

FIAF Cataloguing Commission Occasional Paper No. 2

**ON THE PROBLEM OF NON-ALPHABETIC
CHARACTERS IN FILM TITLES**

(Transcription vs. ordering by computer)

by

**Oskar von Törne
Stiftung Deutsche Kinemathek**

Translated by

Roger Smither and Cooper Graham

October 1992

ON THE PROBLEM OF NON-ALPHABETIC CHARACTERS IN FILM TITLES **(Transcription vs. ordering by computer)**

The occurrence of "non-alphabetic characters"¹ in film titles becomes a problem in work with documentation or cataloguing when titles must be ordered in alphabetic sequence, since "alphabetic" has hitherto tended in practice to mean "in the correct sequence for the **Latin** alphabet". Where non-alphabetic characters are involved, there is no universally accepted place in current sorting procedures. This lack can be made good in two ways:

1. The non-alphabetic characters may be made "alphabetic" by transcription (traditional library procedure);
2. For non-alphabetic characters, a specific comprehensive ordering sequence may be established incorporating not only the letters of the Latin alphabet but an extended character set. In this way "non-alphabetic characters" will become in a higher sense "alphabetic" (computer sorting procedure).

Examples of results with traditional transcription procedure

A non-alphabetic character, or a character failing to find its place in (Latin) alphabetic order, may be correctly transcribed only if the **spoken** form of the character(s) concerned is unambiguously known **in the correct oral context**. This may immediately present two distinct problems:

1. We may not be able precisely to determine the spoken form of the original title through transcription because we do not have appropriate skills in unfamiliar (or sometimes not so unfamiliar) languages;
2. We may not be able precisely to determine the spoken form of the original title because we have no basis for an unambiguous decision between the various possible readings of the character(s) concerned.

¹ Author's note: this ad-hoc designation here means "numbers, symbols and additional signs other than individual letters from non-Latin alphabets" (Regeln für die Alphabetische Katalogisierung).

Problem of Non-Alphabetic Characters, p. 2.

Before we consider the computer-sorting solution to this problem, let us consider a few practical examples:

The non-alphabetic characters that occur most frequently in film titles are figures and numbers from various languages: 1812, 1905 god, 08/15, Zwischen 1/2 11 und 11, Die 999. Nacht, Cell 2455, Kärleck 1-1000. Already in these examples characters other than numbers have made an appearance. Other examples including non-alphabetic and non-numeric characters include 90°, Marguerite : drei, 1 + 1 = 3. There is also the problem that one cannot foresee which characters will appear next. When, for example, there is already a film with the title 40 qm Deutschland, one must assume that in the future the characters m^2 or m^3 will appear.

We are thus confronted with ever increasing transcription problems in many languages. Problems between languages extend over and above the problem of standard readings within the same language. Consider:

Sem' + dva = a. Sem' i dva
or b. Sem' plus dva.

Cell 2455 = a. Cell two thousand four hundred and fifty five
or b. Cell two thousand four hundred fifty five
or c. Cell twenty four fifty five
or d. Cell two four five five
or e. Cell two four double five.

Such uncertainty is not confined to foreign languages.² German also presents instances of uncertainty:

Marguerite : drei = a. Marguerite geteilt durch drei
or b. Marguerite durch drei.

Occasionally, even experts who should have precise knowledge are unable to give a definitive opinion. Thus Heidi Genee was unable to state precisely how the title of her film should be read to conform to the "original":

1 + 1 = 3 = a. Eins und eins ist drei/ ... macht drei/ ... gleich drei
b. Eins plus eins ist drei/ ... macht drei/ ... gleich drei.³

² Translator's note: this paper was originally written in German.

³ Translator's note: there would be equal room for confusion in an English title - 'One and one / One plus one is three / ... makes three / ... equals three'.

Problem of Non-Alphabetic Characters, p. 3.

In the daily work of archives some of these problems resolve themselves when the non-alphabetic characters are not required as significant elements in ordering. Furthermore, in cases of doubt cross-references help minimize the number of mistakes in finding material. While more and more cases arise in which a definitive placing of an item cannot be made because there is no clear-cut transcription of an element required for sorting, nonetheless the sum total of these cases is small enough for the problem to be insignificant from the point of view of practical archive work.

The problem of the correct spoken form of the film title is of greater significance when the perspective is that of film scholarship. In our publications, and especially in the indexes to our publications, pragmatic decisions on the basis of compromise are hardly to be entertained unless it is expressly acknowledged that the spoken form of the film title as it is represented by its placing in the alphabetical sequence of the index may not be a true representation of the original title. But even if use is made of this qualification, the unwelcome fact remains that because of the location in a certain place in the alphabetical sequence there will be a prejudice established in favour of a particular spoken form. That form is thus given a kind of spurious orthodoxy or authority by being written into an alphabetic order.

We can avoid this unwanted effect if we decide on the principle of a solution by computer sorting procedure. The computer does not translate non-alphabetic characters into Latin letters because there is no affordable program to accomplish this effect. It locates numbers and other non-alphabetic symbols according to an established table in its sorting protocols which recognizes a position, in addition to the letters of the Latin alphabet, of every other symbol. If required, a suitable cross reference in the traditional Latin alphabet, in a form similar to that required by traditional library practice, would be put at the relevant location for the title in question:

Eins und eins is drei: **SEE 1 + 1 = 3.**
but also Eins plus eins is drei: **SEE 1 + 1 = 3.**

In just the same way, we could usefully follow this procedure in routine archive work. Confronted by the original title Kärlek 1-1000, we need not wearily look for a transcription (which might in the end still contain errors): instead, we would simply accept that the title would sort before Kärlek hors without worrying whether a different reading of the printed number would result in an alphabetic location **after** Kärlek hors. In precisely the same way, Murder at 3 A.M. would come before Murder at Covent Garden.

Problem of Non-Alphabetic Characters. p. 4.

Certainly, the sorting of film titles according to computer procedures will sometimes lead, because of its uncompromising approach, to results that can be annoying from the perspective of practical archival work. Film titles with ordering elements which are identical when spoken but are different when written will be sorted differently. Examples:

1 x 1 der Ehe (West Germany, 1949) would not - as formerly - come before Einmal eins der Liebe (Germany, 1935), but before the entire Latin alphabet under "1".

Der Mann mit den 7 Gesichtern would come, not immediately after Der Mann mit den sieben Frauen, but before Der Mann mit den drei Frauen because the figure "7" is filed before the letter "d".⁴

In such cases, use must be made of cross references so that the unfamiliar user will not be misled.

We still do not have extensive experience of the use of computer sorting programs. Above all, we know little about the possibilities for writing programs specific to our needs. The usual sorting programs provide a place for every punctuation symbol outside the Latin alphabet, while normal library conventions disregard punctuation altogether. Without amendment, therefore, computer sorting would mean that titles which in their original form were written with quotation marks or begin with an ellipsis - for example "Denn die Elemente hassan ..." (Germany, 1913) or ... denn alle Schuld rächt sich auf Erden (Germany, 1913) - would not appear in their traditional alphabetic sequence but would be filed under the punctuation symbols used. The same would be true of Spanish titles phrased as questions or exclamations, which would be filed in the place determined by the computer sorting procedure for the inverted question or exclamation marks at the start of such titles. In these cases there would need to be a pragmatic compromise to the rule for practical archive work: since these characters have no effect on the spoken form of a title, we could simply leave them out of consideration and place the titles in their customary alphabetic location, even if they are placed in a completely different sequence in computer printout.

⁴Translator's note: for English examples, consider the case of the 1939 Hitchcock film of The 39 Steps which a computer would not file next to the 1978 version, which carried the title The Thirty Nine Steps; on the other hand, use of the computer convention prevents separation of titles which, while written similarly, are spoken differently - for example 1900 (Italy etc., 1976) and 1941 (USA, 1979). When currency is involved, we must prepare ourselves for titles grouped by currency symbols rather than by numbers - \$1,000 a Minute with \$10 raise not with 1001 Arabian Nights. Who is to say which change is more or less logical or helpful?

Problem of Non-Alphabetic Characters, p. 5.

These and other problems remain to be fully worked out. In any case, sorting according to computer procedure (which obviously also raises other problems such as the treatment of spaces and punctuation symbols occurring **within** a title) will require considerable re-thinking both for archive-workers and for their users.

In summary:

Advantages of sorting by computer procedure over traditional transcription-based practice -

1. It avoids the labour-intensive search for (correct) transcriptions;
2. Material is not kept in limbo in the cataloguing process because we cannot find an appropriate transcription - especially for less well-known languages;
3. We avoid the risk of giving a form of endorsement to incorrect "original titles" by committing ourselves in writing to transcriptions that may be imperfect;
4. The change to sorting by computer procedure based on current equipment standards would be an important preparatory step towards the implementation of computerization.

Conclusion:

From the point of view of archives of textual materials, I would declare myself in favour of the principle of sorting by computer procedure.