

Alois Pichler

Transcriptions, Texts and Interpretation

This paper does not deal with Wittgenstein's philosophy, nor does it speak about his Nachlass. Rather, it discusses one aspect of making his Nachlass accessible to machine processing: computer aided transcription, as it is done at the Wittgenstein Archives at the University of Bergen (WAB). This discussion involves questions about text representation and text in general.

The terms "transcribe" and "transcription" are often understood as the process of making - by writing - an exact representation of a document, text, etc. Consequently it may be said that the aim of transcription is to represent the original as correctly as possible. However, I would argue that transcription - in particular transcription as it is done at WAB - is a much more complex process than "to represent the original manuscripts as correctly as possible".

The following discussion will therefore firstly focus on the questions "what is transcription?" and "what is the aim of transcription?". Secondly I will consider in more detail the question of transcription and interpretation. We must recognize that transcription work involves a range of different interpretational activities, which need to be distinguished in order to understand transcription work properly and avoid problems and inconsistencies in this work.

Both parts of my paper serve to illustrate the following general point: Transcribing is not copying, but - as text-editorial work in general - rather selecting and interpreting. Any edition of Wittgenstein is in a strong sense a result of interpretation. Our only option is to formalize interpretation, and to make it explicit.

I

The aim of transcription has often been defined as "to represent the original manuscripts as correctly as possible". This needs a clarification, therefore let us ask some questions:

What does "as correctly as possible" mean? What is the criterion of *correctness*? And what does "to represent" mean? To represent in which medium?

I think, the essential question is not about a true representation, but: Whom do we want to serve with our transcriptions? Philosophers? Grammarians? Or graphologists? What is "correct" will depend on the answer to this question. And what we are actually going to represent, and how, is determined by our research interests (philosophical, grammatical, philological, graphological ... in-

terests), and not by a text, which exists independently and which we are going to depict.

In our transcription work at WAB we do not for example distinguish between features of handwriting such as convex and concave "r"s, or the position of the dot over the "i". We do not record the endings of a line, and we represent only a few grammatical elements of the text. All this would be possible, but you might say: It is of no relevance to distinguish such features. However, the fact is, we have decided, what is relevant for us and what not. We have made certain decisions about what a transcription should contain, and these are in answer to our idea of whose interests we serve.

Even so, our decisions are not final, or at least, ought not to be. We could, for example, insert all line endings; our reason for not doing this at present is not that it might not be of importance to someone, but rather, that it does not belong to our current interests. This again also has to do with practical questions such as financing, time schedules etc. As you can imagine, making a record of all the line endings is an extremely time consuming task.

The sign of a good transcription system is that it has extensibility. This means that it should be possible to revise and adjust the system to serve new interests as and when they arise. Thus we choose to transcribe certain things at an early stage knowing that we can include other things later if necessary. What we choose to include initially is also determined by what is easier to distinguish while emerged in the text. The insertion of codes for wordclasses, for example, would be a simple task more easily performed afterwards, whereas the distinction between different uses of parentheses is best done during the initial transcription process.

With these considerations in mind I conclude: Our aim in transcription is not to represent as correctly as possible the originals, but rather to prepare from the original text another text so as to serve as accurately as possible certain interests in the text. We do not want to produce a photograph of the original - this is the function of a facsimile. "As correctly as possible" can only mean: "as correctly as possible in relation (in answer) to certain research interests".

What these interests are, with regard to the work done at WAB, I will present in the second part of this paper.

II.

To transcribe a text according to specific interests will require that those interests be clearly served by different conventions, e.g. different codes. The ap-

plication of these conventions presupposes in turn a variety of different interpretational decisions. For example, to identify the string *DIE ABHANDLUNG* as an inscription which bears a meaning requires a different interpretational standpoint and knowledge than is required when encoding it as an intertextual reference to Wittgenstein's "Abhandlung", the Tractatus, or as a variant to *MEIN BUCH*.

What are the different interests - on a macro level - we at WAB focus on when transcribing Wittgenstein's manuscripts?

(a) A fundamental aspect of transcription is that of graphic transcription.

This means that we transcribe "a"s as "a"s, "b"s as "b"s, "c"s as "c"s etc., sections as sections, deleted text as deleted, inserted text as inserted, etc. But already here the selective element becomes obvious, because we do not record that the single handwritten letters look different, that the lines which strike out text can vary significantly etc. A totally faithful graphic transcription is not possible, neither is it desirable.¹

Speaking of graphic transcription we must make a point which concerns the role of perception. When transcribing, one first tries to grasp the written word as a Gestalt, and not as a sequence of single letters. Very often - in particular when transcribing Wittgenstein's secret code passages - we cannot see what the single 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.

(b) The next transcription activity I want to distinguish is syntactic transcription.

Syntactic transcription has the particular aim of providing for text processing which produces a syntactically well-formed text. In order to meet this requirement of well-formedness 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. However, doing so, we do not forget about the principle of graphic transcription and, in this case, encode the text in addition as inserted outside regular line (e.g. in the upper margin of the page).

There is a fundamental difference between graphic and syntactic transcription. A graphic transcription of

I-AM ^{FEEL} FINE

¹ I should make it clear that a facsimile does not fall under the notion of "transcription", as it is understood here.

where *AM* is deleted and *FEEL* inserted above would mean the encoding of the deleted text as deleted and the inserted text as inserted, and might look as follows:

I <del/AM> <ins/FEEL> FINE

But to encode the same text syntactically leads to something quite different: Since we probably all understand the inserted word *FEEL* as replacing the deleted word *AM*, 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 (which might be represented as follows):

I [subst|AM| FEEL] FINE

What WAB does is both, graphic and syntactic transcription:

I [subst|<del/AM>| <ins/FEEL>] FINE

Another example of syntactic transcription is the following:

I <del/<npc/IF>> LIKE COWS

The deleted word *IF* does not form a part of the syntactic context, therefore it is encoded by "<npc/.../npc>" (= not part of context), a syntactic code. This is in contrast to cases such as

I LIKE ~~BIG~~ COWS

which can be transcribed - purely graphically - as

I LIKE <del/BIG> COWS

The inclusion of *BIG* does not conflict with the requirement of well-formedness, since *BIG* can form a part of the syntactic context and has no replacing function.

However, it is often left to the transcriber to decide whether a certain inter-linear string is an addition to the text or rather a substitution.

An analysis of WAB's transcription shows that codes which fall under the "syntactic transcription"-category constitute a large part of the codes used in the transcriptions, which in return means that this type of transcriber's interpretation is highly present in the transcriptions. WAB's transcriptions are therefore much more than an ad litteram transcription: they *do* provide for accurate diplomatic printouts, but they also allow for the possibility of printouts of well-formed texts.

(c) A third type of encoding is normalization of orthography.

If we aim at being able to produce both what we call diplomatic and normalized printouts 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.

(d) A fourth activity includes the application of codes for documentation of the source.

These codes contain information concerning material matters such as 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. This presupposes knowledge of the source.

(e) A fifth group concerns documentation of the use of the code system in the transcription work and transcribers' explanations and comments on the text, the transcription process and the use of the code system.

Since the code system is a reflection of work in progress which is updated time and again, the specification of the system used will contain important information for any further work with the transcriptions. It might e.g. be that the transcriber has difficulty applying certain codes (following up certain interests) in a certain manuscript. These matters need proper documentation and explanation; they might in themselves lead to further changes in the code system.

This group would also contain codes indicating uncertainty regarding the correctness of the encoding itself, codes for not clearly legible passages, codes for text which cannot be deciphered at all etc. Some text passages might be very difficult or impossible to transcribe: here the transcriber will make a comment which refers the user back to the original.

Sometimes 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 WAB 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.

(f) Another type of interpretation is again involved in what we call disambiguation: WAB's code system provides possibilities for distinguishing between different functions of the same graph. Therefore the transcriber is required to distinguish whatever should and can be distinguished within practical limits. Parentheses in Wittgenstein's Nachlass, for example, can have quite different functions (besides the conventional use: e.g. suggesting a deletion, indicating a possible substitution etc.); hence parentheses with different functions should be disambiguated and encoded differently.

(g) A final type of encoding serves the retrieval and analysis of various

kinds, such as indexation. These codes concern among other things the registration of compositional features and intertextuality.

What I mean by intertextuality is codes which record 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. Such codes allow for hyperlinks which guide you to variants in other manuscripts etc.

Compositional registration implies distinguishing different types of text within the manuscript, e.g. where something functions as a preface to label it as such, and similarly for the author's own miscellany, editorial instructions, titles, content tables etc.

The classification of interpretation types as presented here is by no means exhaustive in relation to text encoding in general. The classification does not for example include a set of codes for grammatical encoding or for subject indexing. Neither are the codes of WAB's transcription standard exhaustive in their particular areas: it would for example be easy to distinguish further graphically between different kinds of deletions.

In order to provide for consistent and smooth transcription work, it is wise to keep these categories as much apart from each other as possible, which means that the single types must be extractable without hereby interfering with other types. From this it follows e.g. that it is necessary to encode variants which at the same time are insertions, both as variants and as insertions - since the graphic level shall be kept apart from the syntactic level.

The conclusion from these considerations about transcription work at WAB is that transcription work is essentially selective and interpretational in nature, moreover, that any text editing work is interpretational work. Editing Wittgenstein's Nachlass in book form presents the same types of interpretational problems as are encountered in preparing an electronic edition. The difference between a book edition and an electronic edition lies, however, in an electronic edition's potential to be able to make the types of interpretation - and their differences - explicit and extractable, to give the user the possibility to choose between the different levels of interpretation, and to realize them in different ways: e.g.: to choose an ad litteram printout rather than a normalized one, but at the same time have text, which was originally underlined, printed in italics. With regard to these demands a machine-readable version has considerable advantages over a book edition.

Alois Pichler
The Wittgenstein Archives
University of Bergen

Harald Hårfagresgt. 31
N-5007 Bergen
Norway