. iris topolojisi, parmak izi) gibi etkenler uygulanabilirlik imkanlarını kısıtlar. Konuşma ve tuş vuruşu dinamikleri gibi davranışsal biyometriler uygulama kolaylığı ve kabul edilebilirlik açısından fizyolojik biyometrilerden ayrılırlar. Özellikle, sürekli denetim sağlaması nedeniyle kullanıcı kimlik doğrulaması için düşük maliyetli tuş vuruşu dinamikleri kullanımı daha verimlidir. Bu çalışmanın amacı biyometri tabanlı gelişmiş bir kimlik doğrulama ve tanımlama yöntemiyle parola onay gibi düşük düzeyde bir kontrolün birlikte kullanılabilirliğini göstermektir. Çalışmamızda, medikal rapor yazma esnasında kişilerin tuş basma özelliklerini analiz ederek bir benzerlik ölçüsü belirleyen bir altyapı tasarladık. Bu amaçla, editör kullanımıyla elde edilen tuş vuruşu zamanlamalarının sınıflandırılması için prototip oluşturduk. Geliştirilen sınama ortamında beş farklı operatörün dört gruptan oluşan medikal raporlar oluştururken ürettikleri tuş basma paternlerini karşılaştırdık. Sistemimiz yetkili kullanıcıları yüzde yüz doğrulukla tanımladı. Modern sağlık bilgi sistemleri için elektronik imza uygulamaları vazgeçilmezdir, ancak çalışmamız tuş vuruşu dinamiklerinin güvenliği artırmak için kolay ve etkili bir bileşen olduğunu gösterir.

Anahtar Kelimeler: Medikal Rapor; Tuş Vuruşu Dinamikleri; Kimlik Doğrulama; Biyometri; Elektronik İmza

To My Family

ACKNOWLEDGMENTS

I express my sincere appreciation to my supervisor Asst. Prof. Dr. Didem GÖKÇAY for her patience, guidance and insight throughout the study. I greatly appreciate her share in every step taken in the development of the thesis. Thanks go to my domain expert Hasan Odabaşı, Ms for his technical help for the development of the editor interface and evaluation of my results.

I am grateful to all my friends, especially Cumhur Aydınlı, Hüseyin Akarsu, Kenan Akar, and Zeki Güven for sharing their brilliant ideas, advices, and for supporting me though the best and worst of times.

Encouragements of my wife, Reyhan, for my application to the M.S. program after many years and moral assistance during all this period is also gratefully appreciated. I also acknowledge my son, for coping with my frequent absences during the last two months.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

I.1 Overview

Malicious access to computer systems as well as confidential records is an important problem. A number of security mechanisms, ranging from token or password based authentication systems to more advanced biometric systems such as fingerprint, iris and voice recognition, must be performed for authentication into a computer system with increased security. This is especially required for multi-user applications, including personal health records. Medical records may require more security precautions than ordinary documents, because they contain private and vital information.

Limited number of characters in a password creates a serious handicap for the authentication phase. In this way, authentication is vulnerable to outside attacks. People tend to identify keywords that will be easy to remember, such as specific dates, names, locations and few things that have personal significance. The most significant shortcoming in the use of these mechanisms is the impossibility to verify the people who use the access information is the same person who should be. These are weaknesses in existing security systems to remain vulnerable to attacks. [1]

Biometric authentication has became a popular complementary measure as a result of concerns expressed above. Identification of individuals by a specific physiological or behavioral characteristic constitutes the research area of biometrics. For certain people who have already gained access rights to the environment, biometrics provide important distinctive characteristics. It is almost impossible to forget, lose or copy these characteristics, because they are purely owned physical or behavioral features [2].

The cost of advanced biometric systems requiring additional hardware is high. However, most of the security mechanisms can not protect the system from unauthorized use, simply because they perform verification only at the beginning. The mechanisms should provide continuous control for unauthorized use, but in this case they become intrusive the authentication requires additional checks with specific hardware.

Keystroke analysis is a behavioral biometric and is more advantageous than other advanced biometrics because of its intrinsic properties. First of all, it does not matter what the user types, typing is not subject to restrictions. Secondly, a person's keystrokes may be monitored unobtrusively for long periods, allowing for a continuous user identification. Thirdly, obtaining Keystroke Biometrics is not tedious, because computer users often type with the keyboard. Finally, hardware necessity is only the computer itself, there is no additional expenditure. As early as 1980[3] the rhythm of the keystrokes is analyzed to gather significant information regarding the user's identity.

In current practice, most of the medical reports are delivered to counseling patients and clinics by adding a digital signature. Electronic signatures are absolutely indispensable with full implementation of health information systems which became a major national policy. A health information system must meet the responsibilities and security needs required. Also, the signature or marking process should be included in the report creation workflow. However as an extra measure, keystroke dynamics can also be included in this chain, because unlike other biometrics, performing/analyzing typing rhythms does not require additional equipment, and it can be performed effectively on-the-fly.

I.2 Aims and Scope

The basic idea of this study is to propose a framework that provides additional authentication mechanism in a hospital information system by analysing the keystroke timings being generated during the typing of medical reports. The report entry procedure in medical documents can be used as a testbed for advanced authentication and identification issues based on typing characteristics. In this thesis, we built a prototype system which can monitor a similarity measure derived from the typing characteristics of medical reporters. This framework, when trained on a previously defined operator, is able to authenticate the current operator that has been given access to the hospital information system through a low level procedure such as password check. The similarity measure we used here is based on extracting the keystroke timings obtained from an editor in terms of n-grams[1], for each authorized medical reporter. The intercepted n-gram timings can be stored in a database or embedded within the medical reports.

In section 2, we provide a review of existing authentication and identification systems, with special emphasis on kesytroke dynamics. Section 3 includes the explanation of the methods which we used for implementation of a keystroke based medical report evaluation system. We evaluate the descriptions of the attendant medical report operators and the comparison of the results obtained from the nearest neighbourhood algorithm and factor analysis using SPSS[2] in section 4. Finally, the new security enhancement methodology which we propose to operate within the network of a hospital is outlined in section 5.

[1] N-gram: Time required to type n consecutive characters in the medical report. For ex: the word benign has the following 2-grams: be, en, ni, ig, gn and the following 3-grams: ben, eni, nig, ign.

[2] Statistical Package for the Social Sciences

CHAPTER II

BACKGROUND

Electronic medical records require more security precautions than ordinary documents that belong to individuals, because they may contain private and vital information. Medical experts are responsible for the privacy of the information they established or reproduced for the patients. For example some diagnoses, surgical operations, adverse reactions, HIV[3] status or specific kind of injuries can not be shared or distributed.

Supporting information protection laws with the addition of prescriptive clinical experiences, consent and sharing principles is accepted as security policy dedicated to the higher sensitivity regarding vulnerability of personal security. Instead of newly invented propositions, below principles are an adaptation of widely accepted policies within computer-based medical systems[4]:

— Access control (including the “read” and “append” roles) to the clinical information should be provided.
— A new medical record might be opened on several security levels.
— An authority level must be determined from the access control list.
— Any changes, except for emergencies, must be reported to the possessor.
— Useful information can not be deleted before expiration.
— Access logs must be kept with timestamp.
— Without consent, the flow of information is allowed through the more sensitive records only.
— In order to keep privacy, data collection from many patients should be avoided.

[3] Human Immunodeficiency Virus

— A stand alone healthcare system should provide the principles above.

In addition, a well designed health information system provides:
—Computer security for authentication and access control;
—Communication security of a local area network;
—Statistical security for sharing records without providing identification;
—Availability mechanisms against loss of any information.

According to United States government's Office of Technology Assessment main insider threats to healthcare systems are listed below[47]:
— Messages between clinicians could be corrupted or altered because of system failures.
— Unstructured communication methods may result misdiagnoses.
— An attacker might change the clinical information or manipulate reports.
— The physicians might modify the record after noticing malpractice.
— Clinicians may consult each other for precise diagnoses and share reliable information.

For these reasons, protection and security of health-related information gained popularity. In order to ensure verification, accesibility or integrity for the clinical records, there should be an alternative proof of concept. Electronic signature provides additional security for this purpose. Once the security policy is provided, other accesibility issues can be handled easily [5].

II.1 Electronic Signature

The rapid development of communication technologies and easy-to-use applications made approximately 2 billion people online[46]. Therefore, users who connect to the virtual world have created a virtual personality, which is represented by electronic signature. Now all known methods for distinguish between people can be considered as a kind of electronic signature.

Due to widespread use of electronic commerce, legal definitions are accepted as standard descriptions and are included in the international commercial law. Electronic Signature is defined in PIPEDA[4], ESIGN[5], GPEA[6], UETA[7], by Federal Reserve[8], Commodity Futures Trading Commission, Food and Drug Administration, etc. Here we are including a single definition that belongs to PIPEDA:

An electronic signature is "a signature that consists of one or more letters, characters, numbers or other symbols in digital form incorporated in, attached to or associated with an electronic document".

II.2 Electronic Signature Categories

The term electronic signature which stands for all kinds of digital identifiers is a generic expression. There are two main categories:

—Digital signatures

—Biometrics

II.3 Digital Signature

Digital signature is a special kind of electronic signature based on worldwide accepted standards[6]. It can be used to both authenticate the identity of the creator of a document and to assure the integrity of the document[7]. Additionally digital signatures provide accountability and non-repudiation of electronic documents[8]. Digital signatures are embedded into electronic documents through public key infrastructure[45].

Digital signatures are fundamentally based on the asymmetric key algorithm also known as public key cryptography. There are two keys used by a mathematical

---

2 *The* Personal Information Protection and Electronic Documents Act

3 ELECTRONIC SIGNATURES IN GLOBAL AND NATIONAL COMMERCE

4 Government Paperwork Elimination Act

5 Uniform Electronic Transactions Act

*6*The Federal Reserve is the central banking system of the United States algorithm, one for creation of the signature named private key and the other for verification named public key. RSA[9] [9] and DSA[10] [10] are well-known algorithms which are standardized for digital signatures. Some important terms related to these algorithms are as follows:

Private key: used to create the signature and known only to the owner

Public key (PK): used to verify the signature

Hash function: a mathematical function which is used for digital signatures

Main areas within which digital signatures can be used are as follows:

Owner authentication: Through a dedicated digital signature with public and private keys that always refer to the document owner, the owner of the document could be identified.

Document authentication: By means of the private and the public keys, hash results are compared thus the document can be validated.

Affirmative act: By using a private key, digital signature warns the owner about consequences of the confirmation.

Efficient assurance: Digital signature provides more qualified assurance although creation, verification and acception processes demand some extra time.

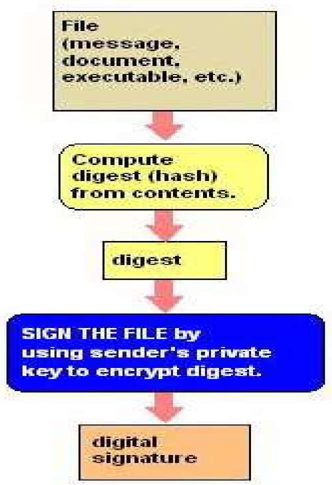

Figure II.1 An Encrypted Digest [11]

In order to clarify the processes associated with digital signature, we present some figures. In FigureII.1 above, summary of a document derived by a hash function and

[9] Rivest, Shamir and Adleman

[10] Digital Signature Algorithm

the encrypted result is shown. In Figure II.2, the validation process of a document encrypted with a hash function and public key is shown. In Figure II.3, privacy has been added to the previous method with the recipient's public key.

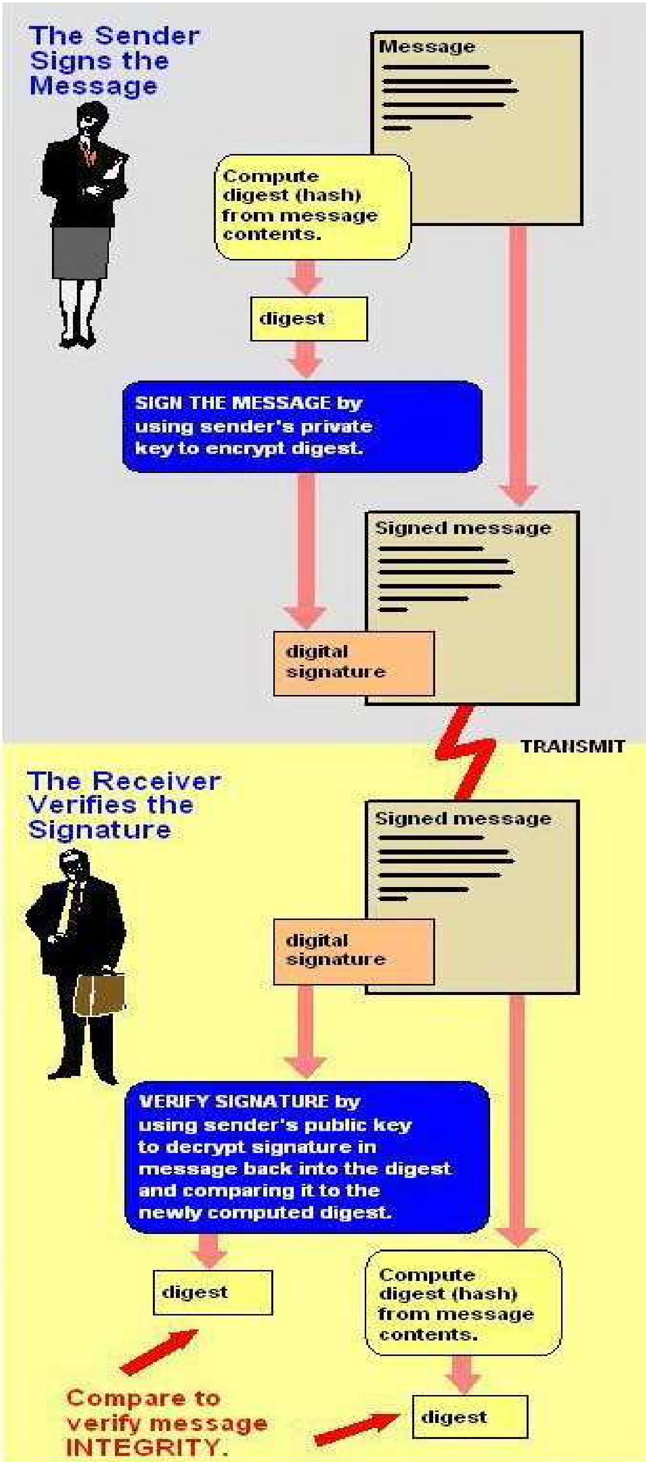

Figure II.2 Integrity, But No Privacy [11]

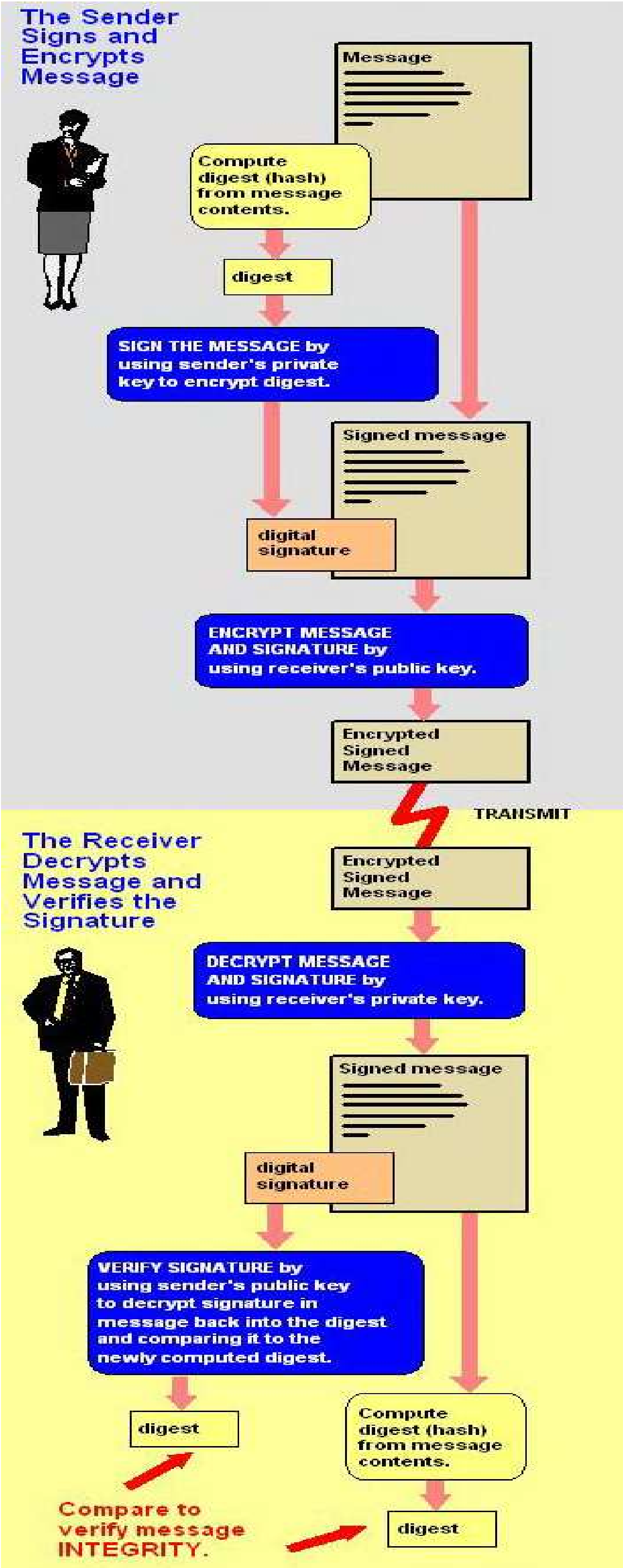

Figure II.3 Message Integrity and Privacy [11]

II.4 Biometrics

II.4.1 Definition:

Biometric signatures are second main class of electronic signatures. People naturally have been the beneficiaries of voice and facial features to identify others since the beginning of humanity, because biometric attributes are unique characteristics. Thus biometrics stand for calculated measures of these naturally owned characteristics. Computer science methods derive numerous kinds of biometric measures to be used for identification. Biometric authentication is established to distinguish a possessor from a forger by using scientific measurements extracted from physiological or behavioral characteristics. The universe of a selected biometric characteristic should have a sufficient variety of features [12]. In addition, if there is a healthy and well managed method of sending and receiving data; duplicating, cheating, sharing or distributing of a biometric can not occur[13]. Here, we would like to present a brief history of biometrics at our times[15].

—1823: Czech Professor Johannes Evangelista Purkinje made fingerprint classification

—1858: Sir William Herschel from England used palm and fingerprints on the contracts in India

—1890s: French anthropologist Alphonse Bertillon developed bertillonage[11] method using some body measurements

—1890s: Sir Francis Galton declared fingerprints study derived from 10 fingers named as Galton's Details

—1901: Sir Edward Richard Henry established fingerprint files in England. 'Henry Classification System' is used since then

—1903: First fingerprints system in the U.S. was implemented in New York State Prison for criminals

—1904: The use of fingerprints in Kansas and Missouri

[11] A method used in the late nineteenth century that identifies people with multiple body measurements such as finger length, width, heigth head circumference. All the measures are recorded and compared for identification. [14]

—**1905**, and **1907:** The use of fingerprints in the U.S. Army and Navy

—**1960s:** Woody Bledsoe from California developed a technique called “man-machine facial recognition” by using feature extraction process

—**1981:** The term “biometrics” is first used in an article in The New York Times

—**1980s**: Dr. John Daugman developed iris recognition technology in Cambridge University

—**1984**: Hand geometry in The University of Georgia (U.S.) dormitory food service areas and signature recognition systems studies in the Stanford Research Institute (U.S.) and the National Physical Laboratory (United Kingdom)

—**1985:** Deployment of access control with retinal scanning systems at the Naval Postgraduate School

-Mid-1980s: Fingerprints are collected for driver license applications in the State of California

—**1986/1987** The International Biometrics Association (IBA) was founded

—**1998:** The International Biometric Industry Association (IBIA) was founded in Washington, DC

—**2001:** The National Biometric Security Project (NBSP) was founded

While devising a system based on biometric measures, some preliminary criteria that should be satisfied according to the individual are: **Universality (**Everyone should have it, **Distinctiveness** (No two should be the same), **Permanence (**It should be invariant over a given period of time), **Collectability (**Should be collectable through digital technology). In addition, **performance, circumvention**, and **acceptability** are preliminary conditions that are important from the system perspective for a biometric measure.

Systems that incorporate a biometric signature mechanism should consist of the following process modules(Figure II.4):

—Sensor module

—Feature extraction module

—Matching module

—Decision-making module

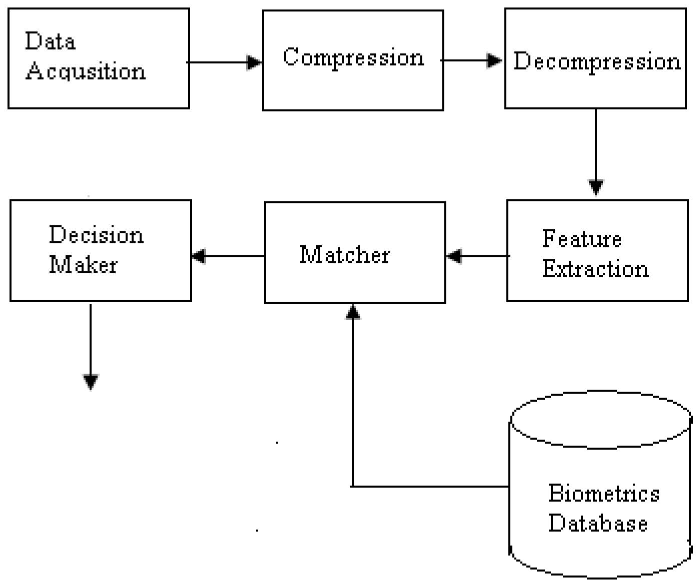

Figure II.4. Biometric Authentication Systems [16]

In order to use biometric identification, some features are extracted from physiological or behavioral characteristics. To identify an individual, the system needs a dedicated template “an individual biometric template” for the subject which is gathered in the enrollment phase, as a part of the sensor module [17]. Access to the system is allowed only if the comparison between the instant features and the template is satisfactory [16].

Given the above operational principles, there exists two main distinct use cases of biometrics[18] (Figure II.5);

—Identification or recognition: One-to-many comparison of a gathered biometric feature and the database consisting of biometric templates for every individual.

—Verification: One-to-one comparison of a new captured biometric and the database consists of the templates for the individual.

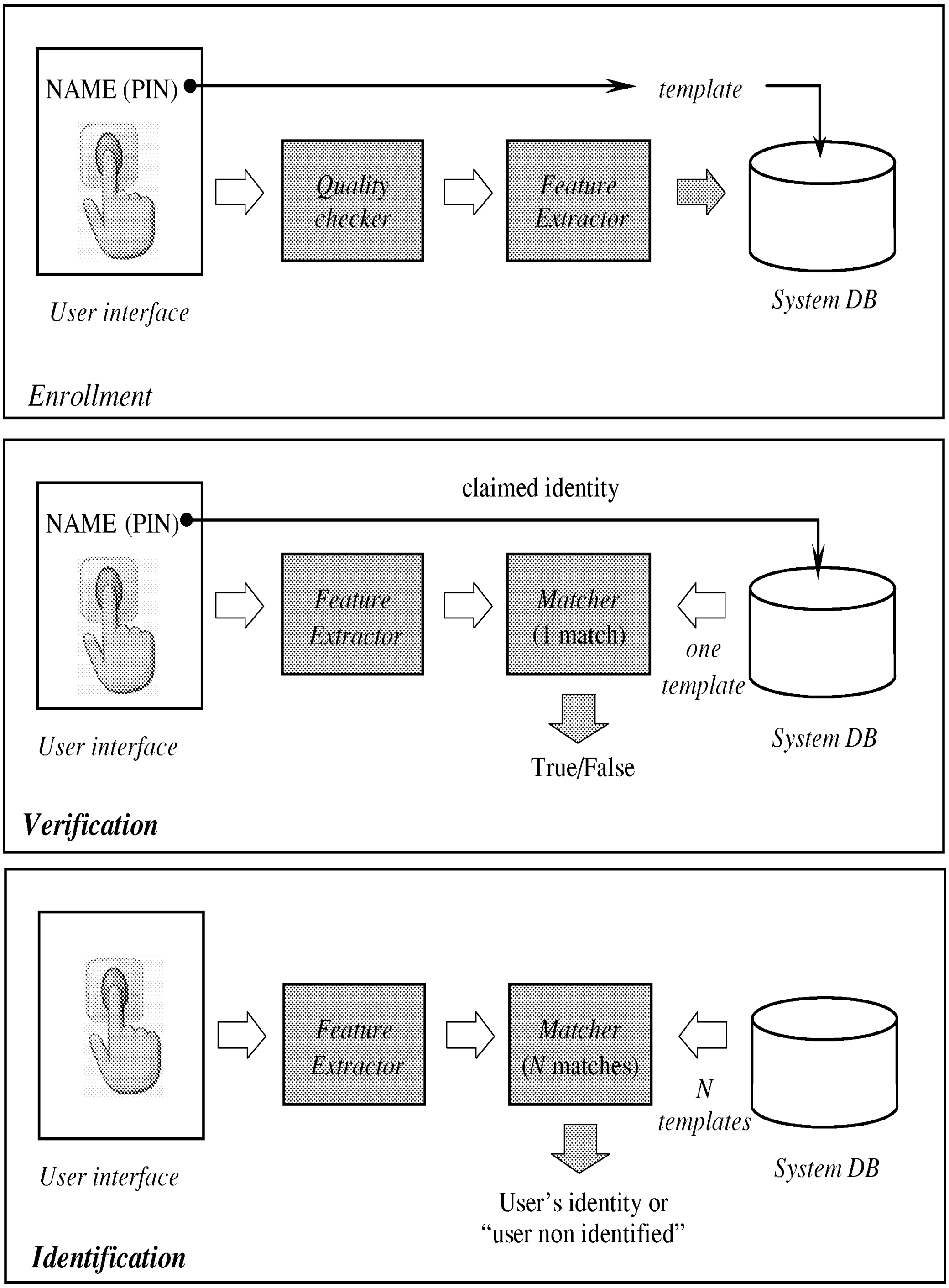

Figure II.5. Modular Representation of the Main Tasks for Biometrics [18]

Considering the requirements and basic components, the biometric methods are to be divided into two main classes(Table II.1):

1. **Physiological Characteristics** use static biometric feature vectors which are always present independent of any action. These methods are considered to be **Static methods.**

2. **Behavioral Characteristics** use dynamic feature vectors which are acquired by efforts made by the person. These methods are also known as **Dynamic methods**.

In table II.1, some physiological and behavioral characteristics are summarized.

Table II.1. Physiological and Behavioral Characteristics Used by Biometrics [17]

Method		Description
Physiological	Face recognition	Uses a camera to take a snapshot of the users face. Key measurements are extracted from the image and are compared to the stored 'faceprint'.
	Facial thermogram	Uses the varying temperatures emanating from different regions of the face to provide the basis for characterizing individuals.
	Fingerprint recognition	Uses optical, capacitive or thermal techniques to assess the characteristic patterns of forks and ridges in the print that can distinguish one person from another.
	Hand geometry	Measures the physical dimensions of the hand (e.g. span, length of fingers) when spread on a flat surface.
	Iris scanning	A snapshot of the user's iris, taken by a camera, is compared with a previously stored image.
	Retinal scanning	Uses a laser to scan the distinctive patterns on the retina at the back of the eye.
	Vain checking	Uses infra-red light to enhance the visibility of characteristic vein patterns in the back of the hand.
Behavioural	Gait recognition	Characterises individuals be the way in which they walk.
	Keystroke analysis	Monitors typing activity in order to determine characteristic rhythms. Can be performed on the basis of know text (e.g. in conjunction with a username and password) or upon keyboard inputs in general
	Mouse dynamics	Monitors mouse-related activity, and attempts to characterise users on the basis of measures such as speed and accuracy.
	Signature analysis	Assesses a handwritten signature captured using a special pen and/or pad. Static analysis simply assesses the resulting pattern, whereas dynamic systems also measure the pressure and speed of the signature.
	Voice verification	A user's voice compared with a previously stored 'voiceprint'. Can be performed on a text-dependent basis, when speaking a known word or phrase, or text-independently.

II.4.2 Physiological Characteristics

There is no need of individual effort or individual action to have physiological characteristics. On the contrary, the person is unable to change the reference feature. Below is a summary of these characteristics.

—Fingerprint: It is the oldest method of all biometrics. Fingerprint can be used for both verification and identification processes.

—Facial Features: Uses the shape and/or location of facial features; the comparison of an image and gathered combination of numerous faces is measured to verify identitiy.

—Hand Geometry: Relative locations and measurements of hand features are used. There is a special method named orthographic scanning for hand geometry.

—Iris Feature: The iris is genetic independent and it has numerous features to gather unique measurements.

—Retina Identification: Retina Identification has genetic independence; but it can be captured via infrared light.

—Vein Recognition: Veins have prominent patterns that can be observed on door handles, biowatches, etc.

—Palmprint: Palmprint recognition may be more reliable but costs much more than fingerprints.

—Voice: Similar algorithms are used with face recognition patterns. The voice has effective features for authentication.

II.4.3 Behavioral Characteristics

Behavioral Characteristics are determined via specific response of the individual and as a result of this they include an enrollment process involving repeated activity for each personality. Below is a summary of the types and definitions of behavioral characteristics.

—**Speaker Recognition**: Speaker recognition is a cost efficient and natural method which is both convenient for recognition and identification. Speaker recognition implementation is easy to settle and it is more popular than other methods.

—**Signature Dynamics**: Dynamic signature analysis measures the changes in pressure, position and velocity of the pen, obtains the manner and stylus used by the owner of the signature and also tracks hand movements of the signatory.

—**Keystroke Dynamics**: Keystroke Dynamics examines specific time intervals spent on the keyboard that makes typing patterns unique. Keystroke Dynamics is a practical and economical method. It is a very new kind of user authentication method but it may be easily implemented by many applications.

Overall, there are several advantegeous and disadvantegeous parts for each biometric method as listed in tables II.2 and II.3.

Table II.2 Comparison of Biometric Methods Regarding General Biometric Requirements[19] (H:High, M:Medium, L:Low)

Biometric characteristic	Universality	Distinctiveness	Permanence	Collectabillity	Performance	Acceptability	Circumvention
Facial thermogram	H	H	L	H	M	H	L
Hand vein	M	M	M	M	M	M	L
Gait	M	L	L	H	L	H	M
Keystroke	L	L	L	M	L	M	M
Odor	H	H	H	L	L	M	L
Ear	M	M	H	M	M	H	M
Hand geometry	M	M	M	H	M	M	M
Fingerprint	M	H	H	M	H	M	M
Face	H	L	M	H	L	H	H
Retina	H	H	M	L	H	L	L
Iris	H	H	H	M	H	L	L
Palmprint	M	H	H	M	H	M	M
Voice	M	L	L	M	L	H	H
Signature	L	L	L	H	L	H	H
DNA	H	H	H	L	H	L	L

Table II.3 Biometric Methods and Features [19]

Fingerprint Recognition	physiological	Scan personal fingerprint and transform to template.	➢ Easy to use. ➢ Fingerprint is unchanged and unique. ➢ Cheap.	➢ Easy to copy. ➢ If hand is dirty or injured, the accuracy will low.
Voice Recognition	behavioral	Record the sound of the user then extract the vocalization feature.	➢ Easy to use.	➢ The result is affected easily by the environment. ➢ Authentication speed is slow.
Iris Recognition	physiological	Iris recognition uses camera technology, and subtle IR illumination to reduce specular reflection from the convex cornea to create images of the detail-rich, intricate structures of the iris.	➢ The highest accuracy. ➢ It is not easy to affect by the outside environment.	➢ Expensive. ➢ Not easy to use. ➢ User excludes this method.
Facial Recognition	physiological	Use webcam or other devices to get human face.	➢ Easy to use. ➢ Fast.	➢ The result is affected easily by the change of the face, eg. Glasses.
Keystroke Dynamic	behavioral	Use keyboard key in keystroke pattern.	➢ Cheap. ➢ Don' t need other device. ➢ Easy to use. ➢ Implement fast.	➢ Affected by the user state, eg. Drunkenness.

Discrimination sensitivity also varies between biometric methods for both physiological and behavioral characteristics. There are a number of publications and studies that indicate the percentages of positive or negative results depending on the constraints or variable conditions (Figure II.6).

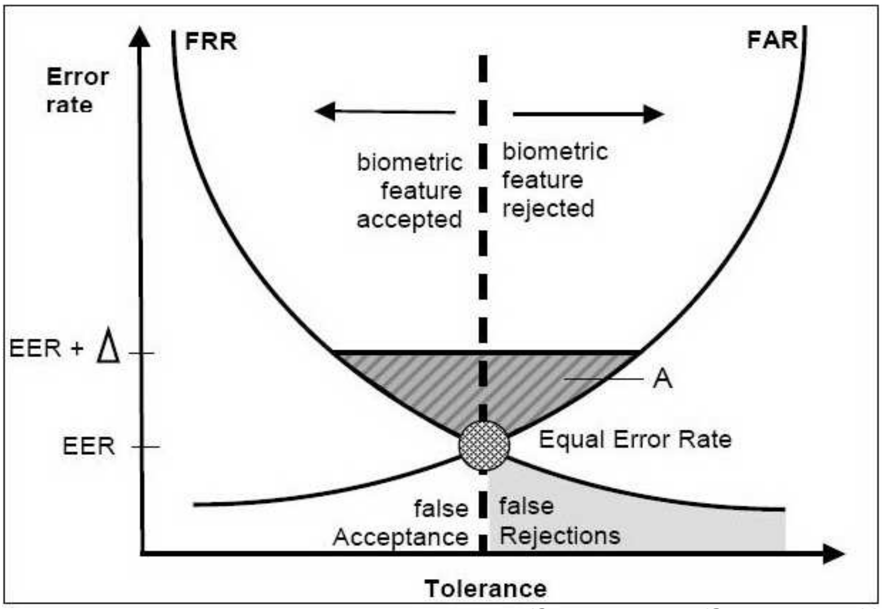

Figure II.6. Determining Decision Point (EER[12]) with FAR[13] and FRR[14] Curves [19]

As seen from the below table II.4, FAR or FRR of behavioral biometric methods are higher than the physiological biometric methods.

Table II.4 Evaluation of Biometric Techniques[20]

Biometric	EER	FAR	FRR	Subject	Comments
Face	NA	1%	10%	37437	Varied light, indoor/outdoor
Finger print	2%	2%	2%	25000	Rotation and exaggerated skin distortion
Hand geometry	1%	2%	2%	129	With rings and improper placement
Iris	.01%	.94%	.99%	1224	Indoor environment
Keystrokes	1.8%	7%	.1%	15	During 6 months period
Voice	6%	2%	10%	30	Text dependent and multilingual

[12] EER- Equal Error Rate: The value indicates that the proportion of false acceptances is equal to the proportion of false rejections according to the predefined treshold.

[13] FAR- False Acceptance Rate: The measure of the likelihood that the biometric security system will incorrectly accept an access attempt by an unauthorized user.

[14] FRR- False Rejection Rate: The measure of the likelihood that the biometric security system will incorrectly reject an access attempt by an authorized user.

An overall account of keystroke dynamics, in terms of advantages and disadvantages can be summarized as follows[21]:

1. Accuracy: Keystroke Dynamics based solutions are more precise than some other biometrics. But error types and error rates may vary according to the principles of the management.

2. Access Speed: The authentication procedure for the user starts from logging into the system. The user is physically and logically at work throughout the time.

3. Fraud Detection: Although Keystroke Dynamics application is installed, additionally, password or ciphers should also be used for authentication every time. By this way Keystroke Dynamics will prevent fraud effectively.

4. Reliability: The only enrollment is through keyboard. Error rates may increase if the keyboard is older, or mechanically challenged.

5. Storage Requirements: A significant advantage is not present in using the data storage. Probably Keystroke Dynamics use less space than visual methods.

6. Enrollment time: The enrollment process does not need a separate location, or particular time, or administration, or any intervention. In fact there is no additional enrollment time either.

7. Interference detection: Keystroke Dynamics based systems do not use any additional equipment repellently and personal information is not stored by the administration.

8. Adoption: There is no obstacle for user adoption, Keystroke Dynamics is naturally available.

9. Cost: Keystroke Dynamics is unrivaled in terms of implementation costs. Any keyboard is enough for instantiating Keystroke Dynamics based authentication.

II.4.4 Keystroke Dynamics

There are many hypotheses on the characteristic stability of keystrokes for each person. Keystroke characteristics provide enough information for establishing feature extraction methods.

Some people may exhibit great differences in the manner of typing. As a countermeasure for this situation, "continuous verification" of an individual in certain periods is possible[22].

Keystroke Biometrics can be a source of attraction for the following reasons:

— A person's keystrokes may be monitored unobtrusively for long periods, allowing for a continuous user identification.

— Obtaining Keystroke Biometrics is not tedious, because computer users often type with the keyboard.

— Hardware necessity is only the computer itself, there is no additional expenditure.

— Internet applications using keystroke biometrics increase the ratio of e-commerce by creating a balance between security and ease of use[23].

Characteristic of a person's typing rhytm provides identity verification with FAR of 0.01% and a FRR of 3,0% approximately. A comparison of these error rates between some other biometrics is shown below (Table II.5). In table II.4 FAR is presented as %7 for 15 subjects according to a 6 months time study but in table II.5 the rate is %0.01. This is because the second measure represents a sophisticated commercial product[25] that claims to provide identitiy verification according to Keystroke Dynamics. Because FRR go beyond the acceptable limit, Keystroke Dynamics alone can not establish the identification task but may be assigned to the verification mission of multi-layer authentication systems[24].

Table II.5 Error Rate Comparisons[21]

Biometrics	FAR	FRR
Fingerprint	~0%	~1%
Voiceprint	~1.6%	~1.8%
Typeprint	~0.01%	~3.0

History

Keystroke Dynamics technique goes back to the early years of telegram. Anecdotally, the operators were able to recognize each other while taking messages. There was even a spacial name of this technique called “Fist of the Sender”. In the second world war, it was used effectively for identity verification. Below figure II.7 shows how this technology has improved over time.

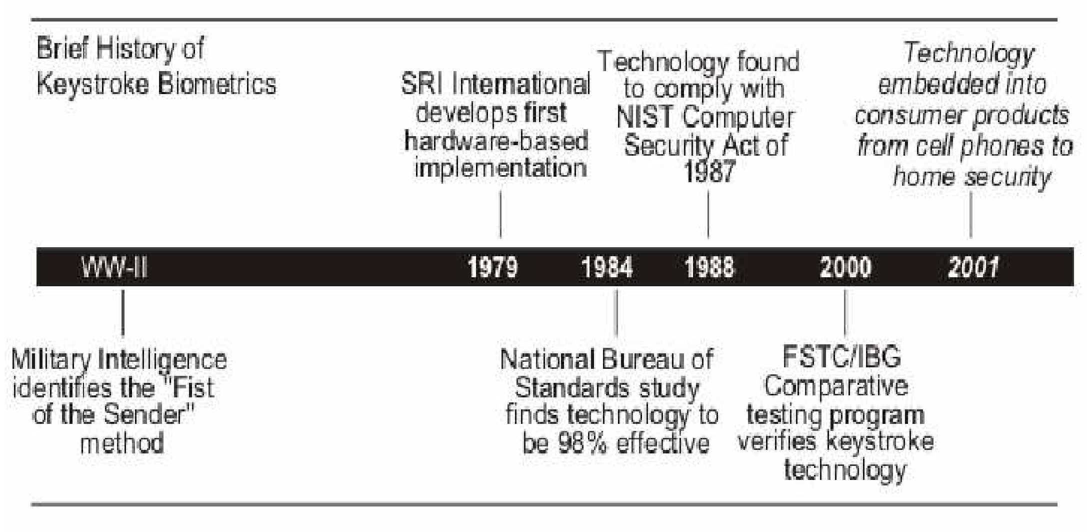

Figure II.7. Keystroke Dynamics Timeline[25]

Innovation and achievement in the *Gaines* [3]study in 1980 is the biggest cause of the interest in keystroke dynamics at our times. The results of this study was not astonishing but the method was a complete novelty. The difference between a reference vector and an instant vector is calculated with a new algorithm which is established by *Joyce & Gupta in 1990*[26]. After that *Gokcay (Unlu)* implemented keystroke rhythm analysis by using two neural network algorithms, back propogation and self organizing feature maps in 1991 [12], later this is replicated by *Obaidat an Maccairolo* [31]. In 1997 Euclidean Distance and probabilistic calculations for the latency times for one-digraph were used by *Monrose and Rubin* [27]. They also studied on identification algorithms in 2000 [28] and the following year they used polynomials and vectors for the keystroke patterns in order to hardening passwords[29].

Table II.6 represents the summary of previous works on Keystroke Dynamics.

TableII.6. Analyses of Previus Works[30]

Author	Year	Measure	Evaluation Method	Test Data	Subjects
Gaines et al. [3]	1980	DG	statistics (T-test)	text, words and phrases	6
Joyce, Gupta [26]	1990	DT vs.FT	statistics	L,P, two well known strings	33
Gokcay (Unlu) [12]	1991	DT vs.FT	neural networks	well known phrase	17
Obaidat, Macchairolo [31]	1993	FT	neural networks	phrases	6
Brown, Rogers [32]	1993	FT	neural networks	names	N/A
Obaidat, Sadoun [33]	1997	DT	neural networks	P	15
Monrose, Rubin [27]	1997	DG	statistics	L,P and first name and surname	63
Berganado et al. [34]	2002	TG	statistics	text, words and phrases	44
Mroczkowski [35]	2004	DT	statistics	L,P	20
DG - di-gram, DT - dwell time, FT - filght time, TG - tri-gram, L - login, P – password					

These studies have used various methods for extracting valid patterns. The vast variety of methods revealed a large number of different results for comparison, as shown in Table II.7.

TableII.7. Several Methods Used[36]

Evaluation Method	Measure	Subjects	FRR	FAR	Author (Year)
Statistical Method	Duration Latency Down-down	10	1.45%	1.89%	Araujo *et al.* (2005)
Statistical model HMM-Gaussian Model	Duration Latency	20	2.54%	2.54%	Jiang (2006)
Statistical method ANN-Fuzzy Logic	Latency	15	2.00%	6.00%	Haidar *et al.* (2000)
Monte Carlo Method Decision Tree	Duration Latency	9	9.62%	0.88%	Sheng *et al.* (2005)
Hidden Markov Model	Duration Latency	20	-	-	Chen and Chang (2004)
	Duration Latency	20	-	-	Chang (2005)
Fuzzy Logic	Latency	varies	7.40%	2.80%	de Ru and Eloff (1997)
Fuzzy C-means	Duration Latency	-	-	-	Mandujano and Soto (2004)
Neural Network K-Nearest algorithm	Latency	-	15.37%	1.03%	Wong *et al.* (2001)
Neural Network, K-Means	Duration Latency	-	-	-	Obaidat and Sadoun (1997)
Genetic Algorithm SVM	Duration Latency	150-400	3.54%	0%	Yu and Cho (2003)
	Duration Latency	150-400	3.69%	0%	Yu and Cho (2004)
SVM	-	-	0.1%	0.02%	Sang et at. (2004)
Similarity Measure	Duration Latency	-	7.6%	-	Monrose and Rubin (2000)
	Latency	10	13.30%	0.17%	Joyce and Gupta (1990)
Vector Analysis	-	15	5%	-	Ibrahim (2003)
Minimum distance Bayesian	-	-	8.10%	2.80%	Bleha *et al.* (1990)

In the last decade several contemporary methods and their combination was introduced. (Table II.8) By applying fuzzy logic algorithms[39], Keystroke Dynamics is investigated in this domain. Moreover, Neuronal Networks based on backpropogation algorithm [12] or support vector machines (SVM) have been tried [40]. This situation indicates that the research topic is appealing and will continue to be developed further[41].

Table II.8. Authors, Methods, Estimated EER Ordered by Years [37]

Article	Method	Estimated EER
Umphress and Williams (1985)	statistical	9.00%
Bleha et al (1990)	Bayes	5.45%
Obaidat and Sadoun (1997)	Neuronal Networks	5.8%
Robinson et al (1998)	Statistical	10.00%
Coltell et al (1999)	Statistical	17.5%
Berganado et al (2002)	Disorder measure	2.00%
Kaholia and Pandit (2003)	Cauchy distribution	2.9%
Guven and Sogukpinar (2003)	Cosine	6.35%
Yu and Cho (2004)	SVM[15]	3.14%
Rodrigues et al (2006)	HMM[16]	3.6%
Clarke and Furnell (2006)	Neuronal Networks	5.0%
Filho and Freire (2006)	HMM and Statistical	6.0%
Hwang et al (2006)	SVM	0.25%
Hoquet et al (2007)	Statistical and fusion	5.0%
Revett et al (2007)	bioinformatic	0.175%

Our study proposes that Keystroke Dynamics can be used for report medical authentication (1:1 tests) and user identification (1:N tests) [38]. The concept may not be customary for behavioral biometrics, but with this proposal, the information security of health records will reach more secure solutions.

[15] Support Vector Machines

[16] Hidden Markov Model

CHAPTER III

IMPLEMENTATION OF A KEYSTROKE BASED MEDICAL REPORT GENERATION SYSTEM

Several text-based forms are produced as a result of a multitude of transactions in health informatics. General medical texts, discharge summaries, adverse event detection, gene and drug data, clinic-specific notes, labaratory results, medical curriculum, pathology reports, and medical imaging reports can be mentioned among these. Physicians, nurses, technicians, secretaries, or implemented hardware create the text documents according to the situation or the way of handling mechanism. Some of the reports contain mostly numerical data while some reports consist of mostly text. These reports can be categorised as structured or pattern based.

Keystroke dynamics can be extracted from personal specific reports which are text-based, containing less numeric, fluently expressed linguistic features as well as medical terminology. Medical imaging reports are the leading types of text-based information that come to mind at this point.

Medical imaging examines the human body for diagnostic, therapeutic or scientific research purposes. Medical imaging reports are mostly generated by nuclear medicine or radiology through magnetic resonance imaging (MRI), photoacoustic imaging, breast thermography, tomography or ultrasound. These report types are all included in our study.

In order to evaluate the keystroke dynamics from a given report, a new software application is developed with Delphi Programming Language. The system is developed for Windows based platforms. In addition, the key feature for the hardware is that the microprocessor must support a high resolution performance counter which means the installed hardware has nano-second precision. To accomplish this constraint, the sofware includes a system check feature. The ‘system check’ (SISTEM KONTROL) button must be clicked first. If a “yes” (EVET) is returned, that means the processor satisfies the nano-second precision. If a result of “no” (HAYIR) is encountered, that means the hardware is not compatible (Figure III.1).

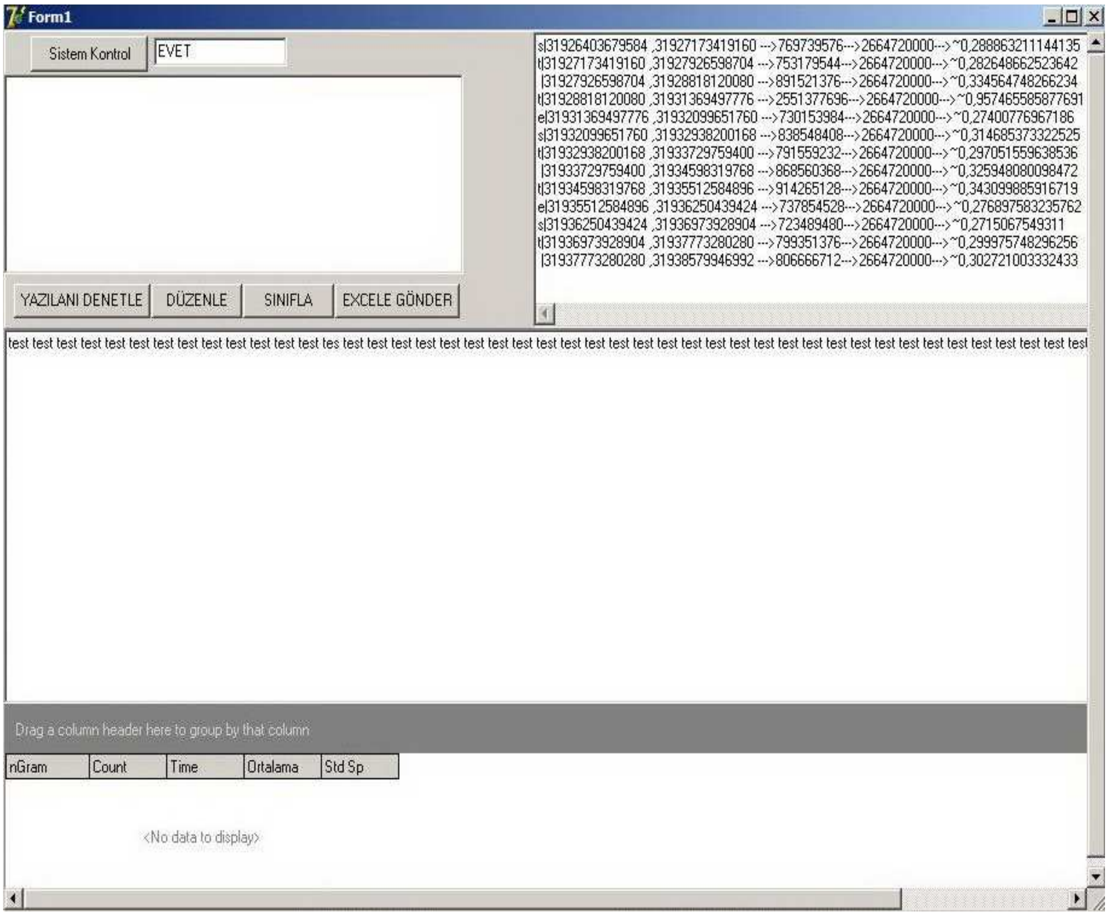

Figure III.1 Initial GUI Screen for the Hardware and Heystroke Test

The medical report is entered by the operator to the space provided in the middle. This space works just as any text-editor, featuring, ‘delete’, ‘backspace’, and ‘carriage return’. Turkish characters are supported by our application as well. In the meantime, transparent to the user, a ‘keystroke collector’ works on the background. Keystroke collector function is written in delphi programming language. Calling this function runs Microsoft Windows QueryPerformanceFrequency API which returns the timing in seconds that corresponds to the difference between the key transactions values on QueryPerformanceCounter over the microprocessor frequency. Since windows is a multitask operating system, if background processes are not turned off, these processes will introduce variance into the keystroke timings. Related description of the microsoft code is as follows[42]:

> ***QueryPerformanceFrequency Function***
> *Retrieves the frequency of the high-resolution performance counter, if one exists. The frequency cannot change while the system is running.*
>
> ***Syntax***
> *Copy*
> *BOOL WINAPI QueryPerformanceFrequency(*
> *__out LARGE_INTEGER *lpFrequency*
> *);*
>
> ***Parameters***
> *lpFrequency [out]*
>
> ***LARGE_INTEGER***
> *A pointer to a variable that receives the current performance-counter frequency, in counts per second. If the installed hardware does not support a high-resolution performance counter, this parameter can be zero.*
> ***Return Value***
>
> ***BOOL***
> *If the installed hardware supports a high-resolution performance counter, the return value is nonzero.*
> *If the function fails, the return value is zero.*

After completion of the medical report, creation of the related keystroke timings in terms of n-grams is performed through a 4 step procedure as illustrated in Figure III.2.

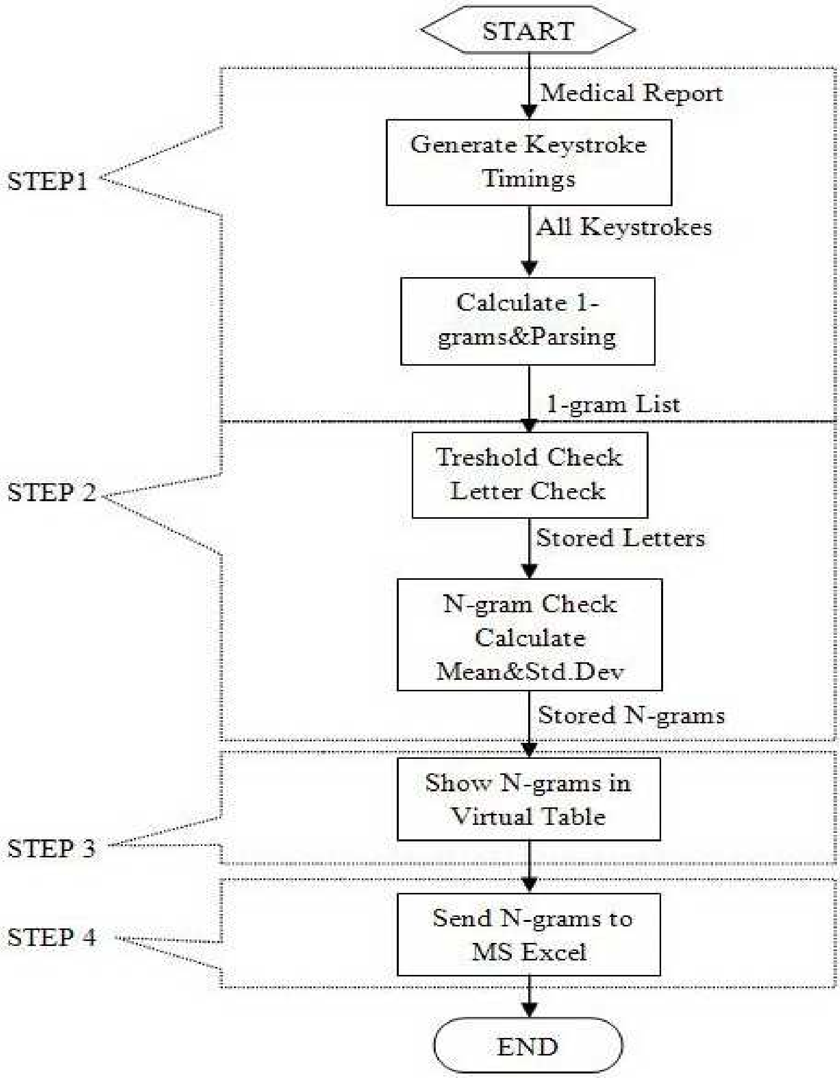

Figure III.2 Flowchart for the Application

Step 1**:** YAZILANI DENETLE: "spell check" is performed. The characters which are not included in the final text are cleared through a spell check process running concurrently with the parsing of the report. For each report, a string list recording the use of individual keys is created.

Step 2: DUZENLE: In the "editing" process, the resulting sequence of characters are rearranged as two, three and four grams of keystrokes (see Figure III.3). The keystroke list obtained by step 1 is re-parsed, and a pointer list is maintained according to a predefined threshold **[43]** filter. Currently, this filter removes keystrokes with a timing higher than 400 ms. This is because any keystroke with such a high value might actually correspond to breaks or pauses in the typing, hence not reflecting the characteristic typing pattern of the operator.

Step 3: SINIFLA: In the process of “classification”, all resulting n-grams are examined in virtual memory and transferred to the virtual table respectively. If the instant n-gram has already been on virtual table, the counter, the mean and standard deviation is updated, otherwise it is inserted. For each n-gram, the mean and standard deviation values are calculated and listed inside the editor in form of a table as seen at the bottom of figure III.3.

Step 4: EXCEL’E GONDER: The data in the n-gram table, will be transferred to Microsoft Excel for permanent storage.

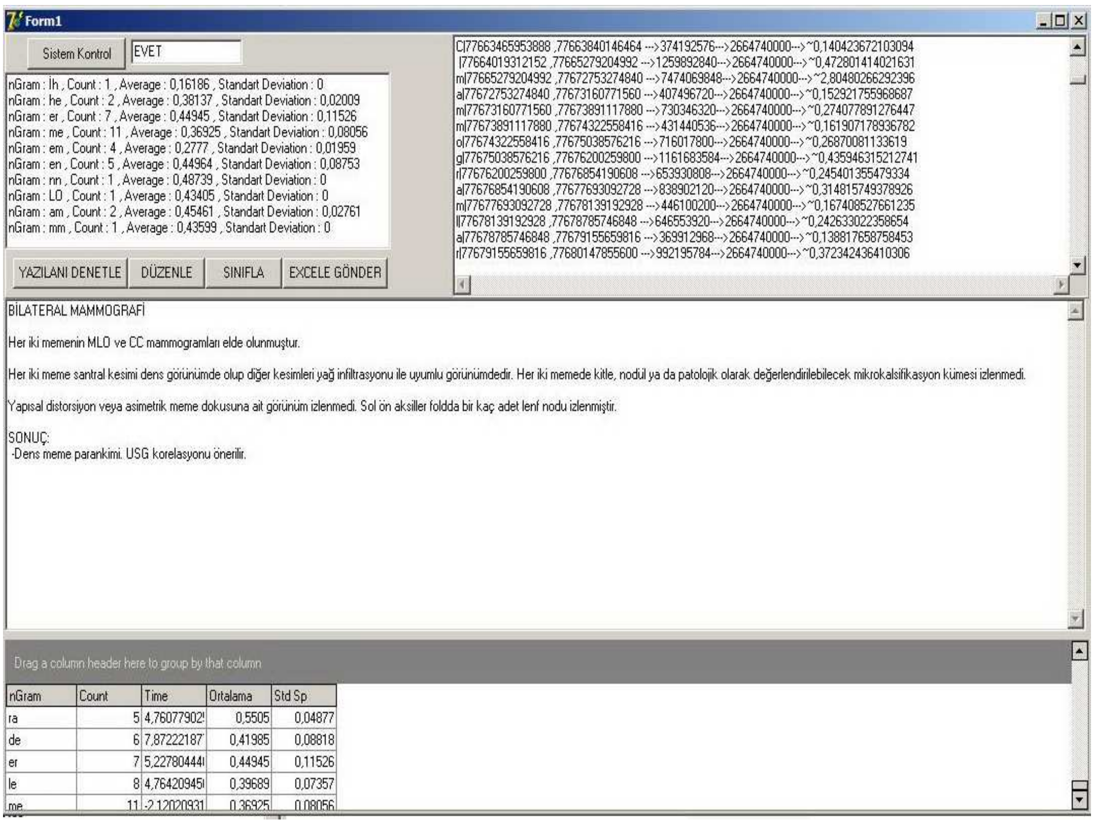

Figure III.3 Keystroke N-gram Test and Classification Sample

In order to test the prototype version of this keystroke based report system, these steps have been adequate (Figure III.3).

All the characters written to the editor except letters will be removed in order to create n-grams of the keystroke rhythms. For each letter, the time spent from previous key release to that letter's key release is calculated and appointed as 1-gram. In other words latency between two consecutive letters and the duration of the current letter provides the time recorded as the core 1-gram. Since 1-grams may not be informative, our software also collects 2-grams, 3-grams and 4-grams as illustrated in Figure III.4.

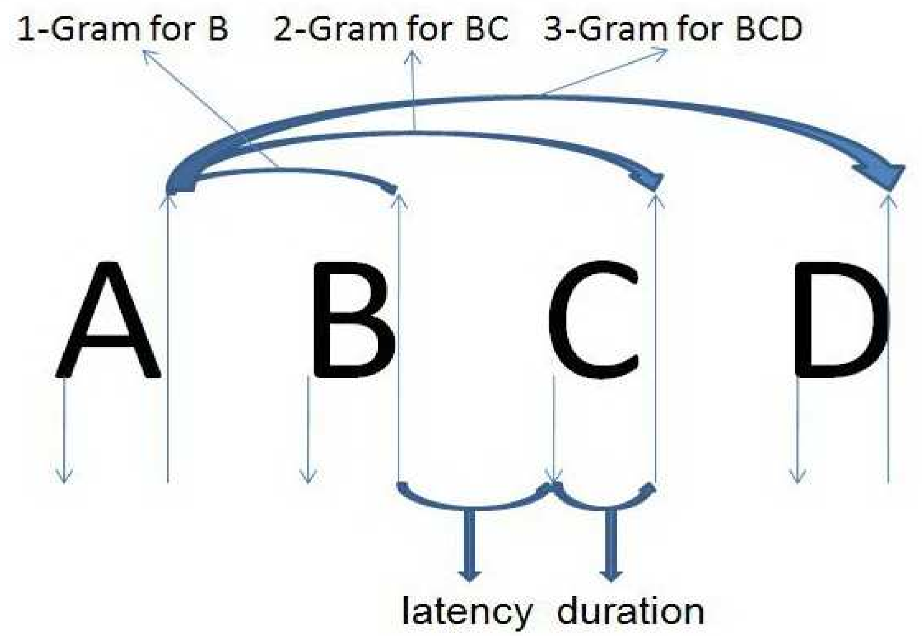

Figure III.4 N-gram Representation

The application's editor allows the user to pause or break for a while. To ensure that the pauses do not disrupt the keystroke measurements, a threshold (400 miliseconds) is defined inside the editor and the measurements exceeding this threshold are discarded. While creating the text, keystrokes which are deleted with backspace or delete keys are not included in resulting n-grams because of two reasons: First, deleted keystrokes do not exist in the final form of the written report; second, consecutive keystrokes are interrupted due to deletions, hence disrupting the operator's characteristic keystroke pattern.

Each operator has his/her own characteristic typing pattern because getting used to the location of each key varies among individuals. In addition, duration (the time between key-press and key-release) and latency (the time between key-release and following key-press) are different for each person. Also right-handed and left-handed writers may have different typing patterns. Besides this, data produced by each finger is not similar due to the anatomy and mechanics involved in pressing keys. According to the characteristics of the language used, keyboard layout will lead to different results. Frequently used syllables or endings may be written faster. It may even be possible to identify the author regardless of the language used, by examining only the spelling errors and corrections. If the operators are trained through a touchtyping course, or have been using text editor for years, their characteristic typing pattern become more stable.

In our prototype editor, we have collected many medical report samples. One sample medical report is as presented in Figure III.5 below.

BİLATERAL MAMMOGRAFİ

Her iki memenin MLO ve CC mammogramları elde olunmuştur.

Her iki meme santral kesimi dens görünümde olup diğer kesimleri yağ infiltrasyonu ile uyumlu görünümdedir. her iki memede kitle, nodül ya da patolojik olarak değerlendirilebilecek mikrokalsifikasyon kümesi izlenmedi.

Yapısal distorsiyon veya asimetrik meme dokusuna ait görünüm izlenmedi.

Sol ön aksiller foldda birkaç adet lenf nodu izlenmiştir.

SONUÇ : *Dens meme parankimi. USG korelasyonu önerilir.*

Figure III.5. Sample Report

For the medical report in Figure III.5, there exist 256 n-grams, composed of 97 2-grams, 91 3-grams and 68 4-grams as shown in Table III.1 below.

Table III.1 All Keystrokes with Average and Standard Deviation

N-GRAM	Count	Average	Std_Dev
ME	11	0,369250	0,080560
LE	8	0,396890	0,073570
ER	7	0,449450	0,115260
DE	6	0,419850	0,088180
EN	5	0,449640	0,087530
RA	5	0,550500	0,048770
EM	4	0,277700	0,019590
ÜM	4	0,317530	0,126960
EME	4	0,392320	0,042320
ML	3	0,468720	0,095600
OL	3	0,362650	0,018170
LU	3	0,496590	0,110330
NM	3	0,507190	0,070760
AL	3	0,413060	0,049350
ES	3	0,420800	0,022390
SM	3	0,536440	0,027500
AS	3	0,423030	0,046120
ON	3	0,331010	0,038930
ED	3	0,399120	0,022870
KA	3	0,411080	0,060110
MEM	3	0,569760	0,033510
MEME	3	0,693820	0,052130
RA	2	0,508620	0,002960
HE	2	0,381370	0,020090
AM	2	0,454610	0,027610
LA	2	0,339450	0,042000
AR	2	0,460680	0,050480
UN	2	0,375070	0,040410
SA	2	0,476790	0,083840
TR	2	0,508020	0,014540
KE	2	0,391270	0,065250
ÖR	2	0,487820	0,037640
MD	2	0,446260	0,033180
ĞE	2	0,405330	0,105600
YA	2	0,424360	0,066240
RL	2	0,606220	0,000000
OK	2	0,363770	0,083780
ET	2	0,353440	0,013000
HER	2	0,583380	0,017090
LAR	2	0,662730	0,091070
OLU	2	0,597500	0,011770

Table III.1 Continued

KES	2	0,657310	0,062340
ESM	2	0,662840	0,039010
ÜMD	2	0,592050	0,026000
MDE	2	0,632110	0,037480
ĞER	2	0,607340	0,111320
LER	2	0,528270	0,037470
MED	2	0,588830	0,026030
RLE	2	0,772100	0,007320
LEN	2	0,746150	0,018710
ENM	2	0,734660	0,010190
NME	2	0,702770	0,024600
KESM	2	0,912800	0,086890
ÜMDE	2	0,777900	0,030300
ENME	2	0,880240	0,011510
İL	1	0,326110	0,000000
LA	1	0,387140	0,000000
AT	1	0,494160	0,000000
TE	1	0,548690	0,000000
ER	1	0,424340	0,000000
AL	1	0,474030	0,000000
MA	1	0,375680	0,000000
AM	1	0,436160	0,000000
MM	1	0,462090	0,000000
MO	1	0,415990	0,000000
OG	1	0,505390	0,000000
GR	1	0,467820	0,000000
İH	1	0,161860	0,000000
NN	1	0,487390	0,000000
LO	1	0,434050	0,000000
MM	1	0,435990	0,000000
MO	1	0,430610	0,000000
RI	1	0,658290	0,000000
LD	1	0,253870	0,000000
MU	1	0,390000	0,000000
UR	1	0,253900	0,000000
RH	1	0,156010	0,000000
AN	1	0,358290	0,000000
NT	1	0,444510	0,000000
NS	1	0,366770	0,000000
RÜ	1	0,531370	0,000000
AĞ	1	0,436140	0,000000
NU	1	0,317510	0,000000

Table III.1 Continued

DR	1	0,441990	0,000000
TL	1	0,462150	0,000000
ÜL	1	0,360730	0,000000
LO	1	0,369210	0,000000
AK	1	0,410140	0,000000
EĞ	1	0,528080	0,000000
EC	1	0,337430	0,000000
CE	1	0,412510	0,000000
EK	1	0,352080	0,000000
LS	1	0,528720	0,000000
SF	1	0,771070	0,000000
FK	1	0,632050	0,000000
KÜ	1	0,487530	0,000000
IS	1	0,684500	0,000000
DS	1	0,523030	0,000000
OR	1	0,311730	0,000000
YO	1	0,331900	0,000000
RK	1	0,560150	0,000000
KU	1	0,358100	0,000000
US	1	0,401650	0,000000
SU	1	0,332010	0,000000
NA	1	0,401790	0,000000
ZL	1	0,407390	0,000000
LL	1	0,456020	0,000000
DD	1	0,398160	0,000000
DA	1	0,488130	0,000000
AD	1	0,545680	0,000000
NU	1	0,606690	0,000000
İLA	1	0,514120	0,000000
LAT	1	0,693290	0,000000
ATE	1	0,736710	0,000000
TER	1	0,730490	0,000000
ERA	1	0,748200	0,000000
RAL	1	0,655830	0,000000
MAM	1	0,678850	0,000000
AMM	1	0,595090	0,000000
MMO	1	0,719160	0,000000
MOG	1	0,664310	0,000000
OGR	1	0,724880	0,000000
GRA	1	0,759900	0,000000
İHE	1	0,557420	0,000000
MEN	1	0,724750	0,000000

Table III.1 Continued

ENN	1	0,628910	0,000000
AMM	1	0,588910	0,000000
MMO	1	0,704690	0,000000
RAM	1	0,727630	0,000000
AML	1	0,724860	0,000000
MLA	1	0,548860	0,000000
ARI	1	0,797100	0,000000
LDE	1	0,441430	0,000000
LUN	1	0,603160	0,000000
UNM	1	0,635070	0,000000
NMU	1	0,577600	0,000000
URH	1	0,262730	0,000000
RHE	1	0,508460	0,000000
SAN	1	0,545990	0,000000
ANT	1	0,649770	0,000000
NTR	1	0,646520	0,000000
TRA	1	0,779680	0,000000
RAL	1	0,655560	0,000000
ENS	1	0,557150	0,000000
ÖRÜ	1	0,716140	0,000000
SML	1	0,661520	0,000000
MLE	1	0,588860	0,000000
RAS	1	0,708040	0,000000
ONU	1	0,490640	0,000000
MLU	1	0,990390	0,000000
DED	1	0,741890	0,000000
EDR	1	0,623540	0,000000
EDE	1	0,548230	0,000000
TLE	1	0,629610	0,000000
OLO	1	0,516250	0,000000
ARA	1	0,704890	0,000000
RAK	1	0,684350	0,000000
DEĞ	1	0,721440	0,000000
EĞE	1	0,686780	0,000000
ERL	1	0,764930	0,000000
LEC	1	0,556640	0,000000
ECE	1	0,522050	0,000000
CEK	1	0,579970	0,000000
OKA	1	0,641020	0,000000
KAL	1	0,632280	0,000000
ALS	1	0,722190	0,000000
LSF	1	0,921190	0,000000

Table III.1 Continued

SFK	1	1,010650	0,000000
FKA	1	0,808380	0,000000
KAS	1	0,603950	0,000000
KÜM	1	0,883350	0,000000
ÜME	1	0,704420	0,000000
MES	1	0,843510	0,000000
ISA	1	0,924400	0,000000
SAL	1	0,762770	0,000000
YON	1	0,493580	0,000000
ASM	1	0,745300	0,000000
SME	1	0,687330	0,000000
MET	1	0,658070	0,000000
ETR	1	0,643640	0,000000
TRK	1	0,779500	0,000000
OKU	1	0,456200	0,000000
KUS	1	0,583530	0,000000
USU	1	0,508220	0,000000
SUN	1	0,560100	0,000000
UNA	1	0,508360	0,000000
ZLE	1	0,597950	0,000000
LLE	1	0,600290	0,000000
DDA	1	0,649610	0,000000
ADE	1	0,715900	0,000000
DET	1	0,585730	0,000000
İLAT	1	0,820270	0,000000
LATE	1	0,935840	0,000000
ATER	1	0,918500	0,000000
TERA	1	1,054350	0,000000
ERAL	1	0,898370	0,000000
MAMM	1	0,837770	0,000000
AMMO	1	0,852150	0,000000
MMOG	1	0,967480	0,000000
MOGR	1	0,883810	0,000000
OGRA	1	1,016960	0,000000
İHER	1	0,756420	0,000000
EMEN	1	0,837090	0,000000
MENN	1	0,822540	0,000000
AMMO	1	0,857610	0,000000
RAML	1	0,970260	0,000000
AMLA	1	0,863670	0,000000
MLAR	1	0,921200	0,000000
LARI	1	1,039740	0,000000

Table III.1 Continued

OLUN	1	0,773330	0,000000
LUNM	1	0,822740	0,000000
UNMU	1	0,805480	0,000000
URHE	1	0,615180	0,000000
RHER	1	0,713470	0,000000
SANT	1	0,837470	0,000000
ANTR	1	0,851780	0,000000
NTRA	1	0,932710	0,000000
TRAL	1	0,947040	0,000000
ESML	1	0,785460	0,000000
SMLE	1	0,857810	0,000000
MLER	1	0,796690	0,000000
MDED	1	0,842960	0,000000
DEDR	1	0,935560	0,000000
EMED	1	0,660720	0,000000
MEDE	1	0,735920	0,000000
LARA	1	0,866350	0,000000
ARAK	1	0,820340	0,000000
DEĞE	1	0,880140	0,000000
EĞER	1	0,894500	0,000000
ĞERL	1	1,117170	0,000000
ERLE	1	0,923480	0,000000
RLEN	1	0,935160	0,000000
LECE	1	0,741260	0,000000
ECEK	1	0,689500	0,000000
OKAL	1	0,791130	0,000000
KALS	1	1,010890	0,000000
ALSF	1	1,114660	0,000000
LSFK	1	1,160770	0,000000
SFKA	1	1,186990	0,000000
FKAS	1	0,996420	0,000000
KÜME	1	1,050690	0,000000
ÜMES	1	0,984770	0,000000
ISAL	1	1,126550	0,000000
ASME	1	0,866390	0,000000
SMET	1	0,906680	0,000000
METR	1	0,961280	0,000000
ETRK	1	0,900590	0,000000
OKUS	1	0,681630	0,000000
KUSU	1	0,690110	0,000000
USUN	1	0,736310	0,000000
SUNA	1	0,733800	0,000000

Table III.1 Continued

ZLEN	1	0,929770	0,000000
LENM	1	0,987340	0,000000
NMED	1	0,946680	0,000000
LLER	1	0,782160	0,000000
ADET	1	0,912130	0,000000

This concludes our implementation of a keystroke based medical report generation prototype. Results obtained from this system is discussed in the next section.

CHAPTER IV

RESULTS and DISCUSSION

In order to evaluate the consequences of the proposed system usage, a three-stage study is conducted. In the preparation phase many experimental texts are used. Some of them have medical content, and some are plain text. By this way the operators are accustomed to the editor. In the data collection phase, the operators begin to write the medical reports in the prototype editor. In the evaluation phase, the operators are classified according to their keystroke typing patterns.

All studies are performed in the native environment of the users, where they are accustomed to work. In addition to this, the medical reports are typed in our editor, and transferred to the information system after the approval of the radiologist. Hence using our system does not generate additional burden to the reporters.

Table IV.1 Properties of Operators Involved in the Experiment

Subject	Age	Preferred Hand	Education (years, univ: 15, college: 13, high school: 11)	Job Experience for Medical Reports (Years)	Gender	Layout and Typing Style	Number of Reports
sa	35	Right	13	11	Female	Turkish F-Touchtyping	12
sy	36	Right	13	12	Female	Turkish F-Touchtyping	12
ec	31	Right	13	9	Female	Turkish F-Touchtyping	12
be	33	Right	15	8	Female	Turkish F-Touchtyping	12
nc	48	Right	13	16	Female	Turkish F-Touchtyping	12

The expertise of operators admitted to our study are equivalent to each other. This situation facilitates the validation of our system. Because the users are all experienced in this domain and familiar to the medical terminology used in the reports. All the medical report writers are female and right handed. On the other hand they all learned touchtyping in the college and they are using the keyboard type of “Turkish F” for years. (Table IV.1)

We collect sixty randomly selected medical reports with the assistance of five experienced medical report writers. The selected medical reports mainly contain textual interpretation and categorized as: computerized tomography, breast thermography, ultrasound and interventional radiology reports. (Table IV.2)

Table IV.2 Types and Distribution of Medical Reports and the Operators

	Number of Reports Enrolled for each operator, be, ec, nc, sa, sy					
Medical Report Type	be	ec	nc	sa	sy	total
breast thermography	3	3	3	3	3	15
ultrasound	3	3	3	3	3	15
computerized tomography	3	3	3	3	3	15
interventional radiology	3	3	3	3	3	15
total	12	12	12	12	12	60

Following the accomplishment of second stage, the obtained data is transferred out to be classified and analysed according to subjects. 12 seperate data files are created for each subject. Each data file contains a structured table consisting of a certain number of n-grams, obtained from the associated medical report. The tables have four columns; named as “n-gram”, “count”, “mean” and “standard deviation” respectively. The ‘n-gram’ column contains the letters of the n-gram; the ‘count’ column contains the number of repetitions, the ‘mean’ column contains the average of time measurements for the belonging n-gram, the ‘standad deviation’ column also contains the calculated standard deviation for the repeated n-gram.

In this study, only 2-grams, 3-grams and 4-grams are examined. This data is presented in Table III.1 above. For further analysis, n-grams are subjected to an extraction criteria which is based on the ratio of standard deviation and time spent similar to a criteria used in[34]. According to this, a two-step procedure is applied: In the first step, outliers are removed. Outliers occur due to involuntary and unexpected delays during the typing of the medical report. Disposing of the outlier process is important for evaluation[23]. In the second step, average and standard deviation are re-calculated for each n-gram. Sometimes this process can be repeated more than once, until all outliers are removed. As a result, from our sample reports, 41 n-grams (comprising of 35 2-grams and 6 3-grams from within approximately 8000–10000 n-grams) are extracted as the characteristic features for further analysis. The total number of 4-grams are not enough to pass this extraction phase. After that, the n-grams which are common to all the reports from multiple subjects are identified as presented in Appendix A. A graphical representation of extracted n-grams and the corresponding timing values for each operator is presented in Figure IV.1. Differentiation of the timings between users can be seen easily from this chart.

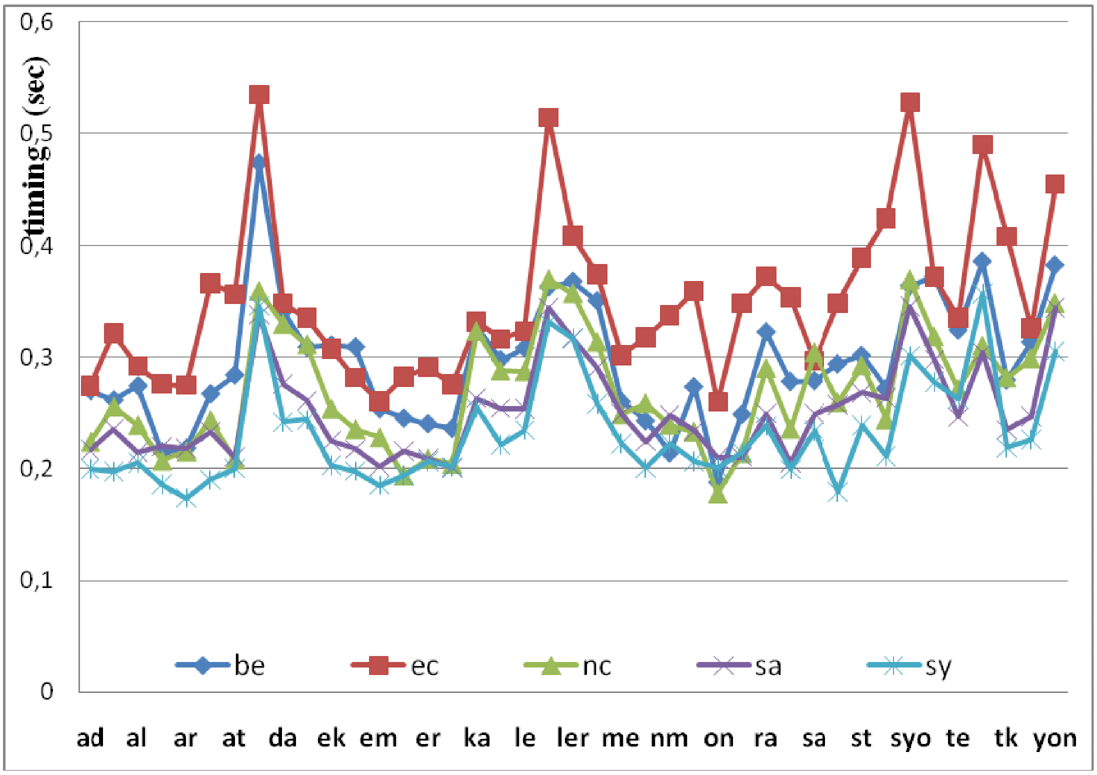

Figure IV.1 Graphical Representation of Featuring n-grams for 5 Operators

The following chart indicates the limit values for a single user 'be' obtained by plotting the minimum and maximum timings of extracted n-grams.(Figure IV.2) Personal keystroke typing boundaries become characteristic after a certain amount of reports enrolled. These charcteristic features can be used in any application to calculate various similarity measures for users.

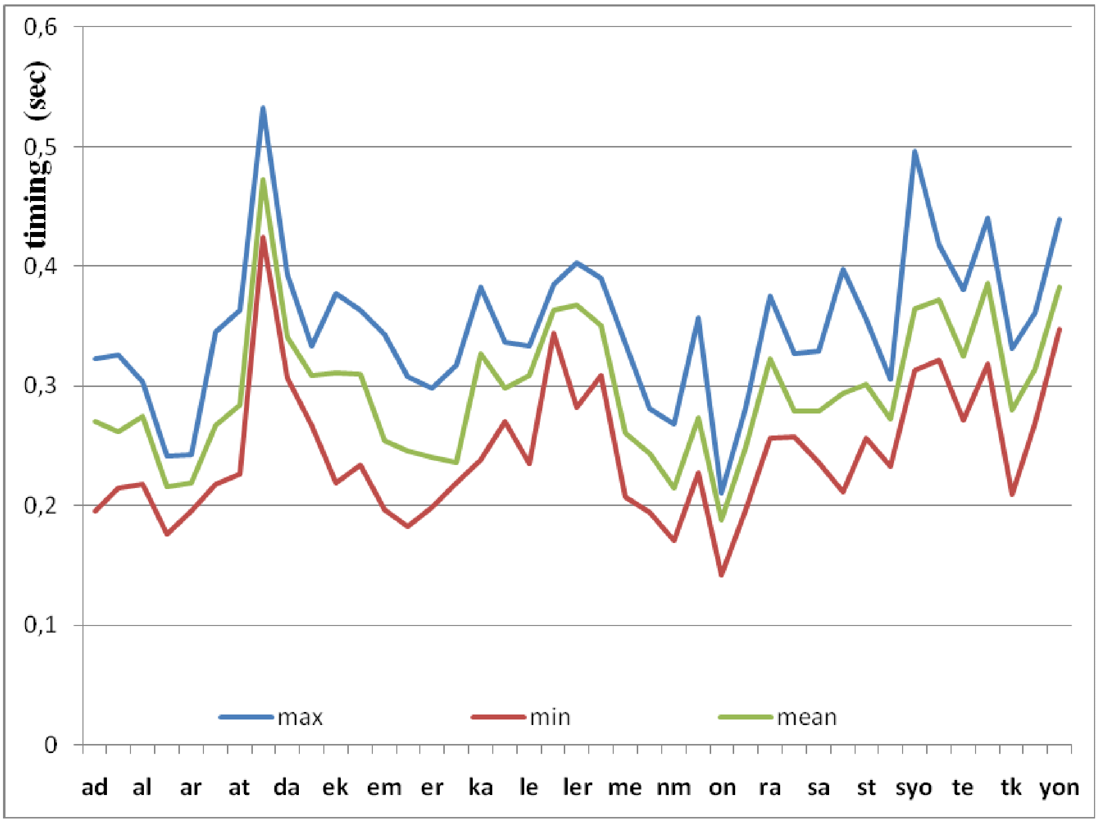

Figure IV.2 Graphical Representation of Featuring N-grams for 'be'

IV.1. Analytical methods to classify medical reporters

Several analytical methods can be employed to demonstrate the identifiability of operators according to the extracted n-grams. We used the Nearest Neighbour algorithm to calculate the similarity between authorized medical reporters and the operator typing the current report. Typed medical reports are divided into two groups for each user. The first group consisting of 9 reports provides the reference values of the average of n-gram timings. These reports are used to extract typical typing characteristics of each reporter as shown in the middle plot in figure IV.2. The reports in the second group are used as test samples. For each one of the 3 reports in the second group, instantaneous n-grams are generated and compared with the reference n-grams of each authorized medical reporter according to the nearest neighbor calculation. Despite some other close candidates, each report's n-gram keystroke characteristics identified the original typist as the nearest neighbour. We used the following formula to identify the nearest neighbor:

$$List_j = SQRT\left[\sum_{i=1}^{N}\left(mean\_n\text{-}gram_{ij} - report\_n\text{-}gram_i\right)^2\right]$$

Where $mean\_n\text{-}gram_{ij}$ is the ith n-gram of the jth medical reporter, $report\_ngram_i$ is the ith n-gram from the given medical report, i is an index for n-grams such that i=1,..,41, j is an index identifying the 5 reporters. Dist j shows the difference of typing characteristics of reporter J from that of the current report. Minimum of this distance identifies the index of the nearest neighbor as follows:

$$J' = \min_j (Dist_j)$$ J' is nearest neighbor

In table below(Table IV.3), the identified nearest neighbors from the test reports are shown.

Table IV.3 Calculated Distances of 3 Reports for Each Operator According to Nearest Neighbor Algorithm

users	be	ec	nc	sa	sy
be	0,21703	0,54561	0,27379	0,35847	0,42407
be	0,19663	0,52978	0,31798	0,43524	0,49999
be	0,19595	0,53150	0,31481	0,43189	0,49650
ec	0,55666	0,33079	0,66753	0,76289	0,83789
ec	0,50789	0,33100	0,62218	0,71547	0,81295
ec	0,49543	0,38880	0,59942	0,69414	0,76288
nc	0,34871	0,69997	0,14811	0,32504	0,37715
nc	0,30809	0,65557	0,19105	0,29941	0,38892
nc	0,29407	0,62393	0,22523	0,31817	0,36851
sa	0,36031	0,68873	0,23908	0,24311	0,33934
sa	0,34454	0,66993	0,24008	0,21565	0,27963
sa	0,36358	0,71199	0,25000	0,19174	0,25641
sy	0,54085	0,90913	0,42507	0,32314	0,18166
sy	0,49750	0,80572	0,46839	0,40763	0,14762
sy	0,49640	0,87522	0,35431	0,25015	0,21098

In order to decide whether all extracted n-grams are representative for identifying the medical reporters, we performed Factor Analysis and Reliability Statistics over the 41 extracted features using SPSS. However, in order to perform factor analysis, some criteria as emasured by KMO measure and Bartlett test should be present in the data. The analysis results of these test are shown in table IV.4. KMO test for convenience of the sample size is 0,708. The Kaiser-Meyer-Olkin measure (KMO) of sampling adequacy tests whether the partial correlations among variables are small. With a value of 0,708, our data satisfies the goodness criterion. In addition, Bartlett test must also be significant. Bartlett's test of sphericity tests whether the correlation matrix is an identity matrix, which would indicate that the factor model is inappropriate. Our data shows a significant value for this test (p=0.000), hence it is appropriate to perform Factor Analysis.

Table IV.4. KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0,708
Bartlett's Test of Sphericity	Approx. Chi-Square	2.524,711
	df	820
	Sig.	0,000

Factor analysis shows the prominant components of the 41 extracted features for identifying individual medical reporters. These components might be composed of multiple features with idividual weighting factors. The result of the factor analysis on our extracted n-grams are presented in table IV.5 below in terms of eigenvalues[17] before and after factor extraction and after rotation. In addition, in figure IV.3, the scree plot is shown. According to this figure, there are currently 8-9 factors with eigenvalues greater than (or equal to) 1. This is in agreement with table IV.5. According to this table, the first factor explains approximately 48% of the variance. As seen from here, 9 factors explain approximately 80% of the total variance. The right part of the table, Rotation, shows the relative importance of the factors. Due to this, factor 1 contribution decreases from 48% to 21% after rotation.

[17] Eigenvalues are a special set of scalars associated with a linear system of equations, as the value of the eigenvalue increases, the contribution of the correspoding factor increases

Table IV.5. Total Variance Explained (Extraction Method: PCA)

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% Var	Cmltive %	Total	% Var	Cmltive %	Total	% Var	Cmltive %
1	19,643	47,910	47,910	19,643	47,910	47,910	8,626	21,039	21,039
2	2,814	6,864	54,774	2,814	6,864	54,774	4,967	12,115	33,154
3	1,950	4,757	59,531	1,950	4,757	59,531	3,775	9,206	42,360
4	1,828	4,458	63,989	1,828	4,458	63,989	3,633	8,862	51,222
5	1,512	3,687	67,675	1,512	3,687	67,675	2,911	7,100	58,321
6	1,360	3,318	70,993	1,360	3,318	70,993	2,804	6,840	65,161
7	1,134	2,765	73,759	1,134	2,765	73,759	2,470	6,024	71,185
8	1,106	2,699	76,457	1,106	2,699	76,457	1,729	4,217	75,402
9	1,052	2,565	79,023	1,052	2,565	79,023	1,484	3,621	79,023
10	0,918	2,239	81,262						
11	0,776	1,893	83,155						
12	0,712	1,737	84,892						
13	0,597	1,457	86,349						
14	0,592	1,444	87,793						
15	0,555	1,354	89,147						
16	0,463	1,128	90,275						
17	0,421	1,027	91,302						
18	0,387	0,944	92,245						
19	0,372	0,908	93,154						
20	0,351	0,855	94,009						
21	0,326	0,795	94,804						
22	0,283	0,690	95,493						
23	0,246	0,600	96,093						
24	0,219	0,534	96,627						
25	0,210	0,511	97,138						
26	0,190	0,464	97,602						
27	0,169	0,412	98,015						
28	0,146	0,356	98,370						
29	0,125	0,306	98,676						
30	0,115	0,281	98,957						
31	0,104	0,253	99,210						
32	0,082	0,199	99,409						
33	0,052	0,127	99,536						
34	0,046	0,113	99,649						
35	0,038	0,092	99,741						
36	0,033	0,080	99,820						
37	0,024	0,058	99,879						
38	0,020	0,048	99,927						
39	0,016	0,040	99,967						
40	0,011	0,026	99,993						
41	0,003	0,007	100,000						

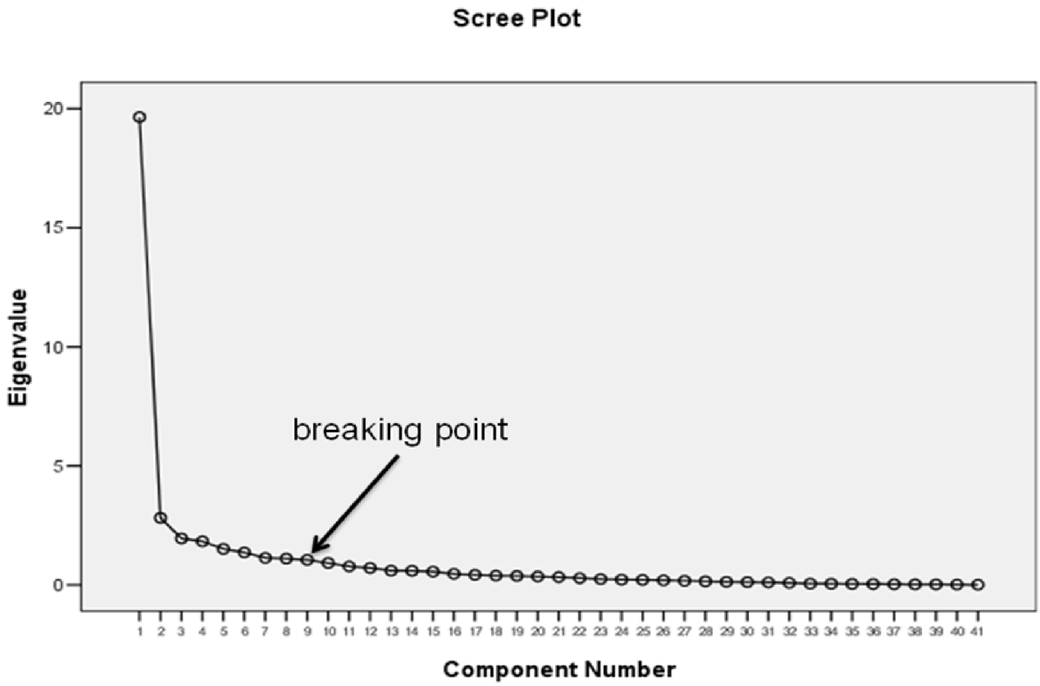

Figure IV.3 Scree Plot

SPSS decided 9 components (Table IV.6). Principal Component Analysis[18] is used for extraction. Most variables are related to the first factor in table IV.6, they have become more balanced after Varimax Rotation with Kaiser Normalization[19] (Table IV.7).

[18] **Principal component analysis (PCA)** involves a mathematical procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components.

[19] **Varimax with Kaiser Normalization**: a change of coordinates used in principal component analysis and factor analysis that maximizes the sum of the variances of the squared loadings.

Table IV.6 Component Matrix

	Component								
	1	2	3	4	5	6	7	8	9
rl	0,843		0,109	-0,128		-0,107			-0,129
as	0,833			0,102		-0,208	0,174		
sy	0,830	-0,310	-0,187			-0,205	0,127	0,154	
yon	0,823	-0,170			-0,110	-0,190	0,129		
le	0,819	0,308	0,152						0,224
syo	0,802	-0,239	-0,240			-0,217	0,201		
ol	0,798	-0,236		-0,305	0,188				
er	0,784	-0,288	0,203	-0,146			-0,304		
nd	0,779	-0,122		-0,229	0,235			-0,173	
at	0,778	-0,210	0,298			-0,105	-0,309	-0,131	-0,188
len	0,762	-0,235	-0,269					-0,112	
st	0,754			-0,203	0,101			0,127	0,122
en	0,754	-0,129						0,442	
ma	0,749	0,150	0,310	0,332					-0,120
ra	0,743	0,174	-0,303		-0,261	0,148			
yo	0,736	0,276			-0,121	-0,213	0,247		0,170
al	0,734	0,213	-0,130	0,230	-0,210			-0,355	-0,238
ate	0,711	-0,246	0,274	-0,319	-0,166	-0,118	-0,141	0,106	-0,111
da	0,707	0,430	0,138	-0,150	0,165	-0,225	0,102		-0,136
an	0,705	-0,177	-0,269	0,419			0,159		-0,123
ar	0,704			0,216	-0,101		0,170	-0,177	
la	0,703	0,320	0,210	-0,141				-0,185	
ter	0,692	-0,335		-0,301	-0,301	0,158			0,123
de	0,687	0,234	-0,184		0,116	-0,232		-0,160	0,333
tk	0,684	-0,294	-0,158	-0,313			0,107	-0,300	
or	0,676	-0,118	-0,376	-0,208		0,320		-0,152	-0,137
el	0,674	0,339		0,195	-0,455		0,182		
ler	0,669		0,219	0,191		0,392		-0,243	0,303
ta	0,668	0,183	0,126	0,154	-0,221		-0,247	-0,139	0,150
te	0,667		0,154	-0,318	-0,379	0,269		0,116	0,123
et	0,641	-0,138	0,108	0,247	0,121		-0,525		
em	0,602	0,322	-0,203		-0,102		-0,115	0,206	0,373
so	0,576	0,318				0,258		0,341	-0,456
ak	0,572	0,307	-0,269	0,176		0,210	-0,331		-0,241
ad	0,565	0,104	0,211		0,187	0,276	0,248	-0,181	
me	0,542	0,353	-0,437		0,192	0,213		0,163	0,123
ka	0,542	0,334	0,212	0,157	0,457	-0,236			
on	0,401	-0,558		0,445		0,164	0,130	0,194	0,134
nm	0,451	-0,521		0,465	0,209	0,314		-0,132	0,119
ek	0,411		0,686			0,210	0,229	0,127	
sa	0,395	0,353		-0,241	0,538	0,345		0,156	0,106

Table IV.7 Rotated Component Matrix

	Component								
	1	2	3	4	5	6	7	8	9
tk	0,818		0,224	0,110			0,115	0,111	-0,141
ol	0,788	0,142		0,233	0,123	0,217	0,228		
nd	0,761	0,186	0,161	0,236		0,153	0,222		
sy	0,722	0,358	0,128		0,371	0,147			0,288
syo	0,704	0,400	0,161		0,346				0,225
len	0,684	0,270	0,267		0,234	0,172	0,148	0,114	
or	0,624		0,435		0,151		0,391	0,221	
ate	0,619	0,147		0,299		0,431		0,311	0,201
rl	0,614	0,279	0,257	0,305		0,354			0,151
yon	0,611	0,453	0,187	0,159	0,252	0,146		0,116	0,215
ter	0,592	0,183	0,116	0,117	0,164	0,191		0,586	
as	0,575	0,435	0,226	0,232	0,244	0,135		-0,104	0,289
st	0,571	0,349		0,249	0,123	0,249	0,213	0,149	0,116
em	0,133	0,695	0,169			0,163	0,329	0,209	
yo	0,347	0,681	0,199	0,344					0,171
de	0,354	0,666	0,216			0,213	0,247	-0,114	-0,182
le	0,277	0,651	0,224	0,401		0,289	0,206		
ta	0,177	0,500	0,374	0,236		0,422		0,118	
al	0,319	0,266	0,791	0,196		0,211			
el	0,157	0,517	0,623	0,256				0,212	0,265
ak	0,115	0,161	0,583			0,343	0,446		0,207
ra	0,367	0,394	0,567			0,111	0,243	0,286	
ar	0,401	0,378	0,405	0,235	0,304				
ek				0,810	0,158	0,123		0,152	0,166
ad	0,345		0,191	0,560	0,136		0,292		
ma	0,220	0,318	0,384	0,545	0,276	0,307		-0,126	0,165
la	0,372	0,354	0,352	0,513	-0,162	0,145	0,112		
ler	0,152	0,289	0,313	0,506	0,392	0,187	0,263	0,204	-0,286
da	0,413	0,432	0,232	0,445	-0,243	0,163	0,236	-0,225	0,181
nm	0,248		0,143	0,122	0,841	0,181			-0,139
on	0,210			0,113	0,824	0,107			0,148
an	0,421	0,338	0,444		0,507				0,267
et	0,242	0,222	0,178		0,308	0,738	0,110		0,110
at	0,570		0,204	0,330	0,128	0,635			
er	0,552	0,216		0,228	0,206	0,563		0,306	
sa	0,168	0,117		0,267			0,817		
me	0,213	0,375	0,276				0,678		0,114
te	0,402	0,241	0,166	0,366		0,111		0,639	
ka	0,199	0,382		0,390		0,300	0,272	-0,483	
so	0,124		0,448	0,347		0,146	0,329		0,597
en	0,418	0,281	0,111	0,125	0,309	0,296	0,245	0,194	0,484

Table IV.9 represents the correlation between the remaining features if any feature is removed. Cronbach Alpha[20] value must be at least 0,7. The result must be close to

[20] It is commonly used as a measure of the internal consistency or reliability of a psychometric test score.

Cronbach's Alpha value of 0,968 (Table IV.8). The value increases to 0,969 after extraction of ‘ek’, and decreases for all the other extractions, but it is not remarkable.

Table IV.8. Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
0,968	0,971	41

Table IV.9. Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cr's Alpha if Item Deleted
ad	11.302,01807	2.929.354,358	0,544	0,866	0,967
ak	11.290,20440	2.910.345,708	0,546	0,880	0,967
al	11.298,07251	2.911.063,610	0,708	0,943	0,967
an	11.325,02435	2.938.582,922	0,687	0,947	0,967
ar	11.326,75627	2.938.111,239	0,675	0,833	0,967
as	11.290,52994	2.858.246,555	0,813	0,956	0,966
at	11.292,83164	2.825.812,077	0,763	0,965	0,966
ate	11.136,19460	2.808.294,935	0,694	0,900	0,967
da	11.240,14008	2.919.505,518	0,680	0,916	0,967
de	11.256,35420	2.937.280,974	0,654	0,926	0,967
ek	11.269,22657	2.902.790,414	0,392	0,836	0,969
el	11.290,08546	2.891.688,731	0,641	0,940	0,967
em	11.318,53624	2.939.600,759	0,566	0,869	0,967
en	11.320,30568	2.926.606,150	0,744	0,930	0,967
er	11.315,47317	2.927.660,276	0,774	0,955	0,967
et	11.322,96610	2.939.361,265	0,620	0,880	0,967
ka	11.246,78690	2.942.938,652	0,508	0,788	0,967
la	11.267,42101	2.926.136,613	0,684	0,833	0,967
le	11.265,22735	2.925.971,367	0,793	0,962	0,967
len	11.162,22663	2.829.578,189	0,739	0,967	0,966
ler	11.189,75724	2.925.360,593	0,654	0,882	0,967
ma	11.232,18257	2.891.594,865	0,730	0,936	0,967
me	11.286,33942	2.947.667,580	0,513	0,878	0,967
nd	11.297,82485	2.912.916,127	0,769	0,966	0,967
nm	11.288,40330	2.933.808,734	0,436	0,914	0,968
ol	11.285,41716	2.859.247,085	0,793	0,947	0,966
on	11.339,41340	2.972.454,538	0,396	0,913	0,968
or	11.298,06975	2.880.397,743	0,666	0,946	0,967
ra	11.253,73760	2.874.733,208	0,714	0,814	0,967
rl	11.289,72960	2.843.449,959	0,834	0,948	0,966
sa	11.270,85473	2.969.089,144	0,380	0,868	0,968
so	11.265,33846	2.859.137,955	0,553	0,933	0,968
st	11.255,16264	2.841.496,383	0,733	0,912	0,966
sy	11.261,99910	2.801.149,117	0,816	0,993	0,966
syo	11.165,13124	2.791.107,871	0,785	0,988	0,966
ta	11.221,23717	2.916.635,108	0,636	0,860	0,967
te	11.257,00278	2.922.607,262	0,659	0,907	0,967
ter	11.177,93260	2.828.101,635	0,677	0,895	0,967
tk	11.262,66846	2.832.507,970	0,667	0,911	0,967
yo	11.264,42646	2.919.543,349	0,708	0,945	0,967
yon	11.179,80640	2.869.423,158	0,809	0,964	0,966

Further details regarding the factor analysis is provided in correlation matrix which is attached in Appendix B and in reproduced correlations and residuals which is attached in Appendix C. These measures are based on the model, not the observed data. Residuals indicate the difference between the correlation matix (Appendix B) and reproduced correlations.

IV.2 A Framework Proposal for Authentication of Medical Reports

Overall, from these results we can infer that, the chosen operators produce reliably distinct keystroke patterns as they type these 4 kinds of medical reports. The operators participating in this study are equivalent in experience and expertise. The consistency of obtained results are encouraging for making proposals about a second level identity verification framework which is illustrated in figure V.1.

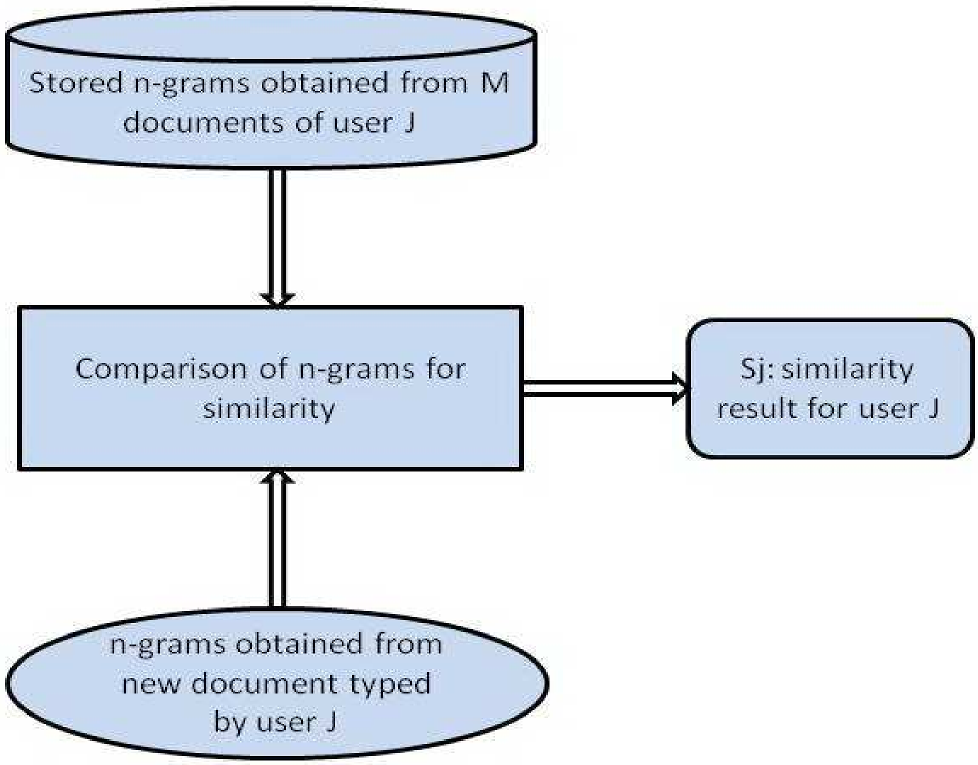

Figure IV.1 Proposed Framework for Similarity

In this figure, the first input to the system consists of stored n-grams obtained from M documents of user J. These n-grams identify user J, because they are composed through computations that involve several keystroke timings of user J. More specifically, these computations involve: gathering the keystroke data bundle, discarding the excess data, re-calculating n-grams, performing feature extraction and finally storing the extracted features in the database. The second input to the system is defined as n-grams obtained from the current document or medical report typed by the current operator. This input is generated after two steps: gathering raw keystroke timings from the current report, and obtaining instantaneous n-grams.

In order to achieve the proposed second level identity verification, n-grams from these two inputs should be compared as illustrated in the above figure. At this stage, stored features of the user J are called from the database and compared to the instant n-grams coming from the current report with pattern recognition algorithms. Finally a “similarity” value is generated for user J (ie. medical operator of the given report). In the event that user J is not the person typing the report which is claimed access into the hospital information system, this similarity value is expected to be low.

Extracted n-grams are combined together according to the prominent components, which are issued as a result of factor analysis summarized beforehands. Euclidean distance measurement can be performed between instantaneous and reference features[21] for obtaining similarity measures. Furthermore, the calculated similarity value may be stored in the database and embedded in the report as a protective or informative measure such as watermark.

IV.3 Limitations of the Study

Keystroke Dynamics is categorized under the behavioral class of biometrics which is based on statistical patterns. Behavioral biometrics is not as reliable as physiological biometrics for authentication or individual measurements. Rather than definitive results such as “TRUE” or “FALSE”, they produce confidence intervals for comparisons.

[21] Instantaneous features come from the current report, while reference features come from the database

This work is developed on the Windows operating system without the use of hardware interrupts or any other low level keystroke timing mechanism. In this system, the keystroke timings may be subject to some errors due to unrelated background processes running simultaneously. For timing each keystroke, a mechanism utilizing operating system kernel modification might produce more reliable results and n-grams with less variance. However this approach will be studied later.

The system has a high percentage of discrimination for 5 operators. After increasing the number of users, or changing users, the sensitivity and specificity of the system should be tested again.

CHAPTER V

CONCLUSION

Information security is not only important for the confidentiality of the patient but also for the experts working at the hospital. Medical operators are among these experts. Users of the hospital information system must first be authenticated by the system to create medical reports. When user authentication is performed with a username and a password, there is a single-level access control. In order to provide legal and personal protection for both parties, a second level identity verification framework which is proposed in section IV.2 is needed. The proposed framework within this thesis collects all the information to distinguish between users who are authorized to create medical report through keystroke dynamics.

Keystroke Dynamics has the advantage of providing security control through continuous measurements at any time, instead of just at the beginning. Therefore identity can be verified continually. This allows for checking against any personal information security breach in contravention of the principles summarized in section II. Although the confidence interval of Keystroke Dynamics is low it has a deterrent effect. For real time applications that involve typing such as remote consultance with medical staff, the proposed framework can be used as a continual identity verification application with additional components as suggested in [24].

In order to show the feasibility of the proposed system, we implemented a testbed and used the nearest neighbor measure to identify the medical reporter who is typing the current report. However, more complicated applications may be developed using various statistical techniques, artificial intelligence or pattern recognition methodology such as neural networks. These approaches will also succeed with high accuracy, because the keystrokes are as characteristic as signature or handwriting.

In future work, the benefits and user acceptance of the framework based on keystroke dynamics and its integration into hospital information system should be established and evaluated. Even more efficient methods for all the free-text editors could be established by using the most common letter combinations in Turkish [44], but this is currently considered as out of the scope of this thesis.

Applications embodying any kind of electronic signature will be inevitable for health records in the near future. Alternatively, we proposed a new security enhancement methodology for improving the reliability of medical reports generated within the network of the hospital. This methodology in contingent upon an existing authentication mechanism, such as username/password. As shown by our test data, the proposed second-level reporter verification mechanism based on keystroke dynamics is feasible.

REFERENCES

1. Jiang, C. H. (2006). Keystroke Statistical Learning Model for Web Authentication, Unpublished master's thesis, Computer Science and Information Engineering in National Chiao Tung University, Hsinchu, Taiwan.

2. Hu J., Gingrich D., Sentosa A., "A k-Nearest Neighbor Approach for User Authentication through Biometric Keystroke Dynamics ," I*EEE International Conference on Communication (ICC)*, China, May 2008

3. Gaines R., Lisowski W., Press S., and Shapiro N.. Authentication by keystroke timing: Some preliminary results. Technical Report Rand report R-256-NSF, Rand corporation, 1980.

4. Kunitz and Associates, Inc. *Final Report of the Task Force on the Privacy of Private-Sector Health Records.* Contract HHS-100-91-0036. Rockville: KAL September 1995.

5. Baykal, N. & Musoğlu, E. (June, 2000). Introduction to Medical Informatics: Lecture Notes, Turkey .

6. Matyas, V. Jr., Riha, Z., "Toward Reliable User Authentication through Biometrics", IEEE Security & Privacy Magazine, May/June 2003, Vol. 1 No. 3, pp 45-

7. Brandner R., van der Haak M. and Hartmann M. *et al.*, Electronic signature for medical documents—integration and evaluation of a public key infrastructure in hospitals, *Methods Inf. Med.* 41 (2002), pp. 321–330.

8. Pharow P, Blobel B. Electronic signatures for long-lasting storage purposes in electronic archives. Int J Med Inform. 2005 Mar;74(2-4):279-87. PMID: 15694634

9. Rivest, R.; A. Shamir; L. Adleman (1978). "A Method for Obtaining Digital Signatures and Public-Key Cryptosystems". *Communications of the ACM* 21 (2): 120–126.

10. Dr. David W. Kravitz

11. "Digital signature." Computer Desktop Encyclopedia. Computer Language Company Inc., 2010. *Answers.com* 15 May. 2010. http://www.answers.com/topic/digital-signature

12. Gokcay (Unlu) D. 1991, "User Identification Through Neural Network Algorithms", Unpublished Master's Thesis in EEE, METU.

13. Conn P., Parodi J.H. and Taylor M., "The Place of Biometrics in a User authentication Taxonomy," Proc. 13th Nat'l Computer Security Conf., Nat'l Inst. Standards and Technology/ Nat'l Computer Security Center, Gaithersburg, Md.,1990.

14.“Bertillonage” Individual Biometrics E-Court Conference 2002 last access 5 June 2010 http://ctl.ncsc.dni.us/biomet_web/BMBody.html

15.“Biometrics History ” last updated 7 August 2006 last access 4 June 2010 http://www.biometrics.gov/Documents/BioHistory.pdf

16. Weber, K.: The Next Step: Privacy Invasions by Biometrics and ICT Implants. Ubiquity. An ACM IT Magazine and Forum, 7(45), 2006.

17. Clarke N. L. and Furnell S. M., "Authentication of users on mobile telephones--a survey of attitudes and practices," *Computers and Security*, vol. 24, no. 7, pp. 519-527, 2005.

18. Jain, AK, Ross, A., Prabhakar, S.: An Introduction to Biometric Recognition. IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Image- and Video-Based Biometrics 14 (2004) 4–20

19. Jain, A, Bolle, R., Pankanti, S., Introduction to biometrics, in: Jain, A., Bolle, R. and Pankanti, S.:Biometrics: Personal Identification in Networked Society, Kluwer Academic Publishers, 1999, pp.1-42.

20. Bhattacharyya D., Ranjan R., Alisherov A.F., Choi M “Biometric Authentication A Review” International Journal of u- and e- Service, Science and Technology Vol. 2, No. 3, September, 2009

21. Olzak T., “Keystroke Dynamics: Low Impact Biometric Verification”, September 2006, http://www.infosecwriters.com/text_resources/pdf/Keystroke_TOlzak.pdf

22. Jain A., Ross A., and Pankanti S., \Biometrics: A tool for information security," *IEEE Transactions on Information Forensics and Security* 1, pp. 125{143, 2006.

23. Villani, M., Tappert, C. M., Ngo,G., Simone, J.,St. Fort,H., and Cha, S., “Keystroke Biometric Recognition Studies on Long-Text Input Under Ideal and Application-Oriented Conditions”, Proc. Student/Faculty Research Day, May 2006.

24. Clarke, N., Furnell, S., Lines, B., and Reynolds, P. 2003. Using keystroke analysis as a mechanism for subscriber authentication on mobile handsets. In Proc. 18th Conf. on Information Security. (IFIP/SEC 2003). Kluwer, Athens, Greece.

25. Checco J C., “Keystroke Dynamics and Corporate Security”, http://www.checco.com/about/john.checco/publications/2003%20keystroke%20Biometrics%20Intro.pdf

26. Joyce R., Gupta G., Identity authentication based on keystroke latencies, Communications of the ACM, v.33 n.2, p.168-176, Feb. 1990.

27. Monrose F. and Rubin A. D.. 1997. Authentication via keystroke dynamics. In *4th ACM Conference on Computer and Communcations Security* (April 1997).

28. Monrose F. and Rubin A. D.. *Keystroke Dynamics as a Biometric for Authentication*, (pdf) Future Generation Computer Systems, (March, 2000).

29. Monrose F., Reiter M. K., and Wetzel S., \Password hardening based on keystroke dynamics," Proceedings of sixth ACM Conference on Computer and Communications Security, CCCS, pp. 73{82, 1999.

30. Cho S. and Hwang S.. Artificial rhythms and cues for keystroke dynamics based authentication. In International Conference on Biometrics, pages 626--632, 2006.

31. Obaidat, M.S., Macchairolo, D.T.: An On-line Neural Network System for Computer Access Security. IEEE Trans. Industrial Electronics, April 1993, Vol. 40, N. 2, pp. 235-241

32. Brown, M., Rogers, S. J.: User identification Via Keystroke Characteristics of Typed Names Using Neural Networks. International Journal of Man-Machine Studies, 1993, vol. 39, no. 6, pp. 999–1014.

33. Obaidat, M. S., Sadoun, B.: An Evaluation Simulation Study of Neural Network Paradigm for Computer Users Identification. Information Sciences Journal-Applications, Elsevier, November 1997, vol. 102, no. 1-4, pp. 239-258.

34. Bergadano, F., Gunetti, D., Picardi, C.: User Authentication Through Keystroke Dynamics. ACM Transacations on Information and System Security, 2002, vol. 5, no.4, 367–397.

35. Mroczkowski, P.: Identity Verification using Keyboard Statistics. [Online accessed: April 26th, 2010]. Available at: http://liu.divaportal.org/smash/record.jsf?pid=diva2:19595

36. Livia C. F. Araujo, Luiz H. R.., Miguel G. Lizarraga, Lee L. Ling, Joao B. T. Yabu-uti (2004). User Authentication through Typing Biometrics Features. Proceedings of the First International Conference on Biometric Authentication, ICBA 2004, Springer-Verlag Heidelberg, International Conference on Biometric Authentication, 694—700

37. Giot R., El-Abed M., and Rosenberger C., "Keystroke Dynamics authentication for collaborative systems," in The IEEE International Symposium on Collaborative Technologies and Systems (CTS), 2009, pp. 172–179.

38. Revett K, A Bioinformatics Based Approach to User Authentication via Keystroke Dynamics, International Journal of Control, Automation, and Systems (2009) 7(1):7-15

39. Hussien B., Bleha S., and McLaren R., "An application of fuzzy algorithms in a computer access security system," *Pattern Recognition Letters*, vol. 9, pp. 39-43,1989.

40.de Oliveira M. VS., Kinto E., Hernandez E. D. M., and de Carvalho T. C., "User authentication based on human typing patterns with artificial neural networks and support vector machines," SBC 2005.

41. Revett, K., Gorunescu, F., Gorunescu, M., Ene, M., Tenreiro de Magalhaes, S., Dinis Santos, H. M. (2007). A machine learning approach to keystroke dynamics based user authentication. *International Journal of Electronic Security and Digital Forensics*, vol. 1, no. 1, pp. 55-70. Inderscience Enterprises.

42. QueryPerformanceFrequency Function, last updated 6 April 2010 last accessed 10 June 2010 *http://msdn.microsoft.com/en-us/library/ms644905(VS.85).aspx*

43. Gunnetti, D., Picardi, C.: Keystroke Analysis of Free Text: ACM, Vol. 8, No. 3 (2005) 312 – 347

44. Gokcay, D., & Gokcay, E. (1995), "Combining Statistics and Heuristics in Language Identification", *Proceedings of the Fourth Annual Symposium on Document Analysis and Information Retrieval*, 1995, Las Vegas

45. Public Key Infrastructure, last accessed 12 June 2010 http://www.itl.nist.gov/fipspubs/fip186.htm#FIPS_TOP

46. "World Internet Users and Population Stats". *Internet World Stats*. Miniwatts Marketing Group. 2009-06-30. http://www.internetworldstats.com/stats.htm. Retrieved 2010-06-06.

47. "Threats to clinical confidentiality" Ross Anderson last updated 12 Jan 1996 last accessed 10 June 2010 http://www.cl.cam.ac.uk/~rja14/policy11/node9.html

APPENDICES

APPENDIX A - N-GRAMS and TIMINGS

n-gram	be01	be02	be03	be04	be05	be06	be07	be08	be09	be10	be11	be12	avg
ad	0,196	0,251	0,205	0,270	0,295	0,322	0,284	0,291	0,298	0,248	0,291	0,291	0,270
ak	0,326	0,229	0,288	0,269	0,232	0,267	0,215	0,270	0,245	0,261	0,271	0,270	0,262
al	0,269	0,240	0,282	0,300	0,218	0,283	0,268	0,285	0,303	0,277	0,285	0,285	0,275
an	0,176	0,195	0,211	0,240	0,184	0,222	0,216	0,225	0,242	0,224	0,225	0,224	0,215
ar	0,235	0,232	0,203	0,209	0,218	0,242	0,195	0,224	0,205	0,212	0,225	0,224	0,219
as	0,242	0,231	0,218	0,324	0,250	0,269	0,315	0,260	0,346	0,237	0,260	0,259	0,268
at	0,347	0,363	0,320	0,279	0,324	0,298	0,297	0,231	0,268	0,227	0,231	0,230	0,284
ate	0,472	0,532	0,504	0,428	0,531	0,457	0,532	0,425	0,460	0,485	0,426	0,425	0,473
da	0,340	0,307	0,325	0,361	0,310	0,322	0,394	0,356	0,330	0,339	0,356	0,355	0,341
de	0,300	0,303	0,313	0,302	0,268	0,332	0,303	0,333	0,298	0,297	0,333	0,332	0,309
ek	0,307	0,307	0,330	0,315	0,377	0,323	0,326	0,321	0,269	0,219	0,321	0,320	0,311
el	0,272	0,272	0,234	0,351	0,275	0,306	0,315	0,363	0,256	0,344	0,364	0,363	0,310
em	0,197	0,225	0,221	0,245	0,241	0,343	0,231	0,250	0,311	0,282	0,251	0,250	0,254
en	0,248	0,183	0,245	0,229	0,215	0,244	0,294	0,248	0,239	0,308	0,248	0,247	0,246
er	0,298	0,274	0,210	0,214	0,259	0,238	0,237	0,238	0,244	0,199	0,239	0,238	0,241
et	0,232	0,229	0,221	0,229	0,230	0,230	0,232	0,220	0,318	0,255	0,220	0,219	0,236
ka	0,312	0,281	0,326	0,382	0,324	0,348	0,238	0,330	0,368	0,356	0,330	0,329	0,327
la	0,271	0,301	0,289	0,336	0,272	0,298	0,294	0,305	0,298	0,302	0,305	0,305	0,298
le	0,310	0,273	0,235	0,330	0,314	0,302	0,317	0,333	0,323	0,298	0,334	0,333	0,308
len	0,363	0,344	0,355	0,377	0,380	0,360	0,357	0,366	0,385	0,344	0,366	0,365	0,363
ler	0,366	0,403	0,282	0,389	0,343	0,368	0,354	0,394	0,383	0,348	0,394	0,394	0,368
ma	0,309	0,343	0,356	0,376	0,312	0,361	0,390	0,356	0,346	0,349	0,356	0,356	0,351
me	0,221	0,208	0,223	0,269	0,218	0,336	0,245	0,282	0,310	0,250	0,283	0,282	0,260
nd	0,279	0,280	0,269	0,245	0,215	0,280	0,268	0,195	0,281	0,216	0,195	0,194	0,243
nm	0,214	0,203	0,216	0,268	0,232	0,231	0,218	0,171	0,239	0,238	0,171	0,171	0,214
ol	0,318	0,334	0,357	0,308	0,290	0,232	0,241	0,248	0,228	0,233	0,249	0,248	0,274
on	0,180	0,200	0,159	0,204	0,210	0,210	0,181	0,195	0,142	0,190	0,195	0,194	0,188
or	0,249	0,232	0,266	0,242	0,197	0,206	0,269	0,265	0,251	0,283	0,265	0,265	0,249
ra	0,256	0,286	0,332	0,291	0,281	0,365	0,326	0,374	0,259	0,355	0,375	0,374	0,323
rl	0,292	0,265	0,257	0,305	0,280	0,282	0,327	0,264	0,268	0,270	0,265	0,264	0,278
sa	0,307	0,244	0,292	0,330	0,236	0,288	0,293	0,279	0,240	0,284	0,280	0,279	0,279
so	0,229	0,306	0,382	0,398	0,232	0,211	0,293	0,319	0,229	0,293	0,319	0,319	0,294
st	0,355	0,306	0,274	0,293	0,269	0,330	0,355	0,290	0,256	0,312	0,291	0,290	0,302
sy	0,299	0,275	0,270	0,269	0,302	0,305	0,233	0,250	0,283	0,280	0,250	0,249	0,272
syo	0,366	0,353	0,351	0,496	0,382	0,345	0,313	0,356	0,350	0,346	0,357	0,356	0,364
ta	0,348	0,382	0,365	0,357	0,418	0,352	0,369	0,383	0,410	0,321	0,383	0,382	0,372
te	0,325	0,381	0,272	0,293	0,343	0,306	0,348	0,347	0,284	0,306	0,347	0,346	0,325
ter	0,383	0,440	0,328	0,319	0,440	0,350	0,384	0,420	0,383	0,348	0,420	0,419	0,386
tk	0,322	0,276	0,300	0,231	0,285	0,332	0,280	0,209	0,296	0,276	0,273	0,279	0,280
yo	0,269	0,287	0,292	0,362	0,318	0,299	0,280	0,356	0,311	0,279	0,357	0,356	0,314
yon	0,372	0,395	0,389	0,420	0,439	0,430	0,350	0,365	0,355	0,347	0,365	0,364	0,383

n-gram	ec01	ec02	ec03	ec04	ec05	ec06	ec07	ec08	ec09	ec10	ec11	ec12	average
ad	0,308	0,294	0,287	0,276	0,284	0,276	0,281	0,264	0,288	0,232	0,274	0,230	0,274
ak	0,401	0,202	0,329	0,291	0,355	0,291	0,322	0,292	0,306	0,384	0,377	0,312	0,322
al	0,387	0,285	0,293	0,275	0,268	0,271	0,271	0,279	0,278	0,274	0,276	0,353	0,292
an	0,262	0,308	0,293	0,329	0,262	0,318	0,272	0,236	0,279	0,268	0,252	0,235	0,276
ar	0,266	0,247	0,257	0,275	0,291	0,267	0,219	0,322	0,281	0,259	0,349	0,268	0,275
as	0,308	0,393	0,327	0,351	0,451	0,351	0,390	0,309	0,376	0,380	0,397	0,370	0,367
at	0,357	0,383	0,383	0,333	0,387	0,333	0,366	0,320	0,383	0,357	0,322	0,360	0,357
ate	0,628	0,477	0,550	0,597	0,567	0,480	0,610	0,438	0,564	0,522	0,480	0,503	0,535
da	0,415	0,366	0,357	0,340	0,353	0,353	0,355	0,348	0,347	0,307	0,309	0,332	0,349
de	0,370	0,308	0,294	0,322	0,315	0,340	0,417	0,332	0,334	0,327	0,310	0,358	0,336
ek	0,310	0,359	0,342	0,261	0,318	0,261	0,310	0,390	0,287	0,269	0,329	0,257	0,308
el	0,253	0,223	0,266	0,322	0,323	0,299	0,275	0,297	0,281	0,295	0,279	0,271	0,282
em	0,216	0,212	0,232	0,343	0,252	0,343	0,273	0,251	0,243	0,233	0,258	0,270	0,261
en	0,265	0,272	0,259	0,325	0,340	0,308	0,285	0,328	0,319	0,234	0,217	0,242	0,283
er	0,365	0,291	0,320	0,338	0,299	0,313	0,250	0,252	0,285	0,274	0,229	0,277	0,291
et	0,389	0,275	0,252	0,272	0,229	0,367	0,213	0,220	0,358	0,237	0,218	0,279	0,276
ka	0,433	0,327	0,358	0,310	0,390	0,310	0,294	0,339	0,374	0,267	0,310	0,275	0,332
la	0,345	0,315	0,359	0,286	0,321	0,287	0,306	0,324	0,319	0,333	0,331	0,274	0,317
le	0,343	0,310	0,321	0,316	0,385	0,341	0,301	0,282	0,371	0,290	0,315	0,309	0,324
len	0,451	0,467	0,520	0,461	0,504	0,465	0,480	0,427	0,659	0,700	0,486	0,550	0,514
ler	0,436	0,428	0,426	0,381	0,374	0,381	0,403	0,398	0,424	0,403	0,430	0,423	0,409
ma	0,523	0,372	0,372	0,350	0,363	0,350	0,304	0,389	0,409	0,385	0,412	0,268	0,375
me	0,239	0,283	0,308	0,310	0,358	0,297	0,344	0,210	0,265	0,356	0,282	0,373	0,302
nd	0,291	0,361	0,388	0,289	0,313	0,290	0,306	0,324	0,348	0,346	0,277	0,285	0,318
nm	0,351	0,387	0,293	0,297	0,314	0,298	0,322	0,347	0,353	0,326	0,413	0,350	0,338
ol	0,327	0,402	0,448	0,447	0,420	0,323	0,376	0,293	0,332	0,299	0,281	0,362	0,359
on	0,244	0,278	0,227	0,229	0,232	0,278	0,251	0,315	0,330	0,263	0,218	0,255	0,260
or	0,255	0,326	0,371	0,381	0,260	0,349	0,383	0,356	0,272	0,373	0,366	0,490	0,349
ra	0,323	0,378	0,448	0,398	0,295	0,399	0,384	0,280	0,362	0,326	0,460	0,419	0,373
rl	0,405	0,365	0,372	0,321	0,369	0,321	0,404	0,333	0,366	0,357	0,349	0,286	0,354
sa	0,270	0,318	0,339	0,337	0,297	0,303	0,312	0,321	0,252	0,281	0,172	0,359	0,297
so	0,347	0,480	0,357	0,363	0,247	0,349	0,266	0,354	0,357	0,332	0,392	0,340	0,349
st	0,398	0,351	0,443	0,595	0,424	0,444	0,279	0,341	0,378	0,382	0,305	0,336	0,390
sy	0,368	0,413	0,373	0,516	0,450	0,516	0,558	0,387	0,433	0,384	0,411	0,294	0,425
syo	0,438	0,536	0,467	0,634	0,454	0,634	0,696	0,484	0,569	0,494	0,528	0,399	0,528
ta	0,455	0,359	0,271	0,414	0,310	0,445	0,369	0,357	0,353	0,326	0,408	0,403	0,372
te	0,238	0,364	0,361	0,375	0,366	0,341	0,264	0,332	0,345	0,394	0,326	0,318	0,335
ter	0,387	0,405	0,535	0,602	0,485	0,495	0,495	0,496	0,524	0,548	0,458	0,450	0,490
tk	0,273	0,364	0,569	0,499	0,341	0,386	0,423	0,409	0,388	0,375	0,432	0,444	0,408
yo	0,306	0,381	0,292	0,346	0,319	0,336	0,337	0,304	0,327	0,380	0,322	0,263	0,326
yon	0,439	0,364	0,436	0,487	0,464	0,487	0,477	0,451	0,509	0,463	0,463	0,414	0,454

n-gram	nc01	nc02	nc03	nc04	nc05	nc06	nc07	nc08	nc09	nc10	nc11	nc12	average
ad	0,220	0,216	0,205	0,192	0,228	0,264	0,232	0,248	0,219	0,198	0,256	0,216	0,225
ak	0,322	0,251	0,241	0,216	0,268	0,232	0,247	0,257	0,339	0,232	0,243	0,226	0,256
al	0,244	0,230	0,245	0,198	0,230	0,208	0,224	0,243	0,228	0,246	0,289	0,288	0,239
an	0,226	0,221	0,217	0,188	0,241	0,216	0,220	0,194	0,177	0,188	0,208	0,195	0,207
ar	0,196	0,193	0,211	0,189	0,191	0,199	0,205	0,278	0,231	0,221	0,280	0,195	0,216
as	0,267	0,296	0,200	0,211	0,322	0,232	0,216	0,219	0,311	0,223	0,218	0,214	0,244
at	0,206	0,196	0,200	0,225	0,212	0,200	0,240	0,176	0,237	0,190	0,194	0,225	0,208
ate	0,347	0,317	0,323	0,323	0,331	0,344	0,520	0,388	0,370	0,331	0,341	0,379	0,359
da	0,368	0,328	0,323	0,321	0,349	0,323	0,302	0,337	0,342	0,306	0,371	0,288	0,330
de	0,324	0,291	0,312	0,331	0,298	0,349	0,257	0,288	0,347	0,309	0,320	0,311	0,311
ek	0,254	0,216	0,246	0,357	0,257	0,200	0,224	0,224	0,352	0,224	0,265	0,231	0,254
el	0,176	0,216	0,220	0,216	0,230	0,299	0,226	0,184	0,276	0,226	0,270	0,290	0,236
em	0,189	0,199	0,182	0,193	0,260	0,256	0,195	0,208	0,380	0,266	0,200	0,211	0,228
en	0,194	0,221	0,174	0,197	0,196	0,184	0,199	0,178	0,238	0,162	0,185	0,204	0,194
er	0,220	0,196	0,196	0,214	0,203	0,217	0,194	0,218	0,235	0,189	0,193	0,236	0,209
et	0,212	0,256	0,181	0,224	0,190	0,224	0,168	0,153	0,232	0,194	0,200	0,219	0,204
ka	0,348	0,368	0,304	0,352	0,424	0,316	0,262	0,292	0,298	0,312	0,336	0,269	0,323
la	0,303	0,278	0,286	0,235	0,306	0,261	0,261	0,276	0,326	0,278	0,291	0,353	0,288
le	0,262	0,264	0,259	0,323	0,287	0,277	0,291	0,263	0,336	0,304	0,293	0,290	0,287
len	0,339	0,404	0,400	0,406	0,344	0,338	0,369	0,352	0,408	0,332	0,367	0,376	0,370
ler	0,355	0,306	0,287	0,388	0,352	0,328	0,398	0,324	0,428	0,344	0,408	0,372	0,357
ma	0,355	0,224	0,291	0,293	0,313	0,331	0,364	0,322	0,334	0,335	0,342	0,261	0,314
me	0,253	0,250	0,258	0,246	0,277	0,238	0,213	0,226	0,335	0,254	0,254	0,184	0,249
nd	0,238	0,240	0,248	0,251	0,282	0,248	0,241	0,272	0,274	0,268	0,285	0,261	0,259
nm	0,260	0,248	0,228	0,222	0,238	0,232	0,237	0,261	0,246	0,227	0,252	0,224	0,240
ol	0,254	0,250	0,224	0,281	0,204	0,220	0,212	0,216	0,250	0,236	0,230	0,225	0,234
on	0,160	0,171	0,181	0,168	0,172	0,171	0,225	0,183	0,195	0,170	0,176	0,166	0,178
or	0,228	0,212	0,214	0,180	0,233	0,187	0,248	0,195	0,232	0,176	0,180	0,288	0,214
ra	0,285	0,286	0,286	0,214	0,293	0,248	0,307	0,277	0,317	0,326	0,336	0,310	0,290
rl	0,226	0,246	0,261	0,216	0,294	0,240	0,174	0,202	0,278	0,233	0,184	0,277	0,236
sa	0,336	0,309	0,285	0,272	0,328	0,326	0,312	0,330	0,287	0,309	0,288	0,272	0,305
so	0,257	0,192	0,216	0,248	0,308	0,320	0,271	0,192	0,284	0,257	0,284	0,289	0,260
st	0,240	0,272	0,219	0,296	0,389	0,322	0,289	0,256	0,316	0,310	0,309	0,299	0,293
sy	0,288	0,220	0,228	0,256	0,236	0,308	0,261	0,232	0,237	0,208	0,214	0,245	0,244
syo	0,408	0,344	0,340	0,382	0,358	0,460	0,395	0,357	0,347	0,332	0,348	0,365	0,370
ta	0,328	0,352	0,339	0,321	0,325	0,261	0,281	0,301	0,377	0,310	0,355	0,274	0,319
te	0,249	0,265	0,258	0,238	0,251	0,241	0,297	0,313	0,326	0,279	0,233	0,307	0,272
ter	0,285	0,312	0,320	0,324	0,300	0,288	0,308	0,329	0,296	0,305	0,304	0,353	0,310
tk	0,256	0,352	0,242	0,352	0,262	0,283	0,208	0,307	0,296	0,291	0,274	0,260	0,282
yo	0,292	0,280	0,299	0,312	0,268	0,272	0,340	0,296	0,323	0,291	0,310	0,307	0,299
yon	0,332	0,312	0,357	0,328	0,307	0,312	0,446	0,373	0,356	0,339	0,366	0,357	0,349

n-gram	sa01	sa02	sa03	sa04	sa05	sa06	sa07	sa08	sa09	sa10	sa11	sa12	average
ad	0,198	0,225	0,205	0,224	0,208	0,168	0,192	0,190	0,296	0,232	0,211	0,253	0,217
ak	0,209	0,229	0,240	0,248	0,196	0,228	0,219	0,232	0,232	0,288	0,282	0,224	0,236
al	0,204	0,224	0,216	0,198	0,227	0,189	0,211	0,217	0,212	0,239	0,240	0,204	0,215
an	0,205	0,220	0,232	0,204	0,245	0,200	0,214	0,226	0,224	0,226	0,234	0,212	0,220
ar	0,182	0,232	0,192	0,172	0,214	0,188	0,207	0,246	0,216	0,316	0,253	0,201	0,218
as	0,218	0,247	0,216	0,240	0,224	0,229	0,229	0,234	0,280	0,231	0,233	0,215	0,233
at	0,240	0,257	0,211	0,219	0,199	0,204	0,192	0,202	0,184	0,202	0,208	0,212	0,211
ate	0,377	0,320	0,308	0,336	0,312	0,348	0,324	0,348	0,343	0,336	0,356	0,352	0,338
da	0,254	0,269	0,263	0,260	0,272	0,253	0,293	0,260	0,356	0,274	0,276	0,285	0,276
de	0,253	0,253	0,230	0,198	0,288	0,248	0,269	0,279	0,286	0,277	0,285	0,271	0,261
ek	0,271	0,232	0,240	0,210	0,256	0,194	0,221	0,203	0,205	0,219	0,205	0,250	0,225
el	0,205	0,193	0,192	0,221	0,231	0,192	0,226	0,203	0,245	0,277	0,220	0,208	0,218
em	0,191	0,223	0,171	0,163	0,193	0,188	0,184	0,258	0,217	0,243	0,196	0,205	0,202
en	0,181	0,226	0,184	0,213	0,219	0,203	0,197	0,224	0,223	0,264	0,230	0,229	0,216
er	0,210	0,213	0,208	0,191	0,200	0,186	0,207	0,210	0,230	0,236	0,210	0,211	0,209
et	0,202	0,199	0,216	0,198	0,184	0,216	0,152	0,248	0,168	0,232	0,200	0,202	0,201
ka	0,236	0,264	0,243	0,232	0,266	0,356	0,250	0,219	0,232	0,302	0,296	0,267	0,264
la	0,265	0,253	0,230	0,272	0,270	0,226	0,229	0,250	0,251	0,269	0,259	0,279	0,254
le	0,253	0,257	0,230	0,222	0,266	0,228	0,250	0,254	0,271	0,287	0,275	0,257	0,254
len	0,308	0,390	0,325	0,304	0,328	0,316	0,344	0,392	0,336	0,361	0,408	0,326	0,345
ler	0,311	0,344	0,280	0,312	0,346	0,248	0,336	0,315	0,338	0,394	0,294	0,288	0,317
ma	0,299	0,318	0,320	0,288	0,295	0,271	0,304	0,289	0,306	0,272	0,276	0,250	0,291
me	0,253	0,276	0,221	0,199	0,233	0,212	0,244	0,242	0,296	0,266	0,280	0,270	0,249
nd	0,221	0,231	0,250	0,220	0,219	0,226	0,208	0,234	0,212	0,225	0,236	0,213	0,225
nm	0,243	0,336	0,228	0,253	0,232	0,228	0,208	0,260	0,224	0,264	0,259	0,243	0,248
ol	0,229	0,264	0,264	0,220	0,217	0,208	0,214	0,224	0,204	0,251	0,271	0,241	0,234
on	0,169	0,197	0,172	0,184	0,264	0,320	0,195	0,204	0,219	0,176	0,196	0,227	0,210
or	0,214	0,229	0,207	0,238	0,190	0,176	0,239	0,183	0,219	0,248	0,195	0,192	0,211
ra	0,252	0,256	0,223	0,176	0,255	0,172	0,227	0,271	0,292	0,295	0,299	0,279	0,250
rl	0,194	0,200	0,188	0,220	0,210	0,208	0,228	0,236	0,224	0,176	0,200	0,187	0,206
sa	0,281	0,266	0,264	0,236	0,171	0,197	0,264	0,352	0,221	0,275	0,246	0,224	0,250
so	0,248	0,175	0,268	0,248	0,240	0,328	0,248	0,232	0,212	0,304	0,336	0,254	0,258
st	0,264	0,302	0,277	0,238	0,280	0,259	0,304	0,239	0,248	0,288	0,261	0,262	0,268
sy	0,258	0,278	0,260	0,276	0,263	0,264	0,244	0,267	0,288	0,248	0,252	0,264	0,264
syo	0,337	0,371	0,336	0,356	0,348	0,343	0,341	0,347	0,348	0,328	0,322	0,372	0,346
ta	0,305	0,284	0,296	0,288	0,365	0,232	0,356	0,283	0,280	0,317	0,286	0,287	0,298
te	0,264	0,249	0,198	0,230	0,243	0,222	0,285	0,241	0,273	0,226	0,286	0,245	0,247
ter	0,310	0,289	0,260	0,264	0,276	0,259	0,349	0,301	0,379	0,248	0,386	0,332	0,304
tk	0,167	0,289	0,296	0,260	0,200	0,228	0,292	0,200	0,278	0,230	0,200	0,183	0,235
yo	0,248	0,246	0,252	0,216	0,280	0,248	0,176	0,238	0,252	0,245	0,240	0,325	0,247
yon	0,339	0,349	0,348	0,331	0,324	0,365	0,256	0,341	0,364	0,336	0,346	0,437	0,345

n-gram	sy01	sy02	sy03	sy04	sy05	sy06	sy07	sy08	sy09	sy10	sy11	sy12	average
ad	0,155	0,194	0,241	0,232	0,246	0,213	0,162	0,164	0,176	0,186	0,160	0,268	0,200
ak	0,168	0,179	0,236	0,217	0,197	0,176	0,214	0,239	0,170	0,200	0,191	0,183	0,197
al	0,198	0,210	0,195	0,190	0,178	0,217	0,241	0,197	0,238	0,216	0,184	0,202	0,206
an	0,186	0,171	0,182	0,249	0,200	0,171	0,184	0,170	0,195	0,148	0,202	0,178	0,186
ar	0,164	0,165	0,164	0,197	0,186	0,182	0,207	0,154	0,160	0,154	0,185	0,172	0,174
as	0,196	0,183	0,181	0,192	0,194	0,181	0,193	0,234	0,191	0,169	0,195	0,175	0,190
at	0,168	0,225	0,260	0,209	0,203	0,187	0,181	0,203	0,176	0,183	0,219	0,192	0,200
ate	0,308	0,383	0,442	0,357	0,374	0,316	0,339	0,342	0,300	0,294	0,373	0,329	0,346
da	0,270	0,314	0,231	0,179	0,211	0,292	0,231	0,247	0,218	0,197	0,280	0,231	0,242
de	0,244	0,266	0,261	0,255	0,208	0,204	0,266	0,233	0,251	0,220	0,230	0,295	0,244
ek	0,192	0,182	0,176	0,184	0,255	0,223	0,234	0,197	0,185	0,221	0,177	0,218	0,204
el	0,198	0,178	0,160	0,163	0,207	0,179	0,225	0,210	0,202	0,202	0,208	0,245	0,198
em	0,194	0,190	0,207	0,187	0,186	0,152	0,204	0,190	0,203	0,177	0,153	0,182	0,185
en	0,192	0,177	0,180	0,215	0,223	0,177	0,226	0,178	0,193	0,180	0,207	0,186	0,194
er	0,185	0,203	0,205	0,231	0,257	0,170	0,197	0,167	0,197	0,258	0,224	0,189	0,207
et	0,178	0,176	0,191	0,256	0,182	0,175	0,222	0,239	0,207	0,211	0,182	0,204	0,202
ka	0,349	0,241	0,253	0,287	0,274	0,319	0,205	0,226	0,224	0,213	0,233	0,232	0,255
la	0,198	0,227	0,253	0,214	0,205	0,232	0,224	0,235	0,215	0,209	0,242	0,209	0,222
le	0,235	0,230	0,248	0,231	0,225	0,239	0,200	0,231	0,226	0,253	0,221	0,281	0,235
len	0,357	0,315	0,336	0,265	0,321	0,321	0,386	0,324	0,328	0,319	0,351	0,357	0,332
ler	0,305	0,304	0,342	0,317	0,339	0,396	0,269	0,297	0,302	0,334	0,265	0,341	0,318
ma	0,242	0,218	0,248	0,346	0,275	0,256	0,263	0,276	0,246	0,264	0,234	0,231	0,258
me	0,231	0,225	0,249	0,180	0,207	0,304	0,200	0,209	0,197	0,232	0,216	0,221	0,223
nd	0,179	0,230	0,245	0,171	0,192	0,199	0,225	0,192	0,199	0,181	0,200	0,195	0,201
nm	0,197	0,229	0,262	0,224	0,240	0,212	0,235	0,219	0,216	0,195	0,207	0,237	0,223
ol	0,187	0,243	0,233	0,207	0,205	0,237	0,218	0,156	0,190	0,165	0,181	0,263	0,207
on	0,219	0,169	0,170	0,392	0,169	0,172	0,199	0,205	0,176	0,176	0,178	0,188	0,201
or	0,200	0,237	0,193	0,163	0,178	0,375	0,270	0,173	0,181	0,174	0,222	0,214	0,215
ra	0,219	0,214	0,227	0,223	0,260	0,189	0,276	0,294	0,234	0,284	0,231	0,212	0,239
rl	0,161	0,190	0,243	0,163	0,181	0,218	0,150	0,254	0,222	0,175	0,234	0,204	0,200
sa	0,218	0,260	0,268	0,284	0,258	0,335	0,195	0,167	0,224	0,203	0,199	0,206	0,235
so	0,176	0,155	0,175	0,184	0,172	0,194	0,173	0,171	0,191	0,176	0,182	0,200	0,179
st	0,230	0,257	0,279	0,226	0,269	0,204	0,244	0,162	0,252	0,269	0,256	0,219	0,239
sy	0,223	0,224	0,182	0,216	0,252	0,201	0,259	0,238	0,213	0,180	0,195	0,155	0,212
syo	0,337	0,319	0,260	0,298	0,262	0,311	0,366	0,341	0,290	0,285	0,299	0,254	0,302
ta	0,268	0,256	0,386	0,272	0,247	0,256	0,304	0,274	0,258	0,272	0,273	0,271	0,278
te	0,217	0,272	0,293	0,289	0,267	0,259	0,292	0,276	0,243	0,259	0,284	0,210	0,263
ter	0,356	0,411	0,450	0,400	0,341	0,248	0,333	0,364	0,351	0,386	0,368	0,286	0,358
tk	0,200	0,259	0,147	0,193	0,115	0,280	0,264	0,227	0,261	0,220	0,199	0,262	0,219
yo	0,237	0,262	0,220	0,219	0,227	0,202	0,255	0,190	0,234	0,236	0,205	0,230	0,226
yon	0,343	0,337	0,311	0,312	0,309	0,259	0,294	0,267	0,317	0,316	0,311	0,289	0,305

APPENDIX B - CORRELATION MATRIX

Correlation Matrix (Determinant = 2,29E-025)

	ad	ak	al	an	ar	as	at	ate	da	de	ek	el	em
ad	1,00	0,21	0,39	0,28	0,44	0,45	0,44	0,37	0,49	0,26	0,34	0,37	0,26
ak	0,21	1,00	0,56	0,38	0,40	0,35	0,38	0,30	0,44	0,49	0,06	0,48	0,40
al	0,39	0,56	1,00	0,64	0,60	0,57	0,60	0,38	0,53	0,54	0,22	0,72	0,44
an	0,28	0,38	0,64	1,00	0,60	0,74	0,45	0,36	0,37	0,43	0,13	0,56	0,37
ar	0,44	0,40	0,60	0,60	1,00	0,57	0,48	0,39	0,46	0,47	0,19	0,57	0,41
as	0,45	0,35	0,57	0,74	0,57	1,00	0,60	0,54	0,64	0,49	0,32	0,56	0,51
at	0,44	0,38	0,60	0,45	0,48	0,60	1,00	0,77	0,52	0,47	0,39	0,35	0,30
ate	0,37	0,30	0,38	0,36	0,39	0,54	0,77	1,00	0,50	0,35	0,37	0,39	0,28
da	0,49	0,44	0,53	0,37	0,46	0,64	0,52	0,50	1,00	0,60	0,34	0,49	0,39
de	0,26	0,49	0,54	0,43	0,47	0,49	0,47	0,35	0,60	1,00	0,16	0,46	0,58
ek	0,34	0,06	0,22	0,13	0,19	0,32	0,39	0,37	0,34	0,16	1,00	0,28	0,18
el	0,37	0,48	0,72	0,56	0,57	0,56	0,35	0,39	0,49	0,46	0,28	1,00	0,54
em	0,26	0,40	0,44	0,37	0,41	0,51	0,30	0,28	0,39	0,58	0,18	0,54	1,00
en	0,44	0,44	0,42	0,56	0,56	0,71	0,58	0,58	0,48	0,39	0,31	0,49	0,50
er	0,41	0,34	0,48	0,46	0,52	0,57	0,80	0,73	0,43	0,45	0,32	0,33	0,40
et	0,26	0,43	0,47	0,46	0,34	0,51	0,68	0,45	0,33	0,42	0,24	0,29	0,44
ka	0,41	0,39	0,42	0,31	0,38	0,52	0,45	0,28	0,63	0,50	0,28	0,32	0,30
la	0,41	0,45	0,57	0,34	0,49	0,55	0,54	0,49	0,63	0,55	0,48	0,54	0,50
le	0,49	0,40	0,59	0,45	0,57	0,71	0,59	0,52	0,70	0,64	0,40	0,70	0,60
len	0,28	0,44	0,53	0,54	0,53	0,62	0,53	0,55	0,42	0,55	0,09	0,45	0,34
ler	0,57	0,35	0,54	0,39	0,46	0,47	0,50	0,36	0,38	0,47	0,46	0,49	0,40
ma	0,51	0,52	0,62	0,58	0,59	0,64	0,63	0,51	0,64	0,45	0,47	0,58	0,36
me	0,36	0,55	0,39	0,36	0,28	0,49	0,22	0,19	0,44	0,47	0,00	0,40	0,51
nd	0,41	0,37	0,51	0,46	0,54	0,68	0,63	0,56	0,57	0,53	0,24	0,31	0,40
nm	0,20	0,26	0,33	0,56	0,43	0,41	0,42	0,23	0,01	0,22	0,25	0,11	0,09
ol	0,46	0,31	0,48	0,50	0,52	0,64	0,74	0,65	0,49	0,52	0,36	0,33	0,33
on	0,20	0,13	0,12	0,45	0,33	0,33	0,34	0,26	0,01	0,22	0,28	0,13	0,08
or	0,49	0,43	0,60	0,51	0,36	0,50	0,46	0,42	0,39	0,40	0,10	0,43	0,36
ra	0,37	0,54	0,70	0,56	0,56	0,55	0,47	0,43	0,49	0,57	0,11	0,65	0,57
rl	0,52	0,45	0,59	0,51	0,51	0,77	0,77	0,72	0,63	0,53	0,38	0,52	0,46
sa	0,34	0,40	0,17	0,11	0,20	0,21	0,19	0,17	0,46	0,37	0,21	0,12	0,35
so	0,28	0,54	0,51	0,42	0,32	0,45	0,38	0,30	0,50	0,24	0,41	0,56	0,34
st	0,39	0,28	0,37	0,52	0,51	0,62	0,59	0,59	0,54	0,42	0,26	0,40	0,48
sy	0,39	0,39	0,49	0,75	0,57	0,81	0,65	0,61	0,48	0,55	0,22	0,47	0,46
syo	0,35	0,38	0,49	0,73	0,52	0,76	0,56	0,54	0,51	0,61	0,20	0,48	0,45
ta	0,32	0,40	0,58	0,51	0,50	0,52	0,55	0,43	0,54	0,54	0,34	0,50	0,48
te	0,35	0,31	0,40	0,30	0,44	0,48	0,44	0,65	0,41	0,30	0,43	0,52	0,41
ter	0,34	0,29	0,39	0,44	0,44	0,47	0,53	0,72	0,31	0,40	0,23	0,42	0,31
tk	0,41	0,28	0,46	0,44	0,47	0,57	0,54	0,50	0,42	0,43	0,22	0,32	0,33
yo	0,36	0,36	0,54	0,50	0,47	0,60	0,41	0,47	0,67	0,66	0,40	0,61	0,52
yon	0,40	0,42	0,56	0,61	0,64	0,65	0,65	0,68	0,49	0,54	0,29	0,58	0,48

	en	er	et	ka	la	le	len	ler	ma	me	nd	nm	ol	on
ad	0,44	0,41	0,26	0,41	0,41	0,49	0,28	0,57	0,51	0,36	0,41	0,20	0,46	0,20
ak	0,44	0,34	0,43	0,39	0,45	0,40	0,44	0,35	0,52	0,55	0,37	0,26	0,31	0,13
al	0,42	0,48	0,47	0,42	0,57	0,59	0,53	0,54	0,62	0,39	0,51	0,33	0,48	0,12
an	0,56	0,46	0,46	0,31	0,34	0,45	0,54	0,39	0,58	0,36	0,46	0,56	0,50	0,45
ar	0,56	0,52	0,34	0,38	0,49	0,57	0,53	0,46	0,59	0,28	0,54	0,43	0,52	0,33
as	0,71	0,57	0,51	0,52	0,55	0,71	0,62	0,47	0,64	0,49	0,68	0,41	0,64	0,33
at	0,58	0,80	0,68	0,45	0,54	0,59	0,53	0,50	0,63	0,22	0,63	0,42	0,74	0,34
ate	0,58	0,73	0,45	0,28	0,49	0,52	0,55	0,36	0,51	0,19	0,56	0,23	0,65	0,26
da	0,48	0,43	0,33	0,63	0,63	0,70	0,42	0,38	0,64	0,44	0,57	-0,01	0,49	-0,01
de	0,39	0,45	0,42	0,50	0,55	0,64	0,55	0,47	0,45	0,47	0,53	0,22	0,52	0,22
ek	0,31	0,32	0,24	0,28	0,48	0,40	0,09	0,46	0,47	0,00	0,24	0,25	0,36	0,28
el	0,49	0,33	0,29	0,32	0,54	0,70	0,45	0,49	0,58	0,40	0,31	0,11	0,33	0,13
em	0,50	0,40	0,44	0,30	0,50	0,60	0,34	0,40	0,36	0,51	0,40	0,09	0,33	0,08
en	1,00	0,64	0,58	0,37	0,33	0,61	0,56	0,38	0,49	0,46	0,49	0,34	0,63	0,45
er	0,64	1,00	0,66	0,33	0,46	0,61	0,58	0,58	0,56	0,26	0,59	0,42	0,72	0,35
et	0,58	0,66	1,00	0,42	0,34	0,54	0,51	0,44	0,52	0,24	0,45	0,41	0,47	0,39
ka	0,37	0,33	0,42	1,00	0,45	0,59	0,31	0,37	0,52	0,35	0,41	0,16	0,39	0,16
la	0,33	0,46	0,34	0,45	1,00	0,60	0,43	0,49	0,61	0,34	0,64	0,12	0,52	0,08
le	0,61	0,61	0,54	0,59	0,60	1,00	0,58	0,67	0,67	0,55	0,54	0,21	0,57	0,18
len	0,56	0,58	0,51	0,31	0,43	0,58	1,00	0,43	0,46	0,47	0,75	0,43	0,67	0,39
ler	0,38	0,58	0,44	0,37	0,49	0,67	0,43	1,00	0,61	0,36	0,44	0,51	0,45	0,35
ma	0,49	0,56	0,52	0,52	0,61	0,67	0,46	0,61	1,00	0,29	0,54	0,43	0,41	0,36
me	0,46	0,26	0,24	0,35	0,34	0,55	0,47	0,36	0,29	1,00	0,38	0,11	0,41	0,04
nd	0,49	0,59	0,45	0,41	0,64	0,54	0,75	0,44	0,54	0,38	1,00	0,39	0,73	0,27
nm	0,34	0,42	0,41	0,16	0,12	0,21	0,43	0,51	0,43	0,11	0,39	1,00	0,37	0,68
ol	0,63	0,72	0,47	0,39	0,52	0,57	0,67	0,45	0,41	0,41	0,73	0,37	1,00	0,29
on	0,45	0,35	0,39	0,16	0,08	0,18	0,39	0,35	0,36	0,04	0,27	0,68	0,29	1,00
or	0,53	0,47	0,35	0,18	0,38	0,37	0,65	0,47	0,27	0,51	0,54	0,31	0,64	0,25
ra	0,52	0,54	0,42	0,25	0,53	0,59	0,55	0,52	0,47	0,54	0,52	0,21	0,52	0,18
rl	0,60	0,67	0,54	0,47	0,61	0,69	0,67	0,47	0,61	0,40	0,68	0,32	0,68	0,29
sa	0,28	0,21	0,16	0,39	0,34	0,34	0,21	0,31	0,30	0,49	0,39	0,15	0,39	-0,04
so	0,53	0,34	0,38	0,31	0,49	0,50	0,28	0,35	0,51	0,39	0,42	0,09	0,40	0,17
st	0,56	0,75	0,43	0,44	0,48	0,65	0,54	0,49	0,53	0,36	0,66	0,29	0,69	0,26
sy	0,72	0,67	0,53	0,35	0,42	0,54	0,67	0,40	0,55	0,41	0,66	0,47	0,75	0,48
syo	0,60	0,57	0,49	0,34	0,43	0,53	0,67	0,42	0,52	0,37	0,65	0,44	0,69	0,45
ta	0,42	0,53	0,53	0,36	0,50	0,65	0,40	0,52	0,57	0,35	0,35	0,22	0,39	0,14
te	0,52	0,61	0,29	0,16	0,55	0,55	0,54	0,47	0,44	0,31	0,49	0,24	0,53	0,25
ter	0,55	0,72	0,35	0,10	0,41	0,47	0,61	0,42	0,36	0,31	0,52	0,37	0,62	0,37
tk	0,43	0,57	0,38	0,23	0,47	0,38	0,64	0,41	0,32	0,29	0,73	0,33	0,73	0,26
yo	0,46	0,46	0,36	0,47	0,62	0,72	0,53	0,49	0,58	0,39	0,54	0,12	0,53	0,20
yon	0,63	0,65	0,48	0,37	0,54	0,63	0,68	0,43	0,57	0,37	0,64	0,36	0,67	0,47

	or	ra	rl	sa	so	st	sy	syo	ta	te	ter	tk	yo	yon
ad	0,49	0,37	0,52	0,34	0,28	0,39	0,39	0,35	0,32	0,35	0,34	0,41	0,36	0,40
ak	0,43	0,54	0,45	0,40	0,54	0,28	0,39	0,38	0,40	0,31	0,29	0,28	0,36	0,42
al	0,60	0,70	0,59	0,17	0,51	0,37	0,49	0,49	0,58	0,40	0,39	0,46	0,54	0,56
an	0,51	0,56	0,51	0,11	0,42	0,52	0,75	0,73	0,51	0,30	0,44	0,44	0,50	0,61
ar	0,36	0,56	0,51	0,20	0,32	0,51	0,57	0,52	0,50	0,44	0,44	0,47	0,47	0,64
as	0,50	0,55	0,77	0,21	0,45	0,62	0,81	0,76	0,52	0,48	0,47	0,57	0,60	0,65
at	0,46	0,47	0,77	0,19	0,38	0,59	0,65	0,56	0,55	0,44	0,53	0,54	0,41	0,65
ate	0,42	0,43	0,72	0,17	0,30	0,59	0,61	0,54	0,43	0,65	0,72	0,50	0,47	0,68
da	0,39	0,49	0,63	0,46	0,50	0,54	0,48	0,51	0,54	0,41	0,31	0,42	0,67	0,49
de	0,40	0,57	0,53	0,37	0,24	0,42	0,55	0,61	0,54	0,30	0,40	0,43	0,66	0,54
ek	0,10	0,11	0,38	0,21	0,41	0,26	0,22	0,20	0,34	0,43	0,23	0,22	0,40	0,29
el	0,43	0,65	0,52	0,12	0,56	0,40	0,47	0,48	0,50	0,52	0,42	0,32	0,61	0,58
em	0,36	0,57	0,46	0,35	0,34	0,48	0,46	0,45	0,48	0,41	0,31	0,33	0,52	0,48
en	0,53	0,52	0,60	0,28	0,53	0,56	0,72	0,60	0,42	0,52	0,55	0,43	0,46	0,63
er	0,47	0,54	0,67	0,21	0,34	0,75	0,67	0,57	0,53	0,61	0,72	0,57	0,46	0,65
et	0,35	0,42	0,54	0,16	0,38	0,43	0,53	0,49	0,53	0,29	0,35	0,38	0,36	0,48
ka	0,18	0,25	0,47	0,39	0,31	0,44	0,35	0,34	0,36	0,16	0,10	0,23	0,47	0,37
la	0,38	0,53	0,61	0,34	0,49	0,48	0,42	0,43	0,50	0,55	0,41	0,47	0,62	0,54
le	0,37	0,59	0,69	0,34	0,50	0,65	0,54	0,53	0,65	0,55	0,47	0,38	0,72	0,63
len	0,65	0,55	0,67	0,21	0,28	0,54	0,67	0,67	0,40	0,54	0,61	0,64	0,53	0,68
ler	0,47	0,52	0,47	0,31	0,35	0,49	0,40	0,42	0,52	0,47	0,42	0,41	0,49	0,43
ma	0,27	0,47	0,61	0,30	0,51	0,53	0,55	0,52	0,57	0,44	0,36	0,32	0,58	0,57
me	0,51	0,54	0,40	0,49	0,39	0,36	0,41	0,37	0,35	0,31	0,31	0,29	0,39	0,37
nd	0,54	0,52	0,68	0,39	0,42	0,66	0,66	0,65	0,35	0,49	0,52	0,73	0,54	0,64
nm	0,31	0,21	0,32	0,15	0,09	0,29	0,47	0,44	0,22	0,24	0,37	0,33	0,12	0,36
ol	0,64	0,52	0,68	0,39	0,40	0,69	0,75	0,69	0,39	0,53	0,62	0,73	0,53	0,67
on	0,25	0,18	0,29	-0,04	0,17	0,26	0,48	0,45	0,14	0,25	0,37	0,26	0,20	0,47
or	1,00	0,59	0,55	0,37	0,39	0,46	0,60	0,62	0,40	0,49	0,53	0,69	0,38	0,49
ra	0,59	1,00	0,55	0,29	0,52	0,50	0,53	0,54	0,50	0,55	0,56	0,50	0,58	0,59
rl	0,55	0,55	1,00	0,25	0,47	0,57	0,71	0,67	0,59	0,59	0,63	0,60	0,56	0,66
sa	0,37	0,29	0,25	1,00	0,38	0,38	0,23	0,27	0,15	0,25	0,19	0,19	0,30	0,19
so	0,39	0,52	0,47	0,38	1,00	0,43	0,36	0,38	0,32	0,37	0,28	0,25	0,46	0,37
st	0,46	0,50	0,57	0,38	0,43	1,00	0,63	0,60	0,47	0,52	0,60	0,57	0,49	0,62
sy	0,60	0,53	0,71	0,23	0,36	0,63	1,00	0,95	0,45	0,45	0,59	0,64	0,57	0,80
syo	0,62	0,54	0,67	0,27	0,38	0,60	0,95	1,00	0,45	0,40	0,55	0,65	0,63	0,76
ta	0,40	0,50	0,59	0,15	0,32	0,47	0,45	0,45	1,00	0,45	0,43	0,34	0,53	0,46
te	0,49	0,55	0,59	0,25	0,37	0,52	0,45	0,40	0,45	1,00	0,75	0,48	0,52	0,50
ter	0,53	0,56	0,63	0,19	0,28	0,60	0,59	0,55	0,43	0,75	1,00	0,58	0,43	0,63
tk	0,69	0,50	0,60	0,19	0,25	0,57	0,64	0,65	0,34	0,48	0,58	1,00	0,33	0,56
yo	0,38	0,58	0,56	0,30	0,46	0,49	0,57	0,63	0,53	0,52	0,43	0,33	1,00	0,72
yon	0,49	0,59	0,66	0,19	0,37	0,62	0,80	0,76	0,46	0,50	0,63	0,56	0,72	1,00

Sig. (1-tailed)

	ad	ak	al	an	ar	as	at	ate	da	de	ek	el	em
ad		0,06	0,00	0,02	0,00	0,00	0,00	0,00	0,00	0,02	0,00	0,00	0,02
ak	0,06		0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,31	0,00	0,00
al	0,00	0,00		0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,05	0,00	0,00
an	0,02	0,00	0,00		0,00	0,00	0,00	0,00	0,00	0,00	0,16	0,00	0,00
ar	0,00	0,00	0,00	0,00		0,00	0,00	0,00	0,00	0,00	0,07	0,00	0,00
as	0,00	0,00	0,00	0,00	0,00		0,00	0,00	0,00	0,00	0,01	0,00	0,00
at	0,00	0,00	0,00	0,00	0,00	0,00		0,00	0,00	0,00	0,00	0,00	0,01
ate	0,00	0,01	0,00	0,00	0,00	0,00	0,00		0,00	0,00	0,00	0,00	0,02
da	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00		0,00	0,00	0,00	0,00
de	0,02	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00		0,11	0,00	0,00
ek	0,00	0,31	0,05	0,16	0,07	0,01	0,00	0,00	0,00	0,11		0,02	0,08
el	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,02		0,00
em	0,02	0,00	0,00	0,00	0,00	0,00	0,01	0,02	0,00	0,00	0,08	0,00	
en	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,00
er	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,00
et	0,02	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,03	0,01	0,00
ka	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,02	0,00	0,00	0,02	0,01	0,01
la	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
le	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
len	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,24	0,00	0,00
ler	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
ma	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
me	0,00	0,00	0,00	0,00	0,01	0,00	0,05	0,07	0,00	0,00	0,49	0,00	0,00
nd	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,03	0,01	0,00
nm	0,06	0,02	0,00	0,00	0,00	0,00	0,00	0,04	0,48	0,05	0,03	0,21	0,24
ol	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01
on	0,06	0,16	0,17	0,00	0,01	0,01	0,00	0,02	0,47	0,05	0,02	0,17	0,27
or	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,21	0,00	0,00
ra	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,21	0,00	0,00
rl	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
sa	0,00	0,00	0,10	0,19	0,06	0,05	0,08	0,10	0,00	0,00	0,06	0,18	0,00
so	0,01	0,00	0,00	0,00	0,01	0,00	0,00	0,01	0,00	0,03	0,00	0,00	0,00
st	0,00	0,02	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,02	0,00	0,00
sy	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,04	0,00	0,00
syo	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,06	0,00	0,00
ta	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
te	0,00	0,01	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00
ter	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,04	0,00	0,01
tk	0,00	0,02	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,05	0,01	0,01
yo	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
yon	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,00

	en	er	et	ka	la	le	len	ler	ma	me	nd	nm	ol	on
ad	0,00	0,00	0,02	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,06	0,00	0,06
ak	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,02	0,01	0,16
al	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,17
an	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
ar	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,01
as	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01
at	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,05	0,00	0,00	0,00	0,00
ate	0,00	0,00	0,00	0,02	0,00	0,00	0,00	0,00	0,00	0,07	0,00	0,04	0,00	0,02
da	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,48	0,00	0,47
de	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,05	0,00	0,05
ek	0,01	0,01	0,03	0,02	0,00	0,00	0,24	0,00	0,00	0,49	0,03	0,03	0,00	0,02
el	0,00	0,01	0,01	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,21	0,00	0,17
em	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,24	0,01	0,27
en		0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
er	0,00		0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,02	0,00	0,00	0,00	0,00
et	0,00	0,00		0,00	0,00	0,00	0,00	0,00	0,00	0,03	0,00	0,00	0,00	0,00
ka	0,00	0,01	0,00		0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,12	0,00	0,11
la	0,00	0,00	0,00	0,00		0,00	0,00	0,00	0,00	0,00	0,00	0,19	0,00	0,28
le	0,00	0,00	0,00	0,00	0,00		0,00	0,00	0,00	0,00	0,00	0,06	0,00	0,09
len	0,00	0,00	0,00	0,01	0,00	0,00		0,00	0,00	0,00	0,00	0,00	0,00	0,00
ler	0,00	0,00	0,00	0,00	0,00	0,00	0,00		0,00	0,00	0,00	0,00	0,00	0,00
ma	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00		0,01	0,00	0,00	0,00	0,00
me	0,00	0,02	0,03	0,00	0,00	0,00	0,00	0,00	0,01		0,00	0,20	0,00	0,39
nd	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00		0,00	0,00	0,02
nm	0,00	0,00	0,00	0,12	0,19	0,06	0,00	0,00	0,00	0,20	0,00		0,00	0,00
ol	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00		0,01
on	0,00	0,00	0,00	0,11	0,28	0,09	0,00	0,00	0,00	0,39	0,02	0,00	0,01	
or	0,00	0,00	0,00	0,08	0,00	0,00	0,00	0,00	0,02	0,00	0,00	0,01	0,00	0,03
ra	0,00	0,00	0,00	0,03	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,06	0,00	0,08
rl	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,01
sa	0,01	0,05	0,12	0,00	0,00	0,00	0,05	0,01	0,01	0,00	0,00	0,12	0,00	0,37
so	0,00	0,00	0,00	0,01	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,24	0,00	0,10
st	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,02
sy	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
syo	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
ta	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,04	0,00	0,13
te	0,00	0,00	0,01	0,12	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,03	0,00	0,03
ter	0,00	0,00	0,00	0,23	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00
tk	0,00	0,00	0,00	0,04	0,00	0,00	0,00	0,00	0,01	0,01	0,00	0,01	0,00	0,02
yo	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,17	0,00	0,06
yon	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00

	or	ra	rl	sa	so	st	sy	syo	ta	te	ter	tk	yo	yon
ad	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00
ak	0,00	0,00	0,00	0,00	0,00	0,02	0,00	0,00	0,00	0,01	0,01	0,02	0,00	0,00
al	0,00	0,00	0,00	0,10	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
an	0,00	0,00	0,00	0,19	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00
ar	0,00	0,00	0,00	0,06	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
as	0,00	0,00	0,00	0,05	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
at	0,00	0,00	0,00	0,08	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
ate	0,00	0,00	0,00	0,10	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
da	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00
de	0,00	0,00	0,00	0,00	0,03	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00
ek	0,21	0,21	0,00	0,06	0,00	0,02	0,04	0,06	0,00	0,00	0,04	0,05	0,00	0,01
el	0,00	0,00	0,00	0,18	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,00
em	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,00	0,00
en	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
er	0,00	0,00	0,00	0,05	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
et	0,00	0,00	0,00	0,12	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00
ka	0,08	0,03	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,12	0,23	0,04	0,00	0,00
la	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
le	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
len	0,00	0,00	0,00	0,05	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
ler	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
ma	0,02	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,00
me	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,01	0,00	0,00
nd	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
nm	0,01	0,06	0,01	0,12	0,24	0,01	0,00	0,00	0,04	0,03	0,00	0,01	0,17	0,00
ol	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
on	0,03	0,08	0,01	0,37	0,10	0,02	0,00	0,00	0,13	0,03	0,00	0,02	0,06	0,00
or		0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
ra	0,00		0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
rl	0,00	0,00		0,03	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
sa	0,00	0,01	0,03		0,00	0,00	0,04	0,02	0,13	0,03	0,08	0,07	0,01	0,07
so	0,00	0,00	0,00	0,00		0,00	0,00	0,00	0,01	0,00	0,01	0,03	0,00	0,00
st	0,00	0,00	0,00	0,00	0,00		0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
sy	0,00	0,00	0,00	0,04	0,00	0,00		0,00	0,00	0,00	0,00	0,00	0,00	0,00
syo	0,00	0,00	0,00	0,02	0,00	0,00	0,00		0,00	0,00	0,00	0,00	0,00	0,00
ta	0,00	0,00	0,00	0,13	0,01	0,00	0,00	0,00		0,00	0,00	0,00	0,00	0,00
te	0,00	0,00	0,00	0,03	0,00	0,00	0,00	0,00	0,00		0,00	0,00	0,00	0,00
ter	0,00	0,00	0,00	0,08	0,01	0,00	0,00	0,00	0,00	0,00		0,00	0,00	0,00
tk	0,00	0,00	0,00	0,07	0,03	0,00	0,00	0,00	0,00	0,00	0,00		0,00	0,00
yo	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00		0,00
yon	0,00	0,00	0,00	0,07	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	

APPENDIX C - REPRODUCED CORRELATIONS

	ad	ak	al	an	ar	as	at	ate	da	de	ek	el	em
ad	0,59	0,27	0,41	0,29	0,41	0,43	0,42	0,34	0,48	0,33	0,47	0,34	0,23
ak	0,27	0,74	0,63	0,45	0,36	0,42	0,42	0,24	0,43	0,43	0,05	0,49	0,44
al	0,41	0,63	0,89	0,62	0,63	0,57	0,57	0,39	0,54	0,54	0,19	0,71	0,41
an	0,29	0,45	0,62	0,83	0,63	0,70	0,46	0,35	0,36	0,46	0,13	0,59	0,38
ar	0,41	0,36	0,63	0,63	0,62	0,62	0,49	0,40	0,46	0,52	0,27	0,58	0,40
as	0,43	0,42	0,57	0,70	0,62	0,80	0,62	0,57	0,64	0,57	0,32	0,57	0,47
at	0,42	0,42	0,57	0,46	0,49	0,62	0,91	0,75	0,54	0,45	0,40	0,33	0,28
ate	0,34	0,24	0,39	0,35	0,40	0,57	0,75	0,83	0,48	0,33	0,41	0,36	0,31
da	0,48	0,43	0,54	0,36	0,46	0,64	0,54	0,48	0,83	0,57	0,38	0,52	0,46
de	0,33	0,43	0,54	0,46	0,52	0,57	0,45	0,33	0,57	0,77	0,08	0,46	0,63
ek	0,47	0,05	0,19	0,13	0,27	0,32	0,40	0,41	0,38	0,08	0,76	0,30	0,12
el	0,34	0,49	0,71	0,59	0,58	0,57	0,33	0,36	0,52	0,46	0,30	0,87	0,54
em	0,23	0,44	0,41	0,38	0,40	0,47	0,28	0,31	0,46	0,63	0,12	0,54	0,72
en	0,33	0,48	0,40	0,59	0,45	0,67	0,57	0,59	0,45	0,40	0,33	0,48	0,49
er	0,38	0,36	0,46	0,43	0,47	0,58	0,81	0,77	0,43	0,48	0,37	0,34	0,42
et	0,21	0,51	0,47	0,47	0,39	0,50	0,72	0,51	0,36	0,46	0,21	0,27	0,39
ka	0,40	0,37	0,39	0,32	0,36	0,52	0,48	0,26	0,66	0,54	0,32	0,27	0,35
la	0,52	0,39	0,61	0,35	0,50	0,55	0,55	0,51	0,70	0,51	0,44	0,58	0,42
le	0,47	0,47	0,59	0,48	0,57	0,67	0,58	0,54	0,70	0,69	0,44	0,65	0,67
len	0,36	0,41	0,55	0,60	0,55	0,64	0,59	0,55	0,42	0,56	0,08	0,42	0,43
ler	0,57	0,38	0,56	0,40	0,54	0,44	0,49	0,35	0,39	0,50	0,49	0,47	0,43
ma	0,51	0,47	0,66	0,57	0,59	0,65	0,65	0,47	0,61	0,47	0,55	0,59	0,38
me	0,31	0,58	0,40	0,35	0,32	0,41	0,19	0,16	0,44	0,53	-0,01	0,41	0,59
nd	0,51	0,36	0,52	0,50	0,53	0,67	0,66	0,59	0,61	0,54	0,25	0,35	0,34
nm	0,32	0,22	0,32	0,54	0,42	0,37	0,42	0,21	-0,02	0,22	0,23	0,12	0,08
ol	0,50	0,34	0,44	0,48	0,49	0,67	0,70	0,69	0,55	0,49	0,30	0,32	0,35
on	0,21	0,09	0,14	0,54	0,36	0,40	0,33	0,26	-0,04	0,13	0,29	0,14	0,11
or	0,44	0,50	0,57	0,50	0,47	0,50	0,45	0,42	0,37	0,40	0,06	0,42	0,34
ra	0,36	0,59	0,71	0,55	0,55	0,55	0,45	0,44	0,48	0,56	0,12	0,69	0,58
rl	0,46	0,45	0,61	0,53	0,56	0,70	0,76	0,72	0,64	0,53	0,37	0,52	0,43
sa	0,44	0,39	0,14	0,07	0,14	0,27	0,18	0,12	0,45	0,34	0,22	0,12	0,36
so	0,37	0,60	0,51	0,42	0,34	0,49	0,39	0,38	0,53	0,19	0,39	0,58	0,33
st	0,43	0,32	0,38	0,43	0,46	0,63	0,61	0,63	0,54	0,52	0,34	0,38	0,47
sy	0,36	0,37	0,49	0,75	0,60	0,80	0,63	0,61	0,50	0,54	0,20	0,47	0,44
syo	0,37	0,36	0,50	0,74	0,61	0,79	0,56	0,54	0,51	0,57	0,16	0,48	0,45
ta	0,31	0,44	0,62	0,43	0,51	0,50	0,57	0,47	0,49	0,58	0,29	0,57	0,52
te	0,42	0,26	0,41	0,30	0,42	0,45	0,48	0,65	0,38	0,31	0,45	0,54	0,44
ter	0,34	0,25	0,39	0,41	0,44	0,49	0,56	0,70	0,28	0,35	0,28	0,42	0,40
tk	0,45	0,27	0,49	0,46	0,49	0,55	0,58	0,56	0,42	0,43	0,13	0,30	0,24
yo	0,42	0,33	0,52	0,51	0,56	0,68	0,42	0,46	0,67	0,62	0,37	0,67	0,60
yon	0,38	0,34	0,54	0,67	0,61	0,76	0,61	0,64	0,53	0,55	0,30	0,57	0,49

	en	er	et	ka	la	le	len	ler	ma	me	nd	nm	ol	on
ad	0,33	0,38	0,21	0,40	0,52	0,47	0,36	0,57	0,51	0,31	0,51	0,32	0,50	0,21
ak	0,48	0,36	0,51	0,37	0,39	0,47	0,41	0,38	0,47	0,58	0,36	0,22	0,34	0,09
al	0,40	0,46	0,47	0,39	0,61	0,59	0,55	0,56	0,66	0,40	0,52	0,32	0,44	0,14
an	0,59	0,43	0,47	0,32	0,35	0,48	0,60	0,40	0,57	0,35	0,50	0,54	0,48	0,54
ar	0,45	0,47	0,39	0,36	0,50	0,57	0,55	0,54	0,59	0,32	0,53	0,42	0,49	0,36
as	0,67	0,58	0,50	0,52	0,55	0,67	0,64	0,44	0,65	0,41	0,67	0,37	0,67	0,40
at	0,57	0,81	0,72	0,48	0,55	0,58	0,59	0,49	0,65	0,19	0,66	0,42	0,70	0,33
ate	0,59	0,77	0,51	0,26	0,51	0,54	0,55	0,35	0,47	0,16	0,59	0,21	0,69	0,26
da	0,45	0,43	0,36	0,66	0,70	0,70	0,42	0,39	0,61	0,44	0,61	-0,02	0,55	-0,04
de	0,40	0,48	0,46	0,54	0,51	0,69	0,56	0,50	0,47	0,53	0,54	0,22	0,49	0,13
ek	0,33	0,37	0,21	0,32	0,44	0,44	0,08	0,49	0,55	-0,01	0,25	0,23	0,30	0,29
el	0,48	0,34	0,27	0,27	0,58	0,65	0,42	0,47	0,59	0,41	0,35	0,12	0,32	0,14
em	0,49	0,42	0,39	0,35	0,42	0,67	0,43	0,43	0,38	0,59	0,34	0,08	0,35	0,11
en	0,80	0,63	0,58	0,34	0,39	0,58	0,56	0,38	0,53	0,46	0,52	0,38	0,61	0,47
er	0,63	0,86	0,68	0,33	0,49	0,61	0,63	0,54	0,53	0,28	0,62	0,44	0,71	0,40
et	0,58	0,68	0,81	0,44	0,32	0,53	0,48	0,43	0,55	0,29	0,43	0,46	0,47	0,40
ka	0,34	0,33	0,44	0,74	0,48	0,58	0,29	0,39	0,59	0,34	0,47	0,18	0,39	0,11
la	0,39	0,49	0,32	0,48	0,71	0,68	0,43	0,53	0,61	0,35	0,56	0,08	0,53	0,01
le	0,58	0,61	0,53	0,58	0,68	0,85	0,51	0,63	0,69	0,50	0,55	0,21	0,55	0,21
len	0,56	0,63	0,48	0,29	0,43	0,51	0,73	0,44	0,42	0,44	0,68	0,43	0,69	0,35
ler	0,38	0,54	0,43	0,39	0,53	0,63	0,44	0,84	0,62	0,36	0,45	0,55	0,45	0,39
ma	0,53	0,53	0,55	0,59	0,61	0,69	0,42	0,62	0,81	0,28	0,50	0,42	0,48	0,37
me	0,46	0,28	0,29	0,34	0,35	0,50	0,44	0,36	0,28	0,75	0,41	0,11	0,39	0,04
nd	0,52	0,62	0,43	0,47	0,56	0,55	0,68	0,45	0,50	0,41	0,78	0,35	0,78	0,25
nm	0,38	0,44	0,46	0,18	0,08	0,21	0,43	0,55	0,42	0,11	0,35	0,86	0,37	0,74
ol	0,61	0,71	0,47	0,39	0,53	0,55	0,69	0,45	0,48	0,39	0,78	0,37	0,83	0,32
on	0,47	0,40	0,40	0,11	0,01	0,21	0,35	0,39	0,37	0,04	0,25	0,74	0,32	0,78
or	0,49	0,49	0,31	0,16	0,42	0,38	0,66	0,44	0,33	0,54	0,63	0,37	0,65	0,23
ra	0,53	0,51	0,40	0,24	0,54	0,61	0,61	0,51	0,48	0,58	0,51	0,23	0,50	0,15
rl	0,63	0,72	0,57	0,46	0,63	0,67	0,64	0,48	0,63	0,36	0,70	0,30	0,73	0,26
sa	0,34	0,21	0,17	0,44	0,33	0,39	0,24	0,37	0,25	0,60	0,39	0,09	0,39	0,00
so	0,59	0,32	0,35	0,34	0,47	0,48	0,28	0,31	0,56	0,45	0,34	0,10	0,37	0,14
st	0,62	0,67	0,48	0,41	0,50	0,63	0,59	0,46	0,49	0,40	0,64	0,30	0,70	0,31
sy	0,73	0,65	0,54	0,39	0,42	0,57	0,74	0,37	0,53	0,40	0,71	0,48	0,74	0,53
syo	0,66	0,57	0,47	0,40	0,43	0,56	0,72	0,37	0,51	0,42	0,70	0,45	0,70	0,47
ta	0,41	0,57	0,55	0,41	0,55	0,68	0,45	0,56	0,59	0,31	0,40	0,24	0,39	0,17
te	0,53	0,64	0,29	0,08	0,53	0,57	0,48	0,53	0,41	0,32	0,46	0,19	0,56	0,22
ter	0,58	0,72	0,39	0,03	0,41	0,48	0,62	0,45	0,34	0,30	0,54	0,34	0,65	0,35
tk	0,42	0,58	0,31	0,22	0,46	0,39	0,69	0,40	0,35	0,32	0,73	0,36	0,74	0,24
yo	0,51	0,45	0,32	0,49	0,62	0,75	0,48	0,47	0,59	0,43	0,52	0,11	0,49	0,18
yon	0,66	0,65	0,49	0,37	0,52	0,65	0,67	0,44	0,57	0,36	0,64	0,38	0,67	0,43

	or	ra	rl	sa	so	st	sy	syo	ta	te	ter	tk	yo	yon
ad	0,44	0,36	0,46	0,44	0,37	0,43	0,36	0,37	0,31	0,42	0,34	0,45	0,42	0,38
ak	0,50	0,59	0,45	0,39	0,60	0,32	0,37	0,36	0,44	0,26	0,25	0,27	0,33	0,34
al	0,57	0,71	0,61	0,14	0,51	0,38	0,49	0,50	0,62	0,41	0,39	0,49	0,52	0,54
an	0,50	0,55	0,53	0,07	0,42	0,43	0,75	0,74	0,43	0,30	0,41	0,46	0,51	0,67
ar	0,47	0,55	0,56	0,14	0,34	0,46	0,60	0,61	0,51	0,42	0,44	0,49	0,56	0,61
as	0,50	0,55	0,70	0,27	0,49	0,63	0,80	0,79	0,50	0,45	0,49	0,55	0,68	0,76
at	0,45	0,45	0,76	0,18	0,39	0,61	0,63	0,56	0,57	0,48	0,56	0,58	0,42	0,61
ate	0,42	0,44	0,72	0,12	0,38	0,63	0,61	0,54	0,47	0,65	0,70	0,56	0,46	0,64
da	0,37	0,48	0,64	0,45	0,53	0,54	0,50	0,51	0,49	0,38	0,28	0,42	0,67	0,53
de	0,40	0,56	0,53	0,34	0,19	0,52	0,54	0,57	0,58	0,31	0,35	0,43	0,62	0,55
ek	0,06	0,12	0,37	0,22	0,39	0,34	0,20	0,16	0,29	0,45	0,28	0,13	0,37	0,30
el	0,42	0,69	0,52	0,12	0,58	0,38	0,47	0,48	0,57	0,54	0,42	0,30	0,67	0,57
em	0,34	0,58	0,43	0,36	0,33	0,47	0,44	0,45	0,52	0,44	0,40	0,24	0,60	0,49
en	0,49	0,53	0,63	0,34	0,59	0,62	0,73	0,66	0,41	0,53	0,58	0,42	0,51	0,66
er	0,49	0,51	0,72	0,21	0,32	0,67	0,65	0,57	0,57	0,64	0,72	0,58	0,45	0,65
et	0,31	0,40	0,57	0,17	0,35	0,48	0,54	0,47	0,55	0,29	0,39	0,31	0,32	0,49
ka	0,16	0,24	0,46	0,44	0,34	0,41	0,39	0,40	0,41	0,08	0,03	0,22	0,49	0,37
la	0,42	0,54	0,63	0,33	0,47	0,50	0,42	0,43	0,55	0,53	0,41	0,46	0,62	0,52
le	0,38	0,61	0,67	0,39	0,48	0,63	0,57	0,56	0,68	0,57	0,48	0,39	0,75	0,65
len	0,66	0,61	0,64	0,24	0,28	0,59	0,74	0,72	0,45	0,48	0,62	0,69	0,48	0,67
ler	0,44	0,51	0,48	0,37	0,31	0,46	0,37	0,37	0,56	0,53	0,45	0,40	0,47	0,44
ma	0,33	0,48	0,63	0,25	0,56	0,49	0,53	0,51	0,59	0,41	0,34	0,35	0,59	0,57
me	0,54	0,58	0,36	0,60	0,45	0,40	0,40	0,42	0,31	0,32	0,30	0,32	0,43	0,36
nd	0,63	0,51	0,70	0,39	0,34	0,64	0,71	0,70	0,40	0,46	0,54	0,73	0,52	0,64
nm	0,37	0,23	0,30	0,09	0,10	0,30	0,48	0,45	0,24	0,19	0,34	0,36	0,11	0,38
ol	0,65	0,50	0,73	0,39	0,37	0,70	0,74	0,70	0,39	0,56	0,65	0,74	0,49	0,67
on	0,23	0,15	0,26	0,00	0,14	0,31	0,53	0,47	0,17	0,22	0,35	0,24	0,18	0,43
or	0,80	0,64	0,55	0,39	0,41	0,48	0,58	0,58	0,30	0,51	0,59	0,69	0,35	0,51
ra	0,64	0,77	0,58	0,28	0,49	0,48	0,55	0,55	0,56	0,57	0,57	0,51	0,56	0,58
rl	0,55	0,58	0,78	0,27	0,49	0,66	0,70	0,66	0,57	0,58	0,61	0,62	0,59	0,70
sa	0,39	0,28	0,27	0,79	0,40	0,37	0,22	0,23	0,12	0,23	0,15	0,24	0,28	0,16
so	0,41	0,49	0,49	0,40	0,83	0,36	0,39	0,36	0,32	0,41	0,29	0,21	0,42	0,39
st	0,48	0,48	0,66	0,37	0,36	0,67	0,67	0,63	0,45	0,56	0,60	0,55	0,56	0,65
sy	0,58	0,55	0,70	0,22	0,39	0,67	0,91	0,88	0,42	0,46	0,61	0,64	0,59	0,81
syo	0,58	0,55	0,66	0,23	0,36	0,63	0,88	0,86	0,41	0,42	0,55	0,64	0,60	0,78
ta	0,30	0,56	0,57	0,12	0,32	0,45	0,42	0,41	0,68	0,45	0,42	0,32	0,55	0,52
te	0,51	0,57	0,58	0,23	0,41	0,56	0,46	0,42	0,45	0,82	0,75	0,49	0,51	0,56
ter	0,59	0,57	0,61	0,15	0,29	0,60	0,61	0,55	0,42	0,75	0,82	0,62	0,43	0,63
tk	0,69	0,51	0,62	0,24	0,21	0,55	0,64	0,64	0,32	0,49	0,62	0,79	0,38	0,59
yo	0,35	0,56	0,59	0,28	0,42	0,56	0,59	0,60	0,55	0,51	0,43	0,38	0,78	0,66
yon	0,51	0,58	0,70	0,16	0,39	0,65	0,81	0,78	0,52	0,56	0,63	0,59	0,66	0,78

Residual[22]

	ad	ak	al	an	ar	as	at	ate	da	de	ek	el	em
ad		-0,06	-0,02	-0,01	0,03	0,02	0,03	0,02	0,01	-0,07	-0,13	0,03	0,03
ak	-0,06		-0,07	-0,07	0,04	-0,07	-0,04	0,06	0,01	0,07	0,01	-0,01	-0,04
al	-0,02	-0,07		0,01	-0,03	-0,01	0,03	-0,01	-0,01	0,01	0,03	0,00	0,03
an	-0,01	-0,07	0,01		-0,03	0,03	-0,01	0,01	0,00	-0,03	0,00	-0,03	0,00
ar	0,03	0,04	-0,03	-0,03		-0,04	0,00	0,00	0,01	-0,05	-0,08	-0,01	0,01
as	0,02	-0,07	-0,01	0,03	-0,04		-0,01	-0,02	0,01	-0,09	0,00	-0,01	0,04
at	0,03	-0,04	0,03	-0,01	0,00	-0,01		0,02	-0,02	0,02	0,00	0,02	0,02
ate	0,02	0,06	-0,01	0,01	0,00	-0,02	0,02		0,02	0,02	-0,04	0,02	-0,03
da	0,01	0,01	-0,01	0,00	0,01	0,01	-0,02	0,02		0,02	-0,04	-0,04	-0,07
de	-0,07	0,07	0,01	-0,03	-0,05	-0,09	0,02	0,02	0,02		0,08	0,00	-0,04
ek	-0,13	0,01	0,03	0,00	-0,08	0,00	0,00	-0,04	-0,04	0,08		-0,03	0,06
el	0,03	-0,01	0,00	-0,03	-0,01	-0,01	0,02	0,02	-0,04	0,00	-0,03		0,00
em	0,03	-0,04	0,03	0,00	0,01	0,04	0,02	-0,03	-0,07	-0,04	0,06	0,00	
en	0,11	-0,04	0,01	-0,04	0,11	0,04	0,01	-0,01	0,03	-0,01	-0,02	0,01	0,01
er	0,03	-0,02	0,02	0,03	0,05	-0,01	-0,01	-0,04	0,01	-0,03	-0,05	-0,01	-0,02
et	0,06	-0,09	0,00	-0,01	-0,05	0,01	-0,04	-0,06	-0,02	-0,04	0,04	0,01	0,05
ka	0,02	0,02	0,03	0,00	0,02	0,00	-0,02	0,02	-0,03	-0,03	-0,04	0,06	-0,05
la	-0,11	0,07	-0,04	0,00	-0,01	0,00	-0,01	-0,02	-0,07	0,04	0,04	-0,04	0,08
le	0,02	-0,06	0,00	-0,02	0,01	0,05	0,01	-0,02	0,00	-0,05	-0,05	0,06	-0,07
len	-0,08	0,03	-0,02	-0,06	-0,02	-0,01	-0,06	0,00	0,00	-0,01	0,01	0,03	-0,09
ler	0,00	-0,02	-0,01	0,00	-0,08	0,03	0,01	0,01	-0,01	-0,02	-0,03	0,02	-0,03
ma	0,00	0,05	-0,04	0,01	0,00	-0,01	-0,02	0,04	0,03	-0,02	-0,07	-0,01	-0,02
me	0,06	-0,02	-0,01	0,01	-0,03	0,08	0,03	0,03	-0,01	-0,06	0,01	-0,02	-0,07
nd	-0,11	0,01	-0,01	-0,03	0,02	0,01	-0,03	-0,03	-0,03	-0,02	-0,01	-0,04	0,06
nm	-0,11	0,04	0,02	0,02	0,01	0,04	0,00	0,03	0,01	0,00	0,02	-0,01	0,01
ol	-0,04	-0,03	0,03	0,02	0,03	-0,03	0,04	-0,04	-0,06	0,04	0,06	0,01	-0,03
on	-0,01	0,04	-0,02	-0,08	-0,04	-0,07	0,01	0,00	0,03	0,09	-0,01	-0,02	-0,03
or	0,05	-0,07	0,03	0,01	-0,11	0,00	0,01	0,00	0,02	0,00	0,04	0,01	0,02
ra	0,01	-0,05	0,00	0,01	0,01	0,00	0,02	-0,01	0,02	0,01	-0,01	-0,05	-0,02
rl	0,06	0,00	-0,02	-0,03	-0,05	0,07	0,00	0,00	-0,02	0,00	0,00	0,00	0,03
sa	-0,10	0,01	0,03	0,05	0,06	-0,06	0,01	0,04	0,01	0,02	-0,01	0,01	-0,01
so	-0,09	-0,06	0,00	0,00	-0,02	-0,04	-0,01	-0,07	-0,03	0,05	0,02	-0,02	0,01
st	-0,03	-0,04	0,00	0,09	0,05	-0,02	-0,02	-0,05	0,00	-0,10	-0,09	0,03	0,01
sy	0,04	0,02	0,00	0,00	-0,03	0,01	0,02	0,00	-0,02	0,01	0,02	0,00	0,02
syo	-0,01	0,02	-0,02	-0,01	-0,09	-0,03	0,00	0,00	0,01	0,04	0,04	0,00	0,00
ta	0,01	-0,04	-0,04	0,08	-0,01	0,03	-0,02	-0,04	0,05	-0,04	0,05	-0,07	-0,04
te	-0,07	0,05	0,00	-0,01	0,02	0,04	-0,04	-0,01	0,03	-0,02	-0,03	-0,02	-0,03
ter	0,00	0,04	0,00	0,03	0,01	-0,02	-0,03	0,03	0,03	0,05	-0,05	0,00	-0,09
tk	-0,04	0,01	-0,03	-0,02	-0,02	0,02	-0,04	-0,06	0,00	0,00	0,09	0,02	0,09
yo	-0,05	0,04	0,02	-0,01	-0,09	-0,08	-0,01	0,00	0,00	0,04	0,02	-0,06	-0,08
yon	0,03	0,08	0,02	-0,07	0,03	-0,11	0,04	0,04	-0,03	-0,01	0,00	0,00	-0,01

[22] Residuals are computed between observed and reproduced correlations. There are 137 (16,0%) nonredundant residuals with absolute values greater than 0.05.

	en	er	et	ka	la	le	len	ler	ma	me	nd	nm	ol	on
ad	0,11	0,03	0,06	0,02	-0,11	0,02	-0,08	0,00	0,00	0,06	-0,11	-0,11	-0,04	-0,01
ak	-0,04	-0,02	-0,09	0,02	0,07	-0,06	0,03	-0,02	0,05	-0,02	0,01	0,04	-0,03	0,04
al	0,01	0,02	0,00	0,03	-0,04	0,00	-0,02	-0,01	-0,04	-0,01	-0,01	0,02	0,03	-0,02
an	-0,04	0,03	-0,01	0,00	0,00	-0,02	-0,06	0,00	0,01	0,01	-0,03	0,02	0,02	-0,08
ar	0,11	0,05	-0,05	0,02	-0,01	0,01	-0,02	-0,08	0,00	-0,03	0,02	0,01	0,03	-0,04
as	0,04	-0,01	0,01	0,00	0,00	0,05	-0,01	0,03	-0,01	0,08	0,01	0,04	-0,03	-0,07
at	0,01	-0,01	-0,04	-0,02	-0,01	0,01	-0,06	0,01	-0,02	0,03	-0,03	0,00	0,04	0,01
ate	-0,01	-0,04	-0,06	0,02	-0,02	-0,02	0,00	0,01	0,04	0,03	-0,03	0,03	-0,04	0,00
da	0,03	0,01	-0,02	-0,03	-0,07	0,00	0,00	-0,01	0,03	-0,01	-0,03	0,01	-0,06	0,03
de	-0,01	-0,03	-0,04	-0,03	0,04	-0,05	-0,01	-0,02	-0,02	-0,06	-0,02	0,00	0,04	0,09
ek	-0,02	-0,05	0,04	-0,04	0,04	-0,05	0,01	-0,03	-0,07	0,01	-0,01	0,02	0,06	-0,01
el	0,01	-0,01	0,01	0,06	-0,04	0,06	0,03	0,02	-0,01	-0,02	-0,04	-0,01	0,01	-0,02
em	0,01	-0,02	0,05	-0,05	0,08	-0,07	-0,09	-0,03	-0,02	-0,07	0,06	0,01	-0,03	-0,03
en		0,01	0,00	0,03	-0,05	0,03	0,00	0,00	-0,04	0,00	-0,04	-0,04	0,02	-0,02
er	0,01		-0,02	0,00	-0,03	0,00	-0,05	0,03	0,02	-0,02	-0,03	-0,02	0,01	-0,05
et	0,00	-0,02		-0,03	0,02	0,01	0,03	0,00	-0,04	-0,05	0,03	-0,05	0,00	-0,01
ka	0,03	0,00	-0,03		-0,04	0,01	0,02	-0,02	-0,07	0,01	-0,06	-0,02	0,00	0,05
la	-0,05	-0,03	0,02	-0,04		-0,08	0,00	-0,04	0,00	-0,01	0,08	0,04	-0,01	0,06
le	0,03	0,00	0,01	0,01	-0,08		0,07	0,05	-0,02	0,05	-0,01	0,00	0,03	-0,03
len	0,00	-0,05	0,03	0,02	0,00	0,07		0,00	0,04	0,03	0,08	0,00	-0,03	0,04
ler	0,00	0,03	0,00	-0,02	-0,04	0,05	0,00		-0,02	0,00	-0,01	-0,04	0,00	-0,04
ma	-0,04	0,02	-0,04	-0,07	0,00	-0,02	0,04	-0,02		0,02	0,04	0,01	-0,06	-0,01
me	0,00	-0,02	-0,05	0,01	-0,01	0,05	0,03	0,00	0,02		-0,03	0,00	0,01	0,00
nd	-0,04	-0,03	0,03	-0,06	0,08	-0,01	0,08	-0,01	0,04	-0,03		0,05	-0,05	0,02
nm	-0,04	-0,02	-0,05	-0,02	0,04	0,00	0,00	-0,04	0,01	0,00	0,05		0,00	-0,06
ol	0,02	0,01	0,00	0,00	-0,01	0,03	-0,03	0,00	-0,06	0,01	-0,05	0,00		-0,03
on	-0,02	-0,05	-0,01	0,05	0,06	-0,03	0,04	-0,04	-0,01	0,00	0,02	-0,06	-0,03	
or	0,04	-0,02	0,05	0,02	-0,04	-0,01	-0,01	0,02	-0,06	-0,03	-0,09	-0,06	-0,01	0,02
ra	-0,01	0,03	0,02	0,01	-0,01	-0,02	-0,06	0,01	0,00	-0,03	0,01	-0,02	0,02	0,04
rl	-0,03	-0,05	-0,03	0,01	-0,02	0,02	0,03	-0,01	-0,02	0,04	-0,02	0,02	-0,05	0,03
sa	-0,06	0,00	-0,02	-0,05	0,01	-0,05	-0,03	-0,05	0,05	-0,12	0,00	0,06	-0,01	-0,05
so	-0,06	0,02	0,03	-0,04	0,02	0,02	0,00	0,04	-0,05	-0,05	0,08	-0,01	0,03	0,03
st	-0,05	0,08	-0,05	0,02	-0,03	0,02	-0,05	0,03	0,04	-0,04	0,02	0,00	0,00	-0,05
sy	-0,01	0,02	-0,01	-0,04	0,00	-0,03	-0,07	0,03	0,01	0,01	-0,05	-0,02	0,00	-0,05
syo	-0,06	-0,01	0,03	-0,06	0,00	-0,03	-0,05	0,05	0,00	-0,06	-0,04	-0,01	-0,01	-0,02
ta	0,01	-0,04	-0,02	-0,05	-0,05	-0,03	-0,05	-0,04	-0,02	0,04	-0,05	-0,01	0,00	-0,02
te	-0,02	-0,02	0,00	0,07	0,02	-0,01	0,06	-0,06	0,03	-0,01	0,03	0,05	-0,03	0,02
ter	-0,03	-0,01	-0,05	0,06	0,00	-0,01	-0,02	-0,04	0,02	0,01	-0,02	0,03	-0,03	0,01
tk	0,01	-0,01	0,07	0,00	0,00	0,00	-0,04	0,00	-0,03	-0,02	0,00	-0,03	-0,01	0,02
yo	-0,05	0,01	0,04	-0,02	0,00	-0,04	0,05	0,01	-0,01	-0,03	0,02	0,01	0,03	0,03
yon	-0,03	0,00	-0,02	0,00	0,02	-0,02	0,01	-0,02	0,00	0,01	0,00	-0,01	0,00	0,04

	or	ra	rl	sa	so	st	sy	syo	ta	te	ter	tk	yo	yon
ad	0,05	0,01	0,06	-0,10	-0,09	-0,03	0,04	-0,01	0,01	-0,07	0,00	-0,04	-0,05	0,03
ak	-0,07	-0,05	0,00	0,01	-0,06	-0,04	0,02	0,02	-0,04	0,05	0,04	0,01	0,04	0,08
al	0,03	0,00	-0,02	0,03	0,00	0,00	0,00	-0,02	-0,04	0,00	0,00	-0,03	0,02	0,02
an	0,01	0,01	-0,03	0,05	0,00	0,09	0,00	-0,01	0,08	-0,01	0,03	-0,02	-0,01	-0,07
ar	-0,11	0,01	-0,05	0,06	-0,02	0,05	-0,03	-0,09	-0,01	0,02	0,01	-0,02	-0,09	0,03
as	0,00	0,00	0,07	-0,06	-0,04	-0,02	0,01	-0,03	0,03	0,04	-0,02	0,02	-0,08	-0,11
at	0,01	0,02	0,00	0,01	-0,01	-0,02	0,02	0,00	-0,02	-0,04	-0,03	-0,04	-0,01	0,04
ate	0,00	-0,01	0,00	0,04	-0,07	-0,05	0,00	0,00	-0,04	-0,01	0,03	-0,06	0,00	0,04
da	0,02	0,02	-0,02	0,01	-0,03	0,00	-0,02	0,01	0,05	0,03	0,03	0,00	0,00	-0,03
de	0,00	0,01	0,00	0,02	0,05	-0,10	0,01	0,04	-0,04	-0,02	0,05	0,00	0,04	-0,01
ek	0,04	-0,01	0,00	-0,01	0,02	-0,09	0,02	0,04	0,05	-0,03	-0,05	0,09	0,02	0,00
el	0,01	-0,05	0,00	0,01	-0,02	0,03	0,00	0,00	-0,07	-0,02	0,00	0,02	-0,06	0,00
em	0,02	-0,02	0,03	-0,01	0,01	0,01	0,02	0,00	-0,04	-0,03	-0,09	0,09	-0,08	-0,01
en	0,04	-0,01	-0,03	-0,06	-0,06	-0,05	-0,01	-0,06	0,01	-0,02	-0,03	0,01	-0,05	-0,03
er	-0,02	0,03	-0,05	0,00	0,02	0,08	0,02	-0,01	-0,04	-0,02	-0,01	-0,01	0,01	0,00
et	0,05	0,02	-0,03	-0,02	0,03	-0,05	-0,01	0,03	-0,02	0,00	-0,05	0,07	0,04	-0,02
ka	0,02	0,01	0,01	-0,05	-0,04	0,02	-0,04	-0,06	-0,05	0,07	0,06	0,00	-0,02	0,00
la	-0,04	-0,01	-0,02	0,01	0,02	-0,03	0,00	0,00	-0,05	0,02	0,00	0,00	0,00	0,02
le	-0,01	-0,02	0,02	-0,05	0,02	0,02	-0,03	-0,03	-0,03	-0,01	-0,01	0,00	-0,04	-0,02
len	-0,01	-0,06	0,03	-0,03	0,00	-0,05	-0,07	-0,05	-0,05	0,06	-0,02	-0,04	0,05	0,01
ler	0,02	0,01	-0,01	-0,05	0,04	0,03	0,03	0,05	-0,04	-0,06	-0,04	0,00	0,01	-0,02
ma	-0,06	0,00	-0,02	0,05	-0,05	0,04	0,01	0,00	-0,02	0,03	0,02	-0,03	-0,01	0,00
me	-0,03	-0,03	0,04	-0,12	-0,05	-0,04	0,01	-0,06	0,04	-0,01	0,01	-0,02	-0,03	0,01
nd	-0,09	0,01	-0,02	0,00	0,08	0,02	-0,05	-0,04	-0,05	0,03	-0,02	0,00	0,02	0,00
nm	-0,06	-0,02	0,02	0,06	-0,01	0,00	-0,02	-0,01	-0,01	0,05	0,03	-0,03	0,01	-0,01
ol	-0,01	0,02	-0,05	-0,01	0,03	0,00	0,00	-0,01	0,00	-0,03	-0,03	-0,01	0,03	0,00
on	0,02	0,04	0,03	-0,05	0,03	-0,05	-0,05	-0,02	-0,02	0,02	0,01	0,02	0,03	0,04
or		-0,05	0,01	-0,03	-0,02	-0,03	0,01	0,04	0,09	-0,02	-0,06	0,00	0,03	-0,02
ra	-0,05		-0,03	0,01	0,02	0,02	-0,02	-0,01	-0,06	-0,03	-0,01	-0,01	0,02	0,00
rl	0,01	-0,03		-0,02	-0,02	-0,09	0,01	0,01	0,03	0,01	0,02	-0,02	-0,04	-0,04
sa	-0,03	0,01	-0,02		-0,01	0,00	0,01	0,04	0,03	0,02	0,04	-0,05	0,02	0,03
so	-0,02	0,02	-0,02	-0,01		0,07	-0,03	0,02	0,00	-0,05	0,00	0,04	0,04	-0,02
st	-0,03	0,02	-0,09	0,00	0,07		-0,04	-0,03	0,03	-0,04	0,00	0,01	-0,07	-0,02
sy	0,01	-0,02	0,01	0,01	-0,03	-0,04		0,07	0,03	-0,01	-0,02	-0,01	-0,02	-0,01
syo	0,04	-0,01	0,01	0,04	0,02	-0,03	0,07		0,05	-0,02	-0,01	0,02	0,02	-0,02
ta	0,09	-0,06	0,03	0,03	0,00	0,03	0,03	0,05		0,00	0,00	0,02	-0,02	-0,06
te	-0,02	-0,03	0,01	0,02	-0,05	-0,04	-0,01	-0,02	0,00		0,00	-0,01	0,01	-0,05
ter	-0,06	-0,01	0,02	0,04	0,00	0,00	-0,02	-0,01	0,00	0,00		-0,04	0,00	0,00
tk	0,00	-0,01	-0,02	-0,05	0,04	0,01	-0,01	0,02	0,02	-0,01	-0,04		-0,05	-0,03
yo	0,03	0,02	-0,04	0,02	0,04	-0,07	-0,02	0,02	-0,02	0,01	0,00	-0,05		0,06
yon	-0,02	0,00	-0,04	0,03	-0,02	-0,02	-0,01	-0,02	-0,06	-0,05	0,00	-0,03	0,06	