

3. What is Transcription, Really?

[ALOIS PICHLER]

The objective of the Wittgenstein Archives at the University of Bergen is to produce a machine-readable version of Wittgenstein's manuscripts. I would like to elucidate the particular concept of transcription which has developed through my work at the Wittgenstein Archives. It is one which I believe will be relevant to the activity of transcribing in general.

Quite different activities and types of interpretation are involved in what we call transcribing. We would all agree that transcribing a manuscript is not the same as copying it. Encoding is essential to transcription. But to say this is not enough. The encoding of any text involves a number of different activities which need to be distinguished. This is all the more evident with manuscript, which contains deletions, alternatives, instructions, etc.

As an example of transcription guidelines I shall present and discuss the encoding guidelines proposed in the Registration Standard of the Wittgenstein Archives, MECS-WIT 1.03. I'll point out that the set of basic rules according to which the encoding of the Wittgenstein manuscripts proceeds, does not have a hierarchic but rather a "democratic" structure. Thus the nature of transcription is in some sense always dynamic and requires decisions on the part of the transcriber. We must also be aware of the fact that there

are passages which simply cannot be transcribed faithfully.

The claim that transcribing is not the same as copying should not be controversial. How should we deal, for example, with the various handwritten tokens of the same grapheme? Copying a manuscript text can ultimately only mean the production of a facsimile. This is not what we mean when we say we transcribe a manuscript.

Transcribing a manuscript means to transfer the various dimensions and dynamics of the handwritten text — such as insertions above line, text written on top of another text, words deleted and replaced by other words — into a one-dimensional medium such as is used in a machine readable version (cf. Claus Huitfeldt: "Multidimensional Texts in a One-Dimensional Medium"). For this purpose we need an encoding system, i.e., we need codes which both retain the text information as faithfully as possible and provide the basis for further text processing, for example, for diplomatic and normalized printouts. In the following I shall discuss some of these encoding activities and propose a preliminary classification of the codes involved in the transcribing of modern manuscripts.

We have codes which can be classified as codes for DOCUMENTATION OF THE SOURCE. These codes contain information which concerns material matters such as the size of the original, writing medium, different hands as well as information about the history of the original, its origin and dates, and references to catalogues.

A second group of codes concerns DOCUMENTATION AND DISCUSSION OF THE CODE SYSTEM used in the transcription work. Since the code system is a reflection of work in progress which becomes updated time and again, the specification of the system used will contain important information for any further work with the transcriptions. It might also be that the transcriber has difficulties applying certain codes in a certain manuscript, or, that for various reasons he doesn't apply certain codes at all. All these matters need proper documentation and explanation; they might in themselves lead to further changes in the code system.

A third code group includes TRANSCRIBER'S COMMENTS AND EXPLANATIONS. In this would belong codes indicating uncertainty about the correctness of the encoding itself, codes for not clearly legible passages, codes for text which can't be deciphered at all, etc. Some text passages are impossible to transcribe

faithfully; here the transcriber will make a comment which refers the user back to the original.

Some text phenomena need an explanation which the transcriber — since he has worked intimately with the text — might be able to give, and which the user might appreciate. In the case of the Wittgenstein Archives this does not imply philosophical commentaries on the text, but rather information about such things as particular orthographic habits or the author's use of markers.

The Wittgenstein Archives' code system provides possibilities for distinguishing different functions of the same grapheme. Therefore the transcriber is required to distinguish whatever can be distinguished within practical limits. Parentheses, for example, can have quite different functions (beside the conventional use those of suggesting a deletion, indicating a possible substitution, etc.). Hence parentheses with different functions should be encoded differently.

The freedom of the transcriber to "add" clarifying information to the original has nothing to do with overinterpretation, even more so, since the transcriber's interferences can easily be separated from the authentic text so long as they are suitably encoded.

A fourth group of codes concerns the REGISTRATION OF INTERTEXTUALITY. What I mean by this is codes which register internal and external references, names of persons, references to published works, relations to other manuscripts, etc., as made by either the author or the transcriber.

A fifth group can be headed COMPOSITIONAL REGISTRATION: this dimension of encoding implies distinguishing different types of text within the manuscript, e.g., where something functions as a preface to label it as such, and similar for the author's own miscellany, editorial instructions, titles, content tables, etc.

Still other codes serve the RETRIEVAL AND ANALYSIS of various kinds, such as indexation.

Then there are codes which serve NORMALIZATION. If we aim at diplomatic and normalized printouts made from the very same transcription, the transcription file has to provide the basis for both; this means in the case of orthographical errors that both the authentic and the normalized versions will be accessible.

A fundamental aspect of transcription is that of PALEOGRAPHIC TRANSCRIPTION. We transcribe "r"s as "r"s, sections as sections, deleted text as deleted, inserted text as inserted, etc., despite the fact that the individual handwritten "r"s look different,

that the lines which strike out text can vary significantly, etc. A totally faithful paleographic transcription is not possible, neither is it desirable.

When we speak of paleographic transcription we must remember a point made by my colleague Ole Letnes: that our transcription work is determined by the topography of printed texts, e.g., the spacing of blanks and punctuation is fixed in certain scales. Another point to be made in this context concerns the role of perception. We must note that a whole word has to be grasped as a Gestalt and not as a mere compilation of single letters. This is probably true even for typed text (at least when reading at speed), but is absolutely undeniable in the case of handwriting. Very often we cannot see what the individual letters are until the whole word has been grasped. But this kind of interpretation in the reading of words and single letters is quite different from the interpretation involved in the encoding of a title as a title.

The last group I want to distinguish concerns what I would call SYNTAGMATIC TRANSCRIPTION. Syntagmatic transcription has the particular aim of producing a "unilinear sequence of orthographically acceptable and grammatically well-formed sentences which together form a coherent text unit." (MECS-WIT 1.03, 1.7) This is the requirement of well-formedness.

In order to comply with the requirement of well-formedness we must often deviate from a certain transcription rule, namely the rule of basic linear order which says: "Manuscripts should be transcribed page by page from the front towards the back page, each page vertically line by line from top to bottom, each line horizontally letter by letter from left to right." (MECS-WIT 1.03, 1.8). In other words, we often have to rearrange the text. In the case of an insertion outside regular lines which adds text to the text in line, for example, we will have to include the inserted text in the line.

There is a fundamental difference between paleographic and syntagmatic transcription. A paleographic transcription of a text which consists of a deleted and an inserted word would mean the encoding of the deleted text as deleted and the inserted text as inserted. But to encode the same text syntagmatically can lead to something quite different: if we understand the inserted word as replacing the deleted, we should — according to the requirement of well-formedness — embed the elements in a substitution code which then suggests a certain relation between them, namely the

relation of substitution. However, it often remains up to the transcriber to decide what a well formed text would be, for example, whether a certain insertion is an addition to the text or rather a substitution.

This classification of interpretation types or of scholarly activities, as involved in transcription work, is exhaustive as far as transcription work at the Wittgenstein Archives is concerned, but it is by no means exhaustive of what concerns text encoding in general. The classification doesn't, for example, include codes for grammatical encoding or for subject indexing. Neither are the codes exhaustive in their particular area; it would, for example, be easy to distinguish further between different kinds of deletions. Neither are the groups of this classification mutually exclusive in a strict sense.

The Encoding Guidelines of MECS-WIT 1.03 give preference to transcriptions which satisfy the criterion of well-formedness, i.e.: "If among two transcriptions basic interpretational considerations do not decide clearly in favor of the one rather than the other, we will decide in favor of the one which comes closest to the ideal of a well-formed text ... Or in other words: Among acceptable transcriptions we will prefer the more well-formed." (MECS-WIT 1.03, 1.7) MECS-WIT 1.8 prescribes a procedure which helps the transcriber to satisfy this criterion. The procedure consists in (A) obeying the rule of basic linear order, unless obeying this rule (D1) does not satisfy the requirement of well-formedness, or (D2) deviates from the chronological and/or compositional order of the text, or, (D3) does not separate the different readings of a certain word, in case the manuscript strongly suggests more than one reading of this word.

On the basis of my transcription experience, I shall discuss first the above given formulation of the requirement of well-formedness; then I shall suggest an alternative model to the hierarchic model of the rule set above (A, D1-D3).

The formulation of the requirement of well-formedness makes a wrong assumption. The assumption is that it is possible to separate the requirement of well-formedness from the "basic interpretational considerations," or from the acceptance of a transcription. This is not the case. When we accept a transcription, then the requirement of well-formedness either contributes to acceptability, or it does not. Neither in the former nor in the latter case would considerations for the requirement of well-formedness arise after the transcription has been accepted.

Contrary to the hierarchical model of the rule set,

as suggested in MECS-WIT 1.08, I would like to propose that the guidelines be seen as a set of equally valid rules. The requirement of well-formedness should be regarded as a rule among these equally valid rules. Of these rules none would make either a sufficient or a necessary condition for what could be considered a faithful transcription. There are times when neither a rule of well formed text nor a rule of basic linear order nor a rule of chronological order (to transcribe the text in the order of its chronological genesis) nor a rule of economy (to transcribe the text without any omissions or additions on either the paleographic or structural levels), etc., can be applied without serious reservations. This means that in applying one rule we are likely to offend against another. When transcribing, no encoding scheme can entirely anticipate every textual idiosyncrasy, and the decision between different possible transcriptions is often left to the transcriber. This decision will depend, on the one hand, on his evaluation of the evidence in the manuscript, and on the other, how he relates this evaluation to the editorial objectives.

References

- Huitfeldt, Claus, "Multidimensional Texts in a One-Dimensional Medium," 1992 (forthcoming).
———, "MECS-WIT - A Registration Standard for the Wittgenstein Archives at the University of Bergen," *Working Papers from the Wittgenstein Archives at the University of Bergen*, 1992 (forthcoming).