ORDER RELATIONS FOR THE CRYPTANALYSIS OF SUBSTITUTION CIPHERS ON THE BASIS OF LINGUISTIC DATA STRUCTURES AS AN OPTIMAL STRATEGY

Norman Neukel

HIKARI LTD

HIKARI LTD

Hikari Ltd is a publisher of international scientific journals and books.

www.m-hikari.com

Norman Neukel, *Order Relations for the Cryptanalysis of Substitution Ciphers on the Basis of Linguistic Data Structures as an Optimal Strategy*

Dr. Norman Neukel
Ministry of Defence
53501 Grafschaft-Gelsdorf
Germany

ISBN 978-954-91999-0-1 (print)
ISBN 978-619-91397-0-7 (online)

Typeset using MS Word.

Mathematics Subject Classification: 03B65, 68T50, 68P25, 90B40, 91A35.

Keywords: Linguistics, cryptanalysis, optimization.

Published by Hikari Ltd

PREFACE

In order to analyze the solution behavior of general substitution ciphers for cryptanalysis, for the first time this book utilizes the methods of set optimization. The book investigates how to proceed optimally in a specific case. Linguistic data structures, such as the frequencies of the different letters, bigrams and trigrams, are divided into sets and described by order relations appropriate to the German and English languages. In the context of cryptanalysis this allows new solutions of existing models within the scope of set-valued order relations. This approach leads to an optimal strategy.

Dr. Norman Neukel
7 June, 2019

Contents

1. Introduction

Set optimization is generally concerned with the investigation of set-valued mapping and comparative set relations. In the set-valued case this is equivalent to the extension of vector optimization. In spite of the fact that the property above acts very naturally, research in this area is very limited. Theoretical mathematical studies are based for example on a Hausdorff metric and a directional derivative in Ha [8]. Vui, Anh and Wangkeeree [23] investigate optimization models and their relationships to well-definedness according to Levitin-Polyak and to stability. The results of academic research in the area of mathematical order relations are described e.g. by Karaman, Soyertem, Atasever Güvenç, Tozkan, M. Küçük and Y. Küçük [9].

Neukel [16] describes practical applications of set optimization in which emission influences lead to urban conflict situations which are investigated on the basis of order relations. For the determination of publicly registered land values Neukel [12] utilizes interval-valued duality with a regression model, resulting in an optimum describing the linking of an interval-valued and a set-valued approach. With the help of a dual equivalence class model extended by an order relation, Neukel [14] develops a decision strategy in the form of an adjusted data set. In practice these models pose new questions and offer new methods for research in other fields of mathematic as well.

In the modern age of communication cryptographic procedures assume central importance in relation to questions of data integrity. It is in fact known that the encoding of information was already practiced in antiquity. Examples known to us are the Scytale of the ancient Greeks and Caesar's cipher, named after its user. The aim of a cryptographic attack can be the direct determination of the plain text from the encoded text – the optimal case. The difficulty for the cryptanalyst is that he has only the encoded text (cipher text only attack). In addition he will have statistical information available about the contents of the communication and may know a few words that could occur in plain text. Under these "classical" conditions it is possible to decipher such codes.

The various publications in the cryptanalysis of substitution ciphers consider a linguistic area purely statistically, together with its properties and its possibilities. In this book optimal sets will be found utilizing sets of maximal and minimal elements and set-valued order relations with respect to self-chosen linguistic-statistic properties. This represents a new approach for information and decision making.

This book is structured as follows: Section 2 assigns definitions and prerequisites with regard to a preorder and partial order to known set order relations. Order relations are considered in a power set on the basis of the "set less" order relation with a natural reference criterion. For other order relations a

comparison with non-empty minima and maxima is defined. Section 3 describes encoding and cryptanalysis methods for the example of the classical method of encryption on the basis of an optimal strategy making use of order relations. The result requires the utilization of new set structures.

2. Set-valued order relations as an optimization approach in cryptanalysis

In the area of set optimization, scientifically oriented publications in mathematics deal with the optimality concept of the minimizer and its variants. However, this definition for set-valued mapping considers only a minimal element of the set and then defines this set as the "best". In general, however, a minimal element does not mean that the entire set is minimal for all sets, and this leads to specific order relations with optimal solutions. Their structure needs the following definition.

Definition 2.1
Let X be a vector space and $A \subset X$.
(a) A is called *convex* if and only if $\{z \in X | z := tx + (1-t)y,\ t \in [0,1]\} \subset A$ holds $\forall\, x, y \in A$.
(b) A is called a *cone* if and only if $tA \subset A\ \forall\, t \geq 0$.
(c) A is called a *convex cone* if and only if A is a cone and convex as a set.

This section discusses order relations in regard to a power set and a non-empty set of minimal and maximal elements. See also Neukel [13], [14] and [16]:

$$\mathcal{P}(Y) := \{A \subset Y | A \text{ is non-empty}\} \text{ with } Y \text{ as an arbitrary vector space}$$

and

$$\mathcal{M} := \{A \in \mathcal{P}(Y) | \operatorname{Min}(A, \preccurlyeq) \text{ and } \operatorname{Max}(A, \preccurlyeq) \text{ are non-empty}\}.$$

With regard to $\mathcal{M}$ the following definitions are required. Furthermore, the known order relations used in this book will now be introduced.

Definition 2.2
Let Q be an arbitrary non-empty set with a binary relation $\preccurlyeq$ and
$A, B, C, D \in Q$ be arbitrarily chosen. This binary relation $\preccurlyeq$ is
(a) *reflexive* when $A \preccurlyeq A$.
(b) *transitive* when it follows from $A \preccurlyeq B$ and $B \preccurlyeq D$ that $A \preccurlyeq D$.
(c) *antisymmetric* when $A \preccurlyeq B, B \preccurlyeq A \Rightarrow A = B$.
(d) *a preorder* when it is both reflexive and transitive.
(e) *a partial order* when it is reflexive, transitive and antisymmetric.

Definition 2.3
Let $\mathcal{A}$ be an non-empty set with a preorder $\preccurlyeq$ and $\bar{A} \in \mathcal{A}$.

(a) $\bar{A}$ is a *minimal element* of $\mathcal{A}$ when

$$A \preccurlyeq \bar{A} \text{ for } A \in \mathcal{A} \Rightarrow \bar{A} \preccurlyeq A.$$

(b) $\bar{A}$ is a *maximal element* of $\mathcal{A}$ when

$$\bar{A} \preccurlyeq A \text{ for } A \in \mathcal{A} \Rightarrow A \preccurlyeq \bar{A}.$$

In this book the following abbreviations will be used for the set of the minimal and maximal elements of $\mathcal{A}$:

$$\operatorname{Min}(\mathcal{A}, \preccurlyeq) := \left\{\overline{A} \in \mathcal{A} \middle| \bar{A} \text{ is a minimal element of } \mathcal{A} \text{ w. r. t. } \preccurlyeq\right\},$$
$$\operatorname{Max}(\mathcal{A}, \preccurlyeq) := \left\{\overline{A} \in \mathcal{A} \middle| \bar{A} \text{ is a maximal element of } \mathcal{A} \text{ w. r. t. } \preccurlyeq\right\}.$$

$\leq$ describes a preorder that is compatible with the linear structure.

Definition 2.3
Let $A, B \in \mathcal{P}(Y)$ be arbitrarily chosen sets.
(a) The *set less* or *Kuroiwa-Nishnianidze-Young (KNY)* order relation $\preccurlyeq_s$ is defined as

$$A \preccurlyeq_s B :\Leftrightarrow (\forall a \in A\ \exists b \in B: a \leq b) \text{ and } (\forall b \in B\ \exists a \in A: a \leq b).$$

(b) The *l-type less* order relation $\preccurlyeq_l$ has the form

$$A \preccurlyeq_l B :\Leftrightarrow (\forall b \in B\ \exists a \in A: a \leq b).$$

(c) The *u-type less* order relation $\preccurlyeq_u$ is defined as

$$A \preccurlyeq_u B :\Leftrightarrow (\forall a \in A\ \exists b \in B: a \leq b).$$

This order relation $\preccurlyeq_s$ was introduced independently of each other by Young [24] in algebra and by Nishnianidze [17] in fixed point theory. Kuroiwa, Tanaka and Ha [11] describe the set less order relation for the first time as a natural criterion. Scientifically oriented "real life" applications relating to set less are elaborated in Neukel [13], [14] and [16].

Chiriaev and Walster [3] employ this order relation in interval arithmetic and SUN Microsystems [22] implements this concept in the f95 FORTRAN compiler. Neukel elaborates a new socio-economic application in [12].

In general $\preccurlyeq_s$ is not antisymmetric (see Neukel [13]).

The definition of the set less order relation considers specific elements of the set *A* and certain elements of the set *B* (see Definition 2.4) and then characterizes these. Other elements can be virtually "freely" distributed. An infinite number of array models can occur. This poses the question how and whether it is necessary to consider various outlier values in a dominating comparison.

A stronger order relation than the last mentioned is described as *certainly less*.

Definition 2.5
Let $A, B \in \mathcal{P}(Y)$ be arbitrarily chosen sets. Then the *certainly less* order relation $\preccurlyeq_c$ is defined as

$$A \preccurlyeq_c B :\Leftrightarrow (A = B) \text{ or } (A \neq B, \forall a \in A\ \forall b \in B: a \leq b).$$

In the following a concept with weak operators will be presented.

Definition 2.6
Let $A, B \in \mathcal{P}(Y)$ be arbitrarily chosen sets. The *possibly less* order relation $\preccurlyeq_p$ is then defined as

$$A \preccurlyeq_p B :\Leftrightarrow (\exists a \in A\ \exists b \in B: a \leq b).$$

Possibly less therefore means that one set is "possibly" smaller than the other. $\preccurlyeq_p$ is not transitive (see Definition 2.2, (b) with e.g. $A := \{4\}$, $B := \{2, 3, 5\}$ and $D := \{3\}$ as subsets of the natural numbers $\mathbb{N}$ with the natural order $\leq$.

For arbitrary sets $A, B \in \mathcal{P}(Y)$ and $C_Y \subset Y$ as a convex cone the following properties apply:
(a) $A \preccurlyeq_l B \Leftrightarrow B \subset A + C_Y$.
(b) $A \preccurlyeq_u B \Leftrightarrow A \subset B - C_Y$.
(c) $A \preccurlyeq_s B \Leftrightarrow B \subset A + C_Y$ and $A \subset B - C_Y$.
(d) $A \preccurlyeq_c B \Leftrightarrow A = B$ or with $B - A \subset C_Y$.
(e) $A \preccurlyeq_p B \Leftrightarrow A \cap (B - C_Y) \neq \emptyset \Leftrightarrow B \cap (A + C_Y) \neq \emptyset$.

The following order relations utilize the minimal and maximal elements of a set. In a topologically real linear space Y minimal and maximal elements exist for every compact set in $\mathcal{P}(Y)$.

With the inclusion of minimal and maximal elements this leads to the definition of the *minmax less* order relation. Here we examine the sets A of $\mathcal{M}$, where $\mathcal{M}$ is assumed to be non-empty.

Definition 2.7
Let $A, B \in \mathcal{M}$ be arbitrarily chosen sets. The definition of the *minmax less* order relation $\preccurlyeq_m$ is then:

$$A \preccurlyeq_m B :\Leftrightarrow \operatorname{Min}(A, \preccurlyeq_s) \preccurlyeq_s \operatorname{Min}(B, \preccurlyeq_s) \text{ and } \operatorname{Max}(A, \preccurlyeq_s) \preccurlyeq_s \operatorname{Max}(B, \preccurlyeq_s).$$

Here the subscript *m* stands for the concept of *minmax*.

Definition 2.8
Let $A, B \in \mathcal{M}$ be arbitrarily chosen sets. The *minmax certainly less* order relation $\preccurlyeq_{mc}$ is then defined as:

$$A \preccurlyeq_{mc} B :\Leftrightarrow$$
$$(A = B) \text{ or}$$
$$\big(A \neq B, \operatorname{Min}(A, \preccurlyeq_c) \preccurlyeq_c \operatorname{Min}(B, \preccurlyeq_c) \text{ and } \operatorname{Max}(A, \preccurlyeq_c) \preccurlyeq_c \operatorname{Max}(B, \preccurlyeq_c)\big).$$

Here the lower index *mc* stands for the concept of *minmax certainly*.

In practice the *minmax certainly less* order relation is extensive. $\mathrm{Min}(A, \preccurlyeq_c)$ must in any case be smaller than $\mathrm{Min}(B, \preccurlyeq_c)$. The properties of the sets of maximal elements are equivalent to those of the sets with minimal elements.

Section 3 will introduce entirely new cryptanalytical evidence for the application of the above order relations.

An example of a linguistic order preference making use of fuzzy set theory and fuzzy logic is given in Sengupta and Pal [21]. This follows from the pessimistic viewpoint of the decision maker. A risk instability is the basic characteristic of pessimistic decision makers, i.e. they act according to the principle "more uncertainty is worse than less uncertainty".

In the following $\leq$ refers to the range of real numbers. Following Sengupta and Pal [21] let us now define $A := [a_L, a_R] = \{x \mid a_L \leq x \leq a_R,\ x \in \mathbb{R}\}$ as a *(closed) interval*, in which $a_L, a_R \in \mathbb{R}$ are the lower limits and the upper limits of A. Here every real number $a \in \mathbb{R}$ is expressed as an interval $[a, a] = \{a\}$ with no length, since $a - a = 0$. An interval can be further represented by its midpoint $a_C := \frac{a_L + a_R}{2}$ and radius $a_W := \frac{a_R - a_L}{2}$:

$$A = [a_L, a_R] \overset{\text{Interval property}}{\widehat{=}} \{x \mid |x - a_C| \leq a_W\} = \left\{x \,\middle|\, \left|x - \tfrac{a_L + a_R}{2}\right| \leq \tfrac{a_R - a_L}{2}\right\}.$$

Case 1 and Case 2 for the elimination of the absolute value lead to the interval definition:

Case 1:

$$x - \frac{a_L + a_R}{2} \leq \frac{a_R - a_L}{2}$$

$$\Leftrightarrow\ x \leq \frac{a_R - a_L}{2} + \frac{a_L + a_R}{2}$$

$$\Leftrightarrow\ x \leq \frac{2a_R}{2} = a_R.$$

Case 2:

$$-x + \frac{a_L + a_R}{2} \leq \frac{a_R - a_L}{2}$$

$$\Leftrightarrow\ -x \leq \frac{a_R - a_L}{2} - \frac{a_L + a_R}{2}$$

$$\Leftrightarrow\ -x \leq \frac{-2a_L}{2} = -a_L$$

$$\Leftrightarrow\ x \geq a_L.$$

With the corresponding definition of arithmetic, on a PC the interval representation often has advantages in terms of speed, but due to rounding errors yields only coarse barriers.

Similarly, one defines *open intervals* $(\cdot,\cdot)$ with $<$ (this is consistent with the definition of real order relations $\leq$ with $\neq$) and *half-open intervals* with $[\cdot,\cdot)$ with $\leq$ and $<$.

Definition 2.9
Let $\mathbb{IR}$ be the set of all closed intervals of the real numbers $\mathbb{R}$. The *acceptance function* $\mathcal{A}: \mathbb{IR} \times \mathbb{IR} \longrightarrow [0,\infty)$ for the intervals $A, B \in \mathbb{IR}$ with $a_C \leq b_C$ is designated $\mathcal{A}(A,B)$ and is defined as

$$\mathcal{A}(A,B) := \frac{b_C - a_C}{b_W + a_W}$$

with $b_W + a_W \neq 0$. $\mathcal{A}(A,B)$ then describes the degree of acceptance for an interval comparison of the form "the interval A is minimal compared with interval B".

In the following the sets of the intervals with $\mathcal{A}(A,B) \geq 0$ and $a_W < b_W$ will be treated as in Sengupta and Pal [21] and defined:
(a) When $\mathcal{A}(A,B) \geq 1 \Rightarrow B$ is strongly preferred compared to A.
(b) When $\mathcal{A}(A,B) = 0 \Rightarrow A$ is strongly preferred compared to B.
(c) When $\mathcal{A}(A,B) \in (0,1) \Rightarrow$ a fuzzy preference exists with regard to A or B.

Definition 2.10
The interval descriptions above apply.
(a) The *fuzzy set* $\bar{B}$ is defined as $\bar{B} := \{(A,B) | \mathcal{A}(A,B) \geq 0 \text{ and } a_W < b_W\}$ and is a rejection of the set B.
(b) The *membership function* $\mu_{\bar{B}}(A)$ of the fuzzy set $\bar{B}$ is defined as follows:

$$\mu_{\bar{B}}(A) := \begin{cases} 1, & \text{when } a_C = b_C \\ \text{Max}\left(0, \dfrac{a_C - (b_L - a_W)}{b_C - (b_L - a_W)}\right), & \text{when } b_C > a_C > b_L + a_W \\ 0, & otherwise \end{cases}.$$

Max$(\cdot,\cdot)$ describes the usual maximum in real numbers and $A = [a_L, a_R]$ an interval.

The values of the membership function $\mu_{\bar{B}}(A)$ lie between 0 and 1. When $\mu_{\bar{B}}(A) = 1$, then B is definitively rejected. When $\mu_{\bar{B}}(A) = 0$, then B is definitively accepted.

The intensity of the pessimism influences the decision theory. One considers a linguistic set of different degrees of pessimism
$\{\dots, \text{very very low}, \text{very low}, \text{low}, \text{moderate}, \text{high}, \text{very high}, \text{very very high}, \dots\}$.

Sengupta and Pal [21] modify this fuzzy preference sequence according to a different numerical-linguistic pessimism.

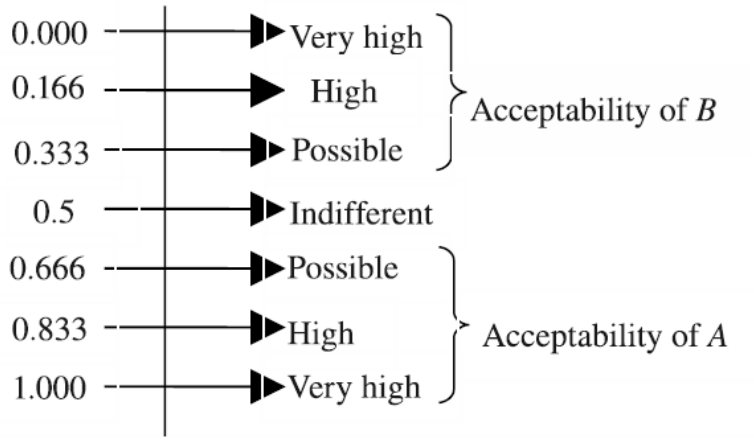

Figure 2.1: A linguistic measure of different degrees of pessimism after Sengupta and Pal [20], Taken from Karmakar and Bhunia [10]

A *modified nonlinear membership function* is defined as

$$\pi_{\bar{B}} := (\mu_{\bar{B}})^p,$$

where $p \in [1/m, m]$ is a sharp value with m a finite number. $\pi_{\bar{B}}$ gives the degree of pessimism of the parameters for an arbitrary decider. p can then assume the values $1/m$ or m when the decider is "absolutely" positive or "absolutely" negative with respect to B, respectively.

When $\pi_{\bar{B}} = (0.333, 0.666)$, as a consequence of the linguistic measure in Figure 2.1 the decider exhibits an indifferent behavior: it prefers neither A nor B. The application of these preference relations is dynamic for two arbitrary intervals. Since the modified nonlinear membership function is dependent on the sharp value p – the degree of pessimism – one obtains different values of the membership function for different degrees of pessimism. This results in the frequent fluctuation in the order of preference according to the values of p.

A description of relationships between interval arithmetic and set optimization is given in Neukel [12] and [13].

3. Classical substitution ciphers of cryptanalysis with set-valued order relations

In the cryptanalysis of a general substitution cipher even with a less extensive alphabet it is no longer possible to try out all codes. Nevertheless, an attack with known plain text can greatly reduce the number of codes coming into question. For example, if one knows the codes for five letters in the Latin alphabet the number of codes to be tested is reduced by the factor 10^7, i.e. from $26! = 4.03 \cdot 10^{26} \approx 2^{88}$ to $21! = 5.11 \cdot 10^{19}$. For the cryptanalysis of such a substitution cipher the knowledge of the cipher text alone is sufficient if the coded message possesses $\preccurlyeq$-redundancy and enough cipher text is available.

Definition 3.1
A language is considered $\preccurlyeq$-*redundant* when there is a natural number r for which not all numerical sequences of length r occur with the same probability.

The concept of redundancy is definied differently in information theory than in this book.

Every natural language already has a high $\preccurlyeq$-redundancy for $r = 1$, i.e. the basis is the frequency distribution of the characters for the particular alphabet. This $\preccurlyeq$-redundancy increases with the length of the character sequences considered. According to Fumy and Rieß [5], in the case of a German or English communication one expects an unambiguous decoding of a simple substitution cipher having a length of 30 characters. This is described by the term *unicity distance*, see Shannon [19].

The frequency distribution of the characters in a natural language text is dependent on the particular language and, to a slight extent, on the content and type of the text. The following tables illustrate typical distributions for the German and English languages (see Bauer and Goos [1] and Beker and Piper [2]), in which the category * considers punctuation marks and blanks.

Since the punctuation marks and blanks contribute significantly to the $\preccurlyeq$-redundancy of a natural language, in practice these categorized symbols are removed from the plain text before encoding. The table below shows the adjusted rounded letter frequencies of Fumy and Rieß [5] for the two languages. Here $\alpha^* := \{*, \mathrm{a}, \mathrm{b}, \ldots, \mathrm{z}\}$ and $\alpha := \{\mathrm{a}, \mathrm{b}, \ldots, \mathrm{z}\}$.

	German		English	
*	15.14	-	19.23	-
a	04.58	05.40	06.60	08.15
b	01.60	01.89	01.21	01.49
c	02.67	03.15	02.25	02.78
d	04.39	05.17	03.43	04.25
e	15.35	18.10	10.26	12.70
f	01.36	01.60	01.80	02.23
g	02.67	03.15	01.63	02.02
h	04.36	05.14	04.92	06.09
i	06.38	07.52	05.63	06.97
j	00.16	00.19	00.12	00.15
k	00.96	01.13	00.62	00.77
l	02.93	03.45	03.25	04.03
m	02.13	02.51	01.94	02.41
n	08.84	10.42	05.45	06.75
o	01.90	02.24	06.06	07.51
p	00.50	00.59	01.56	01.93
q	00.01	00.01	00.08	00.10
r	06.86	08.08	04.84	05.99
s	05.39	06.35	05.11	06.33
t	04.73	05.57	07.31	09.06
u	03.48	04.10	02.23	02.76
v	00.74	00.87	00.79	00.98
w	01.42	01.67	01.91	02.36
x	00.01	00.01	00.12	00.15
y	00.02	00.02	01.59	01.97
z	01.42	01.67	00.06	00.07
total	**100.00**	**100.00**	**100.00**	**100.00**
$\mathrm{Min}(\alpha^*.\preccurlyeq)$	00.01		00.06	
$\mathrm{Min}(\alpha.\preccurlyeq)$		00.01		00.07
$\mathrm{Max}(\alpha^*.\preccurlyeq)$	15.35		19.23	
$\mathrm{Max}(\alpha.\preccurlyeq)$		18.10		12.70

Figure 3.1: Frequencies of letters and sets of the minimal and maximal elements

In German and in English the most frequently occurring letter is "e". Extreme differences in frequencies between the two languages are rare. The values for the letters "y" (approximately 100 times more frequent in English) and "z" (approximately 24 times more frequent in German) are exceptions. The distinctly lower frequency of interspaces and punctuation marks in German (see *) is a consequence of the higher average word length in the German language. The similarity of the frequency distributions for the two languages is also clearly shown in the classification of letters according to Fumy and Rieß [5] into set I := "very frequent", set II := "frequent", set III := "average", set IV := "relatively seldom", and set V := "very seldom" (see Figure 3.2).

Set	German	English
I	{e}	{e}
II	{n, r, i, s, t, d, h, a}	{t, a, o, i, n, s, h, r}
III	{u, l, c, g}	{d, l}
IV	{m, o, b, z, w, f}	{c, u, m, w, f, g, y, p, b}
V	{k, v, p, j, y, q, x}	{v, k, j, x, q, z}

Figure 3.2: Sets of letters

In addition to the frequency distribution of the individual characters one also considers the distributions of pairs of characters (so-called *bigrams*), character triplets (*trigrams*), etc. and takes these into account in the cryptanalysis. The statistical properties of a natural language provide useful data. When the word boundaries in the cipher text are known, various two- or three-letter words can then be quickly identified. One therefore attempts to identify frequently occurring prefixes or endings or to utilize certain or specific distributions of the initial and terminal letters. The most frequently occurring initial letters in English are "t, a, s, w, h, i, o, b, and c" and the most frequently occurring terminal letters are "e, d, s, t, n, r, y, m, and g" (see Beker and Piper [2]). A further starting point are the bigrams of identical letters. In English the letters "t, o, s, and e" are frequently doubled. Furthermore, specific patterns exist, e.g. in "b-an-an-a". Finally, there are different methods to distinguish vowels from consonants. An indication here is the fact that vowels have a broader range of neighboring characters than consonants.

It is evident that especially shorter texts need not necessarily possess all the characteristic properties of the particular language discussed up to this point. Longer texts can also deviate significantly from these statistics (see Perec [18], „*Anton Voyls Fortgang*“, a 363 page long novel in which the letter "e" does not occur. On the other hand, however, it is not possible to avoid every characteristic of speech and still express oneself in this language.

The statistical cryptanalysis of the German and English languages discussed up to now will now be substantiated on the basis of new set-valued order relation based algorithms.

init 1 Set-valued order relations.

2 Letter frequencies determined for $a, \ldots, z$ with the German and English languages.

while Allocation of letter frequencies **do**

1 Selection of letters as sets to be compared between German and English after Bauer and Goos [1] and Beker and Piper [2].

2 Determination of optimal sets utilizing order relations for comparing:
- 2.1 in the German language
- 2.2 in the English language
- 2.3 in the German and English languages.

return Optimal solutions according to the specific properties of the order relations in the linguistic context as a strategy concept.

init was already carried out in Section 2 and in this section. With regard to **while**, 1, the classification of letters according to Bauer and Goos [1] and Beker and Piper [2] applies, with the following definition: $a := 1, b := 2, \ldots, z := 26$.

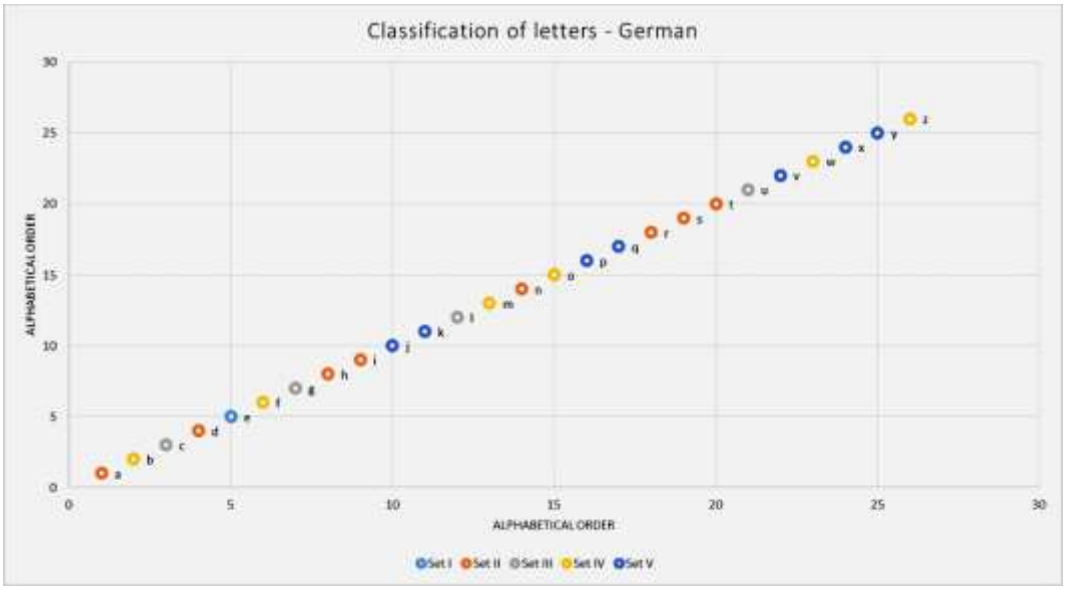

Figure 3.3: Sets of letters for the German language

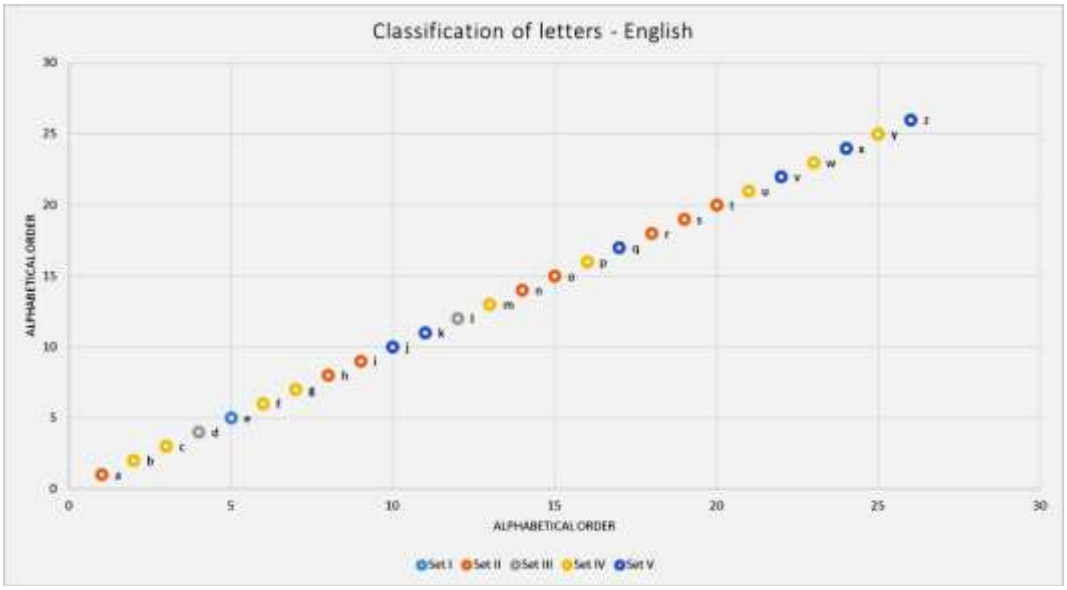

Figure 3.4: Sets of letters for the English language

The letter sets I, II, III, IV and V are subsequently compared with the introduced order relations. The result is subject to special letters finding an optimal set value. Thus, a sequence and evaluation based on these order relations can be created.

For **while**, 2, 2.1 the following new results for the statistics of the German language apply:

Set	Order relation	Set	Special case
I	$\preccurlyeq_p, \succcurlyeq_p$	II	$\mathrm{Min}(\mathrm{II}, \preccurlyeq_c) \preccurlyeq_c \mathrm{Min}(\mathrm{I}, \preccurlyeq_c)$, $\mathrm{Max}(\mathrm{I}, \preccurlyeq_c) \preccurlyeq_c \mathrm{Max}(\mathrm{II}, \preccurlyeq_c)$
I	$\preccurlyeq_p, \succcurlyeq_p$	III	$\mathrm{Min}(\mathrm{III}, \preccurlyeq_c) \preccurlyeq_c \mathrm{Min}(\mathrm{I}, \preccurlyeq_c)$, $\mathrm{Max}(\mathrm{I}, \preccurlyeq_c) \preccurlyeq_c \mathrm{Max}(\mathrm{III}, \preccurlyeq_c)$
I	$\preccurlyeq_p, \succcurlyeq_p$	IV	$\mathrm{Min}(\mathrm{IV}, \preccurlyeq_c) \preccurlyeq_c \mathrm{Min}(\mathrm{I}, \preccurlyeq_c)$, $\mathrm{Max}(\mathrm{I}, \preccurlyeq_c) \preccurlyeq_c \mathrm{Max}(\mathrm{IV}, \preccurlyeq_c)$
I	$\preccurlyeq_c$	V	
II	$\preccurlyeq_{mc}$	III	
II	$\preccurlyeq_{mc}$	IV	
II	$\preccurlyeq_{mc}$	V	
III	$\preccurlyeq_u, \succcurlyeq_l$	IV	$\mathrm{Min}(\mathrm{IV}, \preccurlyeq_c) \preccurlyeq_c \mathrm{Min}(\mathrm{III}, \preccurlyeq_c)$, $\mathrm{Max}(\mathrm{III}, \preccurlyeq_c) \preccurlyeq_c \mathrm{Max}(\mathrm{IV}, \preccurlyeq_c)$
III	$\preccurlyeq_{mc}$	V	
IV	$\preccurlyeq_l, \succcurlyeq_u$	V	$\mathrm{Min}(\mathrm{IV}, \preccurlyeq_c) \preccurlyeq_c \mathrm{Min}(\mathrm{V}, \preccurlyeq_c)$, $\mathrm{Max}(\mathrm{V}, \preccurlyeq_c) \preccurlyeq_c \mathrm{Max}(\mathrm{IV}, \preccurlyeq_c)$

Figure 3.5: Order relations for the German language

The stronger the order relation, the more optimal (in the meaning of minimal, practical, favorable or simple) the allocation of the different letters. The order relations $\preccurlyeq_{mc}$ and $\preccurlyeq_c$ give the strongest separation of the sets of letters and give the "best" optimum as the decision strategy for **return**. In contrast the set comparisons with $\preccurlyeq_l$, $\preccurlyeq_u$ and $\preccurlyeq_p$ offer specific optima for the sets of minimal and maximal elements. Furthermore, bigram and trigram statistics can also be evaluated as word patterns for initial letters and terminal letters of words or for groups of letters. Minimal elements take into account the order of the alphabet.

The table below shows, for example, that "ch" can be "practically" allocated due to $\preccurlyeq_{mc}$ and the low rank. According to **return,** "ch" is an optimal solution. In the German language "ch" is the optimal bigram followed by "un", "ic", "ng" and "au". It is apparent that set V is not found in these bigrams. The allocation of rank is taken from Bauer and Goos [1] and Beker and Piper [2]. From rank 20 it is increasingly difficult to find order relations.

Rank	Bigram	Set membership of bigram letters	Order relations of the sets of letters up to now, see figure 3.5
1	en	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
2	er	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
3	ch	II, III	$\preccurlyeq_{mc}$
4	nd	II	-
5	ei	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
6	de	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
7	in	II	-
8	es	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
9	te	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
10	ie	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
11	un	II, III	$\preccurlyeq_{mc}$
12	ge	I, III	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
13	st	II	-
14	ic	II, III	$\preccurlyeq_{mc}$
15	he	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
16	ne	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
17	se	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
18	ng	II, III	$\preccurlyeq_{mc}$
19	re	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
20	au	II, III	$\preccurlyeq_{mc}$
21	di	II	-
22	be	I, IV	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
23	ss	II	-
24	ns	II	-
25	an	II	-
26	si	II	-
27	ue	I. III	$\preccurlyeq_p$. $\succcurlyeq_p$, special case
28	da	II	-
29	as	II	-
30	ni	II	-

Figure 3.6: New bigram statistics for the German language

Rank	Trigram	Set membership of bigram letters	Order relations of the sets of letters up to now, see figure 3.5
1	ein	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
2	ich	II, III	$\preccurlyeq_{mc}$
3	nde	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
4	die	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
5	und	II, III	$\preccurlyeq_{mc}$
6	der	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
7	che	I, II, III	$\preccurlyeq_p$, $\succcurlyeq_p$, special case, $\preccurlyeq_{mc}$
8	end	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
9	gen	I, II, III	$\preccurlyeq_p$, $\succcurlyeq_p$, special case, $\preccurlyeq_{mc}$
10	sch	II, III	$\preccurlyeq_{mc}$
11	cht	II, III	$\preccurlyeq_{mc}$
12	den	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
13	ine	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
14	nge	I, II, III	$\preccurlyeq_p$, $\succcurlyeq_p$, special case, $\preccurlyeq_{mc}$
15	nun	II, III	$\preccurlyeq_{mc}$
16	ung	II, III	$\preccurlyeq_{mc}$
17	das	II	-
18	hen	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
19	ind	II	-
20	enw	I, II, IV	$\preccurlyeq_p$, $\succcurlyeq_p$, special case, $\preccurlyeq_{mc}$
21	ens	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
22	ies	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
23	ste	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
24	ten	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
25	ere	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
26	lic	II, III	$\preccurlyeq_{mc}$
27	ach	II, III	$\preccurlyeq_{mc}$
28	ndi	II	-
29	sse	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
30	aus	II, III	$\preccurlyeq_{mc}$

Figure 3.7: New trigram statistics for the German language

In Figure 3.7 set V does not occur and set IV only once. The allocation of rank is again taken from Bauer and Goos [1] and Beker and Piper [2]. An optimal trigram for the cryptanalysis is, for example, "ich" (rank 2) or "und" (rank 5) – both show the order relation $\preccurlyeq_{mc}$.

For **while**, 2, 2.2 it follows that one now obtains new results for the statistics of the English language:

Set	Order relation	Set	Specifically
I	$\preccurlyeq_p, \succcurlyeq_p$	II	$\mathrm{Min}(\mathrm{II}, \preccurlyeq_c) \preccurlyeq_c \mathrm{Min}(\mathrm{I}, \preccurlyeq_c)$, $\mathrm{Max}(\mathrm{I}, \preccurlyeq_c) \preccurlyeq_c \mathrm{Max}(\mathrm{II}, \preccurlyeq_c)$
I	$\preccurlyeq_p, \succcurlyeq_p$	III	$\mathrm{Min}(\mathrm{III}, \preccurlyeq_c) \preccurlyeq_c \mathrm{Min}(\mathrm{I}, \preccurlyeq_c)$, $\mathrm{Max}(\mathrm{I}, \preccurlyeq_c) \preccurlyeq_c \mathrm{Max}(\mathrm{III}, \preccurlyeq_c)$
I	$\preccurlyeq_p, \succcurlyeq_p$	IV	$\mathrm{Min}(\mathrm{IV}, \preccurlyeq_c) \preccurlyeq_c \mathrm{Min}(\mathrm{I}, \preccurlyeq_c)$, $\mathrm{Max}(\mathrm{I}, \preccurlyeq_c) \preccurlyeq_c \mathrm{Max}(\mathrm{IV}, \preccurlyeq_c)$
I	$\preccurlyeq_c$	V	
II	$\preccurlyeq_l, \succcurlyeq_u$	III	$\mathrm{Min}(\mathrm{II}, \preccurlyeq_c) \preccurlyeq_c \mathrm{Min}(\mathrm{III}, \preccurlyeq_c)$, $\mathrm{Max}(\mathrm{III}, \preccurlyeq_c) \preccurlyeq_c \mathrm{Max}(\mathrm{II}, \preccurlyeq_c)$
II	$\preccurlyeq_{mc}$	IV	
II	$\preccurlyeq_{mc}$	V	
III	$\preccurlyeq_u, \succcurlyeq_l$	IV	$\mathrm{Min}(\mathrm{IV}, \preccurlyeq_c) \preccurlyeq_c \mathrm{Min}(\mathrm{III}, \preccurlyeq_c)$, $\mathrm{Max}(\mathrm{III}, \preccurlyeq_c) \preccurlyeq_c \mathrm{Max}(\mathrm{IV}, \preccurlyeq_c)$
III	$\preccurlyeq_{mc}$	V	
IV	$\preccurlyeq_{mc}$	V	

Figure 3.8: Order relations for the English language

Figures 3.5 and 3.8 illustrate the (only minor) differences between German and English. In 8 of 10 categories of comparison an analogous result is obtained. For set V in English exclusively strong order relations $\preccurlyeq_c$ and $\preccurlyeq_{mc}$ are applied. The result is a decision strategy with respect to **return**.

In the following the bigram and trigram statistics of the English language are evaluated.

The table below shows that in English there are significantly less minimal elements than in German. The order relation $\preccurlyeq_{mc}$ occurs in only 10 % of the cases in Figure 3.9 – the first time at position 17. For these bigrams there is no set V. The allocations of rank are taken from Bauer and Goos [1] and Beker and Piper [2]. Compared with Figure 3.6, in English it is difficult from the very beginning to find order relations.

Rank	Bigram	Set membership of bigram letters	Order relations of the sets of letters up to now, see figure 3.8
1	th	II	-
2	he	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
3	in	II	-
4	er	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
5	an	II	-
6	re	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
7	ed	I, III	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
8	on	II	-
9	es	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
10	st	II	-
11	en	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
12	at	II	-
13	to	II	-
14	nt	II	-
15	ha	II	-
16	nd	II, III	$\preccurlyeq_l$, $\succcurlyeq_u$, special case
17	ou	II, IV	$\preccurlyeq_{mc}$
18	ea	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
19	ng	II, IV	$\preccurlyeq_{mc}$
20	as	II	-
21	or	II	-
22	ti	II	-
23	is	II	-
24	et	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
25	it	II	-
26	ar	II	-
27	te	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
28	se	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
29	hi	II	-
30	of	II, IV	$\preccurlyeq_{mc}$

Figure 3.9: New bigram statistics for the English language

Rank	Trigram	Set membership of trigram letters	Order relations of the sets of letters up to now, see figure 3.8
1	the	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
2	ing	II, IV	$\preccurlyeq_{mc}$
3	and	II, III	$\preccurlyeq_l$, $\succcurlyeq_u$, special case
4	her	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
5	ere	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
6	ent	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
7	tha	II	-
8	nth	II	-
9	was	II, IV	$\preccurlyeq_{mc}$
10	eth	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
11	for	II, IV	$\preccurlyeq_{mc}$
12	dth	II, III	$\preccurlyeq_l$, $\succcurlyeq_u$, special case
13	hat	II	-
14	she	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
15	ion	II	-
16	int	II	-
17	his	II	-
18	sth	II	-
19	ers	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
20	ver	I, II, V	$\preccurlyeq_p$, $\succcurlyeq_p$, special case, $\preccurlyeq_c$, $\preccurlyeq_{mc}$
21	tth	II	-
22	ter	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
23	hes	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
24	edt	I, II, III	$\preccurlyeq_p$, $\succcurlyeq_p$, special case, $\preccurlyeq_l$, $\succcurlyeq_u$
25	est	I, II	$\preccurlyeq_p$, $\succcurlyeq_p$, special case
26	thi	II	-
27	had	II, III	$\preccurlyeq_l$, $\succcurlyeq_u$, special case
28	oth	II	-
29	all	II, III	$\preccurlyeq_l$, $\succcurlyeq_u$, special case
30	ati	II	-

Figure 3.10: New trigram statistics for the English language

The allocation of rank in Figure 3.10 is taken once again from Bauer and Goos [1] and Beker and Piper [2]. Here, an optimal trigram in the English language is "ing", with the order relation $\preccurlyeq_{mc}$ for the sets of the minimal and maximal elements. The trigram "ver" even contains the order relation $\preccurlyeq_c$. All sets I, II, III, IV and V are utilized by the English trigram letters.

In the following table the indices *D* and *E* for the particular set refer to the German and English languages, respectively. The distribution of letters in Figure 3.2 applies again. For the German – English comparison with **while**, 2, 2.3:

Set	Order relation	Set	Specifically
I_D	$=$	I_E	
I_D	$\preccurlyeq_p, \succcurlyeq_p$	II_E	$\mathrm{Min}(\mathrm{II}_E, \preccurlyeq_c) \preccurlyeq_c \mathrm{Min}(\mathrm{I}_D, \preccurlyeq_c)$, $\mathrm{Max}(\mathrm{I}_D, \preccurlyeq_c) \preccurlyeq_c \mathrm{Max}(\mathrm{II}_E, \preccurlyeq_c)$
I_D	$\preccurlyeq_p, \succcurlyeq_p$	III_E	$\mathrm{Min}(\mathrm{III}_E, \preccurlyeq_c) \preccurlyeq_c \mathrm{Min}(\mathrm{I}_D, \preccurlyeq_c)$, $\mathrm{Max}(\mathrm{I}_D, \preccurlyeq_c) \preccurlyeq_c \mathrm{Max}(\mathrm{III}_E, \preccurlyeq_c)$
I_D	$\preccurlyeq_p, \succcurlyeq_p$	IV_E	$\mathrm{Min}(\mathrm{IV}_E, \preccurlyeq_c) \preccurlyeq_c \mathrm{Min}(\mathrm{I}_D, \preccurlyeq_c)$, $\mathrm{Max}(\mathrm{I}_D, \preccurlyeq_c) \preccurlyeq_c \mathrm{Max}(\mathrm{IV}_E, \preccurlyeq_c)$
I_D	$\preccurlyeq_c$	V_E	
II_D	$\preccurlyeq_p, \succcurlyeq_p$	I_E	$\mathrm{Min}(\mathrm{II}_D, \preccurlyeq_c) \preccurlyeq_c \mathrm{Min}(\mathrm{I}_E, \preccurlyeq_c)$, $\mathrm{Max}(\mathrm{I}_E, \preccurlyeq_c) \preccurlyeq_c \mathrm{Max}(\mathrm{II}_D, \preccurlyeq_c)$
II_D	$\preccurlyeq_{mc}, \succcurlyeq_{mc}$	II_E	$\mathrm{Min}(\mathrm{II}_D, \preccurlyeq_c) = \mathrm{Min}(\mathrm{II}_E, \preccurlyeq_c)$, $\mathrm{Max}(\mathrm{II}_D, \preccurlyeq_c) = \mathrm{Max}(\mathrm{II}_E, \preccurlyeq_c)$
II_D	$\preccurlyeq_l, \succcurlyeq_u$	III_E	$\mathrm{Min}(\mathrm{II}_D, \preccurlyeq_c) \preccurlyeq_c \mathrm{Min}(\mathrm{III}_E, \preccurlyeq_c)$, $\mathrm{Max}(\mathrm{III}_E, \preccurlyeq_c) \preccurlyeq_c \mathrm{Max}(\mathrm{II}_D, \preccurlyeq_c)$
II_D	$\preccurlyeq_{mc}$	IV_E	
II_D	$\preccurlyeq_{mc}$	V_E	
III_D	$\preccurlyeq_p, \succcurlyeq_p$	I_E	$\mathrm{Min}(\mathrm{III}_D, \preccurlyeq_c) \preccurlyeq_c \mathrm{Min}(\mathrm{I}_E, \preccurlyeq_c)$, $\mathrm{Max}(\mathrm{I}_E, \preccurlyeq_c) \preccurlyeq_c \mathrm{Max}(\mathrm{III}_D, \preccurlyeq_c)$
III_D	$\succcurlyeq_{mc}$	II_E	
III_D	$\preccurlyeq_{mc}$	III_E	$\mathrm{Max}(\mathrm{III}_D, \preccurlyeq_c) = \mathrm{Max}(\mathrm{III}_E, \preccurlyeq_c)$
III_D	$\preccurlyeq_u, \succcurlyeq_l$	IV_E	$\mathrm{Min}(\mathrm{IV}_E, \preccurlyeq_c) \preccurlyeq_c \mathrm{Min}(\mathrm{III}_D, \preccurlyeq_c)$, $\mathrm{Max}(\mathrm{III}_D, \preccurlyeq_c) \preccurlyeq_c \mathrm{Max}(\mathrm{IV}_E, \preccurlyeq_c)$
III_D	$\preccurlyeq_{mc}$	V_E	
IV_D	$\preccurlyeq_p, \succcurlyeq_p$	I_E	$\mathrm{Min}(\mathrm{IV}_D, \preccurlyeq_c) \preccurlyeq_c \mathrm{Min}(\mathrm{I}_E, \preccurlyeq_c)$, $\mathrm{Max}(\mathrm{I}_E, \preccurlyeq_c) \preccurlyeq_c \mathrm{Max}(\mathrm{IV}_D, \preccurlyeq_c)$
IV_D	$\succcurlyeq_{mc}$	II_E	
IV_D	$\preccurlyeq_l, \succcurlyeq_u$	III_E	$\mathrm{Min}(\mathrm{IV}_D, \preccurlyeq_c) \preccurlyeq_c \mathrm{Min}(\mathrm{III}_E, \preccurlyeq_c)$, $\mathrm{Max}(\mathrm{III}_E, \preccurlyeq_c) \preccurlyeq_c \mathrm{Max}(\mathrm{IV}_D, \preccurlyeq_c)$

Figure 3.11: Combined order relations for the German and English languages

IV_D	$\succcurlyeq_{mc}$	IV_E	$\mathrm{Min}(\mathrm{IV}_D, \preccurlyeq_c) = \mathrm{Min}(\mathrm{IV}_E, \preccurlyeq_c)$
IV_D	$\preccurlyeq_{mc}$	V_E	$\mathrm{Max}(\mathrm{IV}_D, \preccurlyeq_c) = \mathrm{Max}(\mathrm{V}_E, \preccurlyeq_c)$
V_D	$\succcurlyeq_c$	I_E	
V_D	$\succcurlyeq_{mc}$	II_E	
V_D	$\succcurlyeq_{mc}$	III_E	
V_D	$\succcurlyeq_{mc}$	IV_E	
V_D	$\preccurlyeq_{mc}$	V_E	

Figure 3.11 (continued): Combined order relations for the German and English languages

Figure 3.11 illustrates the linguistic differences between the German and English languages utilizing a new mathematical-scientific approach based on the multitude of strong order relations $\preccurlyeq_c$ and $\preccurlyeq_{mc}$. Comparisons with other languages result accordingly in another structure with regard to order relations. Viewed historically, German and English belong to the same family of languages, and English also belongs to the Germanic languages. Both languages use the same alphabet.

The classical substitution cipher in general utilizes an analog mapping from the plain text alphabet to the cipher alphabet for the encryption of every plain text character. Consequently, the statistical properties of the plain text characters are transferred to the allotted cipher text characters and the set-valued order relations describing the frequencies. Bigrams, trigrams etc. usually allow the successful cryptanalysis of a cipher text. These characteristic distributions can be adapted with the application of *homophonic* substitution ciphers. This is achieved by allocating a set of diverse arbitrary cipher text characters, and not a fixed cipher text character, to each plain text character. With each coding operation a selection is made from this set. Here the cordinality of the cipher text alphabet exceeds that of the plain text alphabet. An expansion of the cipher text – in practice usually not desirable – is thus tantamount to a price for improved reliability.

With a homophonic substitution cipher a set of cipher text characters is first allocated to every plain text character. The image sets for these mappings possess a pair-wise empty intersection. For the encryption a cipher text character is substituted (mapped) for each plain text character (mapped) that is chosen (possibly randomly) from the respective associated set. The above substitution ciphers are therefore homophonic with single-element image sets. Independently of this, homophonic substitution ciphers always attempt to expand the cipher text. Each cipher text character maps only one plain text letter, and the encryption is therefore unambiguous.

If one chooses the powerfulness of the image set for a character to be approximately proportional to the relative frequency of this character in the plain

text, a cryptanalyst will initially have little success on the basis of frequency statistics. However, it must be noted that a "flat" distribution of the cipher text characters by no means automatically results in corresponding distributions of the bigrams or trigrams: Inferences with regard to the plain text are in principle possible (see Günther [7]). The particular linguistic properties of the plain text always remain, independently of the homophonic substitution cipher. The utilization of order relations for such comparable investigations becomes more difficult.

Let an arbitrary homophonic substitution cipher possess $n \in \mathbb{N}$ different cipher text characters and the plain text be German or English without special characters. Assuming that each of the 26 plain text letters can be mapped to at least one cipher text character, the theoretical key space is of the magnitude

$$\binom{n}{26} \cdot 26! \cdot 26^{n-26}.$$

When $\binom{n}{k} < \frac{n^k}{k!}$ for $k = 26$ is estimated upwards:

$$\begin{aligned} &\binom{n}{26} \cdot 26! \cdot 26^{n-26} \\ &< \frac{n^{26}}{26!} \cdot 26! \cdot 26^{n-26} \\ &= n^{26} \cdot 26^{n-26} \\ &= n^{26} \cdot \frac{26^n}{26^{26}} \\ &= \left(\frac{n}{26}\right)^{26} \cdot 26^n. \end{aligned}$$

According to Stanica [20] the following downwards estimate exists with $m > 1$ and $k \geq 2$:

$$\binom{mk}{k} \geq 1.0844 \cdot e^{-\frac{1}{8k}} \cdot k^{-\frac{1}{2}} \cdot \frac{m^{m(k-1)+1}}{(m-1)^{(m-1)(k-1)}}.$$

4. Summary and outlook

For the first time, this book investigates substitution ciphers utilizing set-valued order relations as a mathematical argument. Linguistic properties are statistically newly structured and described for the example of the German and English languages. This approach requires individual optimal sets. For the example of the classical encryption process this strategy model can be extended to other languages. The scientific comparison of different sets yields evidence for linguistic properties and origin. The mathematical basis for a new prioritization with this process is the set less order relation.

In this book the currently established practice – (deductive) evaluations of the cryptanalysis according to statistical frequency – is extended to include mathematical set optimization. In the reduction path, the application of set-valued order relations extracts a differentiated individual result that depends on the particular language and serves as a design and analysis principle.

References

[1] F.L. Bauer, G. Goos, *Informatik – first part*, Springer, 1973.

[2] H. Beker, F.Piper, *Cipher Systems*, Northwood Books, 1982.

[3] A. Chiriaev, G.W. Walster, *Interval Arithmetic Specification*, technical report, 1998.

[4] A. Dhavare, R.M. Low, M. Stamp, Efficient Cryptanalysis of Homophonic Substitution Ciphers, *Cryptologia,* **37** (2013), no. 3, 250 - 281. https://doi.org/10.1080/01611194.2013.797041

[5] W. Fumy, H.P. Rieß, *Kryptographie – Entwurf und Analyse symmetrischer Kryptosysteme*, Oldenbourg, 1988.

[6] H.F. Gaines, *Cryptoanalysis - A Study of Ciphers and Their Solution*, Dover Publications, 1956.

[7] C.G. Günther, A Universal Algorithm for Homophonic Coding, Advances in Cryptology – Eurocrypt 88, *Lect. Notes in Comp. Sci.,* **330** (1988), 405 - 414. https://doi.org/10.1007/3-540-45961-8_37

[8] T.X.D. Ha, A Hausdorff-type Distance, a Directional Derivative of a Set-valued Map and Applications in Set Optimization, *Optimization*, **67** (2018), no. 7, 1031 – 1050, https://doi.org/10.1080/02331934.2017.1420186

[9] E. Karaman, M. Soyertem, İ. Atasever Güvenç, D. Tozkan, M. Küçük, Y. Küçük, Partial Order Relations on Family of Sets and Scalarizations for Set Optimization, *Positivity*, **22** (2018), no. 3, 783 - 802. https://doi.org/10.1007/s11117-017-0544-3

[10] S. Karmakar, A.K. Bhunia, A Comparative Study of Different Order Relations of Intervals, *Reliable Computing,* (2011), 38 - 72.

[11] D. Kuroiwa, T. Tanaka, T.X.D. Ha, On Cone Convexity of Set-Valued Maps. *Nonlinear Analysis,* **30** (1997), no. 3, 1487 - 1496. https://doi.org/10.1016/S0362-546X(97)00213-7

[12] N. Neukel, A New Approach for the Determination of Publicly Registered Land Values on the Basis of Interval-valued Duality Theory and Regression, *Appl. Math. Sciences,* **11** (2017), no. 16, 783 - 805. https://doi.org/10.12988/ams.2017.7128

[13] N. Neukel, *Dualitätskonzepte und Sozio-ökonomische Paradigmen mit Ordnungsrelationen zur Mengenoptimierung*, Dissertation, FAU Erlangen-Nuremberg, 2016.

[14] N. Neukel, Extension of a Dual Equivalence Class Model and its Application to the Socio-Economy of the German Real Estate Market, *Applied Mathematical Sciences,* **11** (2017), no. 47, 2325 - 2340. https://doi.org/10.12988/ams.2017.78254

[15] N. Neukel, *Compensation Payments for Aircraft Noise in an Urban Building Conflict Situation on the Basis of a Set-Valued Conjugate Duality*, Hikari Ltd. 2018.

[16] N. Neukel, Order Relations of Sets and its Application in Socio-Economics, *Appl. Math. Sciences*, **7** (2013), no. 115, 5711 - 5739. https://doi.org/10.12988/ams.2013.37419

[17] Z.G. Nishnianidze, Fixed Points of Monotonic Multiple-Valued Operators, *Bull. Georgian Acad. Sci.*, **114** (1984), no. 3, 489 - 491.

[18] G. Perec, *Anton Voyls Fortgang*, Zweitausendeins, 1986.

[19] C.E. Shannon, Communication Theory of Secrecy Systems, *Bell Sys. Tech. Journal,* **28** (1949), no. 4, 656-715.

[20] P. Stanica, Good Lower and Upper Bounds on Binomial Coefficients, *Journal of Inequalities in Pure and Appl. Math.*, **2** (2001), no. 3, 1 – 12.

[21] A. Sengupta, T.K. Pal, On Comparing Interval Numbers, *European Journal Operational Research,* **127** (2000), 28 - 43. http://dx.doi.org/10.1016/S0377-2217(99)00319-7

[22] SUN Microsystems. Inc., *Interval Arithmetic Programming Reference*, 2000.

[23] P.T. Vui, L.Q. Anh, R. Wangkeeree, Levitin–Polyak Well-posedness for Set Optimization Problems Involving Set Order Relations, *Positivity*, (2018), 1 - 18. https://doi.org/10.1007/s11117-018-0627-9

[24] R.C. Young, The Algebra of Many-Valued Quantities, *Math. Ann.,* **104** (1931), 260 - 290.