

Special Workshop

WITTGENSTEIN'S *NACHLAß*:

THE WITTGENSTEIN ARCHIVES AT THE
UNIVERSITY OF BERGEN

DAS WITTGENSTEIN-ARCHIV DER
UNIVERSITÄT BERGEN

Michael A.R. Biggs

Graphical Problems in Wittgenstein's *Nachlaß*

Abstract

This paper describes some features of Wittgenstein's *Nachlaß* which lie on the boundary between text and graphics. These deliberately ambiguous features are introduced as part of Wittgenstein's method of using "text-experiments" comparable to "thought-experiments". These features present problems of interpretation, and therefore of representation, in print or as encoded text for computer-aided analysis because their ambiguity must be maintained. The avoidance of the problem by the use of names is rejected with reference to Wittgenstein's discussion of universals in BLB. The resultant system of encoding graphics for computer-aided analysis at the Wittgenstein Archives is briefly described. Finally, in any representation of the *Nachlaß*, the need to maintain the integration of graphics and text is emphasised.

Graphical Problems in Wittgenstein's *Nachlaß*

The primary objective of the transcriptions being undertaken at the Wittgenstein Archives at the University of Bergen is to present material which is capable of textual analysis by computer. However, Wittgenstein's philosophical works are well known for containing graphics. In my catalogue of the graphics in the published works (Biggs and Pichler 1993) I identified 479 instances, and work in progress compiling a database of graphics in the manuscripts in the *Nachlaß* (MSS 101-182 as defined in von Wright 1982) contains approximately 2500 instances.

In some cases these graphics may be included or excluded without affecting the ability to analyse the text. But there are some graphics which act as word-substitutes. Therefore even if exclusively textual analysis is desired, some encoding is necessary in such places to indicate the presence of a graphic. In addition to the unequivocal use of graphics there are many instances of features which lie at the very margin of the distinction between graphics and text. These features are a symptom of Wittgenstein's philosophical method and this paper focuses on them and considers the implications for the mark-up of the body text.

The general impression one gains from the *Nachlaß* is that the scripts are working documents, full of corrections, amendments and compositional instructions. Some of these annotations made by Wittgenstein on his own writings are graphical in character. For example, there are many occurrences of arrows indicating the repositioning of text on a single page. He also uses a series of marks and ciphers to stand for editorial instructions such as moving sections to locations in other scripts, indentation of new sections, etc. All this commentary, both graphical and textual, on the base text may be referred to as the *meta-text*. The meta-text contains interesting information on Wittgenstein's compositional process, but consists of signs which are not to be reproduced in the reorganised or edited text itself. We can thus distinguish between the *diplomatic* version of the text, and the *normalised* version; the former reproducing the meta-text either typographically or in facsimile, and the latter in which the instructions contained in the meta-text are implemented.

The term "graphic" is itself somewhat vague, and may be used to describe a wide variety of phenomena. Wittgenstein's method, particularly in the later writings, is to question our philosophical assumptions about the way that language is operating, and one means by which this is achieved is to present text-experiments, rather like thought-experiments. In these experiments a new sign is introduced. It is often a simple pictorial graphic with more than one reading:

[1]

It is possible, if undesirable, to omit this pictorial graphic in a normalised version of the text because it serves as an *illustration* to the text. In other words, the text makes sense with or without it but the illustration demonstrates or reinforces what is said in the text. Graphics which are *syntactically integrated* in the text cannot be omitted, though there are examples where its exact orthography is not important, e.g.:

[2]

In an earlier folio of the same manuscript, in addition to noting the presence of an arrow, representing its orientation is also necessary to preserve the sense, e.g.:

[3]

Here, the arrow is a conventionalised pointing sign. In a typographic representation of the manuscript, the exact orientation and appearance may be substituted by a "type" which is capable of indicating the sense of opposition in the second example, e.g. "this: [4], and not this: [5]". Felicitous transcription in this case could be regarded as maintaining sense without necessarily maintaining appearance.

Substituting the name of a graphic for its graphical representation may be possible in some instances, though normally the inability to differentiate types from tokens results in an inability to name a graphical form. General names do not give information other than marking the presence of a graphic, and when substituted may not integrate satisfactorily with the body text, e.g. compare:

Wenn wir jemandem ein gewisses Zeichen etwa zeigen, so fällt ihm eine Ziffer ein... (RFM 1974 VI §44)

and;

When we show someone a certain sign e.g. [6], a numeral occurs to him... (RFM 1978 VI §44)

as normalised versions of;

[7]

The use of general names is misleading if subsequent textual analysis interprets all graphical representations as general representations.

there is a tendency rooted in our usual forms of expression, to think that the man who has learnt to understand a general term, say, the term "leaf", has thereby come to possess a kind of general picture of a leaf... one which only contains what is common to all leaves. (BLB p.17f.)

The use of general pictorial graphics or genre-pictures is an established practice of non-denotative visual representation (Goodman 1976 §1.5). However, here Wittgenstein contrasts it with our temptation to assume the necessity of denotation by linguistic universals, e.g. Goethe's *Urpflanze*.

Even if names are sought for each graphic, Wittgenstein introduces specific text-experiments based on ambiguity of reading and naming. Such a graphic is the ambiguous character [8] in MS 132 p.183 (RPP-I §539, etc.). Its function is similar to the binary switching of the duck-rabbit illustration. As an ambiguous pictorial graphic, the duck-rabbit is compared with understanding the meaning of a word (PI p.214). Such examples illustrate the role of a sign within a community of users by presenting our community with a kind of *double-entendre*. The use of a visual metaphor with two clearly opposed interpretations, coupled with the thought-experiment of "aspect-blindness", questions our assumptions about the connection between signs and their use.

Character ambiguity is also found in the representation [10], which has a typographical form identical to both a mirrored Latin character R and the Cyrillic upper case character Ya. However, the mirrored R and Ya must be differenti-

ated for analytical purposes. Differentiation in this case is achieved not by detailed graphical analysis but by the textual context, e.g. Russian text in MS 166 pp.57-65.

Wittgenstein uses whole words in "mirrored writing" in PI and LW (p.198 and §599 respectively), to comment on our changed perception of script when reversed. The corresponding manuscript passage for PI (MS 144 p.46) does not include a sample, only a meta-textual instruction to include mirrored writing. In the manuscript for LW (MS 137 p.135a) there is inverted writing with a meta-textual instruction to represent it as mirrored writing. In these cases it is difficult to know what should be represented in an edited or normalised representation since the text of PI is preceded by the experiment to "hold the drawing of a face upside down...". Four figures then follow in the comparative pairings: (a) is to (b) as (c) is to (d). However, in the published edition of PI, (a) is a 180° rotation of (b) while (c) is a horizontal reflection of (d). This provides an example of the inappropriate graphical representation of the pair (a),(b) or (c),(d).

[11] [12] [13] [14]

Ambiguous characters may also be used as non-conventional signs. For example, Wittgenstein rotates a Latin character A (MS 152 p.29) through 90° increments. It appears at one point to be similar to the logical notation ". However, this sign may alternatively be the inverted Latin character A, or a sign used habitually in the meta-text to indicate text to be moved to another location in the script, e.g.:

[15]

Whether a form is a graphical sign or other character representation is not always clear. It may also be desirable to substitute standard characters, e.g. ", for graphical representations to facilitate on-screen or printed representation. However, for computer-aided analysis an encoded character representation must indicate the difference between logical "for all" and "inverted or rotated Latin A" or any other sign, even if the printer instruction is identical. This may be achieved by a combination of graphical and *character disambiguation* codes.

Character disambiguation is also necessary to distinguish between a mirrored upper case Latin E (e.g. [16] in MS 132 p.30) and the logical notation \$; a rotated upper case Latin M (e.g. [17] in MS 132 p.40) and the upper case Greek S. Additionally some novel composites are proposed, for example the combination of the upper case Latin I and the numeral 3 to create an upper case Latin B (e.g. [18] in MS 132 p.97). Each encoding might differentiate between a graphic and a standard character, or cue the graphical interpretation of a particular orthographic token.

Thus far, the distinction between graphics and text has been made on the basis of whether a feature belongs to a conventional sign set. The ambiguous character formed a marginal case and extremes were represented by the Latin alphabet as non-ambiguous orthographic types, and overtly pictorial graphics such as the duck-rabbit. However, the sign set of the Latin alphabet must also be arranged in a left-to-right linear distribution. Where the distribution of the signs on the page departs from this convention it may again be appropriate to assign a "graphics" code. For example, a table, whether ruled or unruled, may be regarded as a graphic.

The left-right reading convention itself forms part of a text-experiment. For example, our practice of writing p to the fifth place as 3.14159 is contrasted with an alternative spatial technique in Z §699 (MS 131 p.144, also published in RPP-I §330):

[19]

This "alter[s] its appearance to the point of unrecognizability by us". The example makes the comparison with a practice which we do recognize, that of translating a lyric poem into another language, or solving the problem

How is this joke (e.g.) to be translated (i.e. replaced) by a joke in the other language? (Z §698)

Felicity is here presented as the appropriate replacement of one joke by another, which is equivalent to the replacement of a textual practice with a graphical one.

We have a graphical calculus to demonstrate the comprehensiveness of lists. This graphical practice may be clearly seen in the standard layout of the first two columns of a truth-table:

[20]

The technique of systematically varying the permutation of the elements enables the author to demonstrate comprehensiveness. It is also seen in other tabular arrangements, e.g. PG p.348:

A B C
A C B
B A C
B C A
C A B
C B A

The graphical layout shows that no line is repeated and proves that there are therefore six permutations of the set of letters, i.e. that there is "a *single* pattern"

(RFM-I §28). However, alternative graphical practices may prove, for example, that $2+2+2=4$:

[21]

Indeed, Wittgenstein draws our attention to the implicit graphical proof in many of our mathematical concepts including counting. (§§28-36).

In addition to the problems of interpreting and therefore of representing graphics on-screen or on paper, there is the problem of providing users with a compatible system of search and retrieval. Although some file formats can contain text based information in the file, e.g. TIFF (Tagged Image File Format), most graphical images are unanalysable by methods which are comparable to the analysis of the text string. The Wittgenstein Archives use codes specific to the representation of graphics which allows them to be located within the text string. As indicated above this sometimes labels a graphical layout of otherwise normal typographical characters, e.g. tables. To differentiate basic graphical forms an embedded keyword system has been devised which groups graphics according to their appearance. This grouping allows indexing and retrieval of graphics independently from viewing and analysing the images themselves.

The keyword system currently divides the full range of graphics into 17 primary classifications, e.g. *Schema*, which comprises all forms of tabular layout. Secondary keywords differentiate between variants on this form, e.g. ruled Schema such as truth tables, etc. A tertiary keyword is available to differentiate the number of similar elements in the whole graphic, e.g. two such Schema. By using up to three keywords, graphics of a similar physical appearance may be easily searched and retrieved. This indexing currently identifies 106 basic appearance types.

Graphics having identical primary, secondary and tertiary keywords will not necessarily be tokens of the same type, e.g. a truth table and a set of dots in the same pattern. However, it will normally enable comparisons to be made between structurally similar graphics and their textual context. Combined use of graphical and non-graphical codes in searches is also possible. For example, ambiguous characters such as [22] and [23] have the same graphics keywords but the latter also has a Russian character code when it could be read as the Cyrillic Ya.

Conclusion

It is possible to omit some, but not all, of the 2500 graphics in a normalised representation of MSS 101-182. For computer-aided analysis codes need to be

inserted to show the presence of a graphic and, often, information about the general nature of the graphic. Conventional signs may be named, e.g. arrows, but may also need secondary codes to indicate orientation in order to preserve the sense of the textual context.

Although the majority of graphics are clearly pictorial, there are also examples which lie close to, or cross, any boundary which might be drawn in an attempt to divide the two. Graphics may also encompass the non-conventional page layout of standard text. Encoding must not proscribe a particular interpretation as conventional or unconventional notation as such signs are often introduced for their ambiguity. For the purposes of analysis it is desirable to be able to identify all graphics whether pictorial in character or not. It is also desirable to represent as much as possible in text strings, on screen or on paper without resorting to graphics files. The Wittgenstein Archives has therefore decided to use a graphics code which may be attached to a graphics file, or to a feature which, whilst representable by typographical means, nonetheless has graphical characteristics. The use of further keywords enables combined text and graphics searches to be undertaken which maintains the complete integration of the two elements, apparent in the *Nachlaß* itself.

References

- Biggs, M.A.R. and Pichler, A. (1993) *Wittgenstein: Two Source Catalogues and a Bibliography*, Working Papers from the Wittgenstein Archives at the University of Bergen No.7, Bergen, Norway: University of Bergen.
- Goodman, N. (1976) *Languages of Art*, Second edition, Indianapolis, USA: Hackett Publishing Co.
- Von Wright, G.H. (1982) *Wittgenstein*, Oxford: Basil Blackwell, 35-62. First published as: Von Wright, G.H. (1969) 'The Wittgenstein Papers' *Philosophical Review*, **78**, 483-503.

Michael Biggs
Moor Lane Cottages
Moor Lane
Rickmansworth
WD3 1LR Herts
England
E-mail: m.a.biggs@herts.ac.uk

Graphics corresponding to the numbered entries in square brackets in the text,
e.g. [1]

[1]



MS 144 p.38

[2]

in dieser → Richtung

MS 134 f.59

[3]

lies: → , + nicht dies: ←

MS 134 f.57

[4]



[5]



[6]



[7]

Vorgang: Wenn wir
jemandem ein fest.
bes. Zeichen zeigen
gegen ~~so~~ so fällt
ihm eine Ziffer ein;

MS 164 p.133

[8]

F

[9]

not used

[10]

Я

[11]



(a)

[12]



(b)

[13] *pleasure*
(c)

[14] *Pleasure*
(d)

[15]

MS 130
p.55

[16] *H*

[17] *M*

[18] *B als B*

[19]

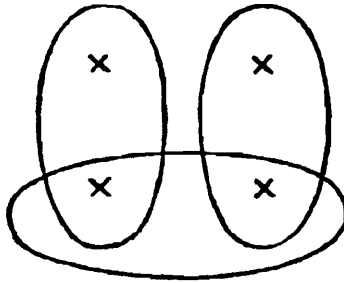
Z §699

Graphical Problems in Wittgenstein's Nachlaß

[20]

T	T	
T	F	
F	T	
F	F	

[21]



RFM-I §38

[22] \mathcal{F}

[23] $\mathcal{Я}$

Wittgenstein's *Nachlaß* Revisited

Introductory Note

Less than half of Wittgenstein's writings have been published to date. The existing publications are scattered among a multitude of different sources in monographs, anthologies and journal articles. By far the majority of them have appeared posthumously, and they follow no uniform set of editorial principles.

The aim of the Wittgenstein Archives at the University of Bergen is to make Ludwig Wittgenstein's entire *Nachlaß*¹ available for research by preparing it for publication in computerized form. The publication of this electronic version has already begun, and will be completed within a few years. It will contain a complete catalogue as well as transcriptions and a facsimile of the entire *Nachlaß*.

1. The Wittgenstein Archives at the University of Bergen

The Wittgenstein Archives was founded in 1990 and is organized as a research project at the Department of philosophy at the University of Bergen. Our current staff comprises 6 full-time and one part-time staff members. We are funded by the University of Bergen and other Norwegian sources.

We enjoy regular communication with the Wittgenstein Trustees², who have entrusted the University of Bergen with the exclusive right to publish the entire Wittgenstein *Nachlaß* in machine-readable form. We also cooperate with Oxford University Press, the Brenner Archives at the University of Innsbruck and the Oxford Text Archive at Oxford University.

The Wittgenstein Archives is not an archive in the traditional sense so much as a computer archive. We do not hold any of the original documents relating to Wittgenstein. However, we are one of the few places where general access is

¹ In this paper, the term *Nachlaß* refers to the philosophical writings catalogued by von Wright 1982. The catalogue does not include letters, unauthorized lecture notes, notes on conversations etc.

² Professor G.E.M. Anscombe (Cambridge), Sir Anthony Kenny (Oxford), Professor Peter Winch (Urbana-Champaign, Illinois), and Professor Georg Henrik von Wright (Helsinki).

provided to a complete set of xerox and microfilm copies of the entire *Nachlaß*.

In addition to our main tasks concerned with the preparation of Wittgenstein's *Nachlaß* for publication, we receive guest researchers, give seminars, publish a series of working papers and take part in international work on text encoding³.

2. Current Wittgenstein Editions

Wittgenstein is one of the best selling philosophers of our time. The new editions of his works still appearing more than 44 years after his death, the frequent reprints of old editions, as well as the burgeoning quantities of secondary literature confirm that he is still very much on the agenda of modern philosophy.

Wittgenstein's published works have not been released in a single collected edition. Except for TLP, WV and three minor publications, all editions have been published posthumously.

Therefore it is not surprising that there has been considerable debate over the relationship between these published sources and their basis, i.e. the 20,000 page '*Nachlaß*' that Wittgenstein left behind when he died. Philosophers have asked themselves whether the published editions are reliable, and whether - or to what extent - they give a complete picture of Wittgenstein's philosophy. In order to answer these questions the following three observations are relevant:

First, the published editions have been guided by no uniform and consistent set of editorial principles. Some of the editions are page by page reproductions documenting almost every aspect of the original manuscripts in minute detail. But by far the majority contain no detailed information about the state of their sources, even though they have sometimes involved quite considerable editorial interference.

Second, there is no one-to-one relationship between the posthumous editions and items in the *Nachlaß*. Many of the published editions are made on the basis of several disparate *Nachlaß* items, and some parts of the *Nachlaß* have been included in different contexts and/or in different editions. The lack of detail in the documentation of sources for many of the editions has made it difficult to form an impression of these relationships⁴.

³ For a full account of the background and history of the Wittgenstein Archives, see WAB 1990-93.

⁴ Such documentation is provided and discussed in numerous commentaries - see e.g. von Wright 1982 and 1992, Kenny 1984, Hilmy 1987.

Third, large parts of the *Nachlaß* are unpublished to date. According to Pichler, between two thirds and three quarters of the *Nachlaß* have not been published.⁵

Especially the last point has occasionally been used to create an impression that the published editions are calculated to serve a particular view of Wittgenstein's philosophy or that things not amenable to this view have been withheld from publication. However, most of Wittgenstein's *Nachlaß* has been available in scores of university libraries since the Cornell photocopy was made available in 1967. It is true that the Cornell copy itself is not complete, that parts of it have been covered up, and that it is of a technically low quality. But it is not infrequent for scholars to have been granted access to the original documents.⁶

It should also be stressed that Pichler's estimate, that only between a quarter and a third of the *Nachlaß* has been published, does not prevent Pichler himself (and others, including the present writer⁷) from maintaining that virtually all the philosophically important aspects of Wittgenstein's *Nachlaß* are present in the posthumous publications. The *Nachlaß* contains a large number of revisions and rewritings of basically identical texts. To publish all of this page by page would lead to a massive duplication of almost identical texts.

On the other hand, many of those who have worked with the *Nachlaß* have experienced that both the relationships between different versions, where they can be studied in their original context, and many of the smaller minute details such as deletions, insertions, substitutions, cross-references etc. shed valuable light on interpretational issues.

All considered, there seems to be a need to provide better access to the *Nachlaß*, e.g. in the form of a new, detailed and comprehensive complete edition. But in order to solve the problems outlined above, such an edition would have to satisfy a number of requirements, among which I would emphasise the following:

1. It must apply a uniform and consistent set of editorial principles throughout the entire *Nachlaß*. The edition must document manuscript details according to the highest possible standards of text-critical accuracy.
2. The edition must document the exact *Nachlaß* sources for each and every part.
3. It must provide references to already published editions of the

⁵ Biggs and Pichler p 12.

⁶ Cf Huitfeldt 1991

⁷ Pichler 1993 p 55, Huitfeldt 1994a p 280

same material wherever relevant.

4. It must cover the entire *Nachlaß*, i.e. it must somehow do justice to the various versions of basically identical texts, and account for the relationships between them.

5. It must be open-ended, i.e. allow for the inclusion of new material not yet discovered.

In combination, requirements 1 and 4 seem particularly difficult to cope with in a printed edition. Either one may publish the entire *Nachlaß* page by page and end up with an extremely repetitious and bulky publication, which would require rather extensive cross-referencing in order to help the reader orient himself. Or one may choose a set of base texts, and document parallel versions of closely similar material according to text-critical conventions. This would not only lead to an extremely complicated apparatus, but also necessitate quite extensive editorial intervention, thus replicating many of the problems with the existing editions.

I have argued elsewhere⁸ that it is unclear whether a book edition is feasible in the light of these requirements. Is there any reason to believe that a computerized version will do any better?

3. Aims and Scope of the Wittgenstein Archives

The main objectives of the Wittgenstein Archives are easily summarised: We will produce a complete transcription and a complete facsimile of the Wittgenstein *Nachlaß* and make them available electronically.

To date we have transcribed 8000 pages of the 20,000 pages constituting the *Nachlaß*. When the present paper is published (August 1995), publication of these transcriptions will already have begun.

Publication of the facsimile, which has been entrusted to Oxford University Press, will follow later. This will be distributed on CD-ROM discs.

These source materials will be supplemented with access tools which enables the user to use them in conjunction with each other, as a seamlessly integrated whole.

Users will normally access the material via a catalogue. From the catalogue, one will identify the item (or items) one wants to work with. This may then be displayed as a transcription on the screen, in which one can search for words or phrases, consult alphabetic and frequency word lists, concordances etc.

⁸ Huitfeldt 1994a p 281 f, Huitfeldt 1994b p 40

Once the facsimile becomes available a user with a CD-ROM player and a copy of the facsimile CD-ROM may click a button and see a facsimile reproduction of the relevant page whenever needed - e.g. when the transcription refers to some uniquely interesting feature.

According to current budgets and plans transcriptions of all 20,000 pages will be complete in 1999. This means that in three or four years time philosophers anywhere in the world will have access to the entire Wittgenstein *Nachlaß*.

4. Basic Principles of the Electronic Edition⁹

A basic, minimal requirement guiding the transcription work at the Wittgenstein Archives is that transcriptions should provide a suitable basis for the production both of (1) a detailed and strictly diplomatic (literal) version and (2) a normalized and simplified version of each and every manuscript. Transcriptions should also include all the information needed for certain kinds of computer-assisted processing and analysis, such as the production of (3) complete or selected graphword lists, segmentation according to chronological criteria, and cross-referencing. In addition, transcriptions should include information facilitating (4) lemmatization, grammatical analysis, and segmentation according to semantical criteria.

These objectives often imply conflicting demands. In particular, the first easily conflicts with the others. Yet for any given manuscript, we want one and only one transcription to serve all our purposes.

Therefore, we prepare the manuscripts in a primary format which can be manipulated by specially designed computer software. This enables us to produce secondary versions which satisfy the demands set forth above. In the primary format the relevant text features are marked explicitly in accordance with a rigorous formal syntax¹⁰.

Transcription is basically conceived of as copying the text of a manuscript page by page, line by line, and letter by letter. However, since manuscripts contain intralinear insertions, overwritings, substitutions etc., it is often necessary to deviate from this general rule. Deviations may take the form of rearrangement, simple or reiterative substitution, suspension, or extension of text elements. All such deviations are marked by special codes.

⁹ The points made in this section are discussed more fully in Huitfeldt 1994a, 1994b and 1995b

¹⁰ Cf Huitfeldt 1995a and 1995b

An all-important aim of manuscript transcription is to represent the original manuscripts as accurately as possible. But it is not entirely easy to spell out in exact detail what a correct representation in the form of a transcription is.¹¹

All aspects and parts of transcription presuppose interpretation. All interpretational decisions may in principle be subject to doubt or conflict, and some will in fact be subject to doubt or conflict even among the most competent of scholars. Even the most basic operations in transcription, reading the graphs of a manuscript and identifying them as instantiations of graphemes, involves informal and largely unrecognised skills. The transcriber has to draw on all sorts of background knowledge which we can have no hope to be able to specify in any exact way.

This is not to say that transcription is an entirely private, subjective, and arbitrary process. Sound interpretational decisions are informed by knowledge of what is likely to be seen as philosophically relevant by informed and competent readers, i.e., transcription presupposes philosophical scholarship.

Therefore, transcribers at the Wittgenstein Archives are highly qualified philosophers, and we cooperate closely with colleagues involved in relevant activities, such as the Forschungs-institut Brenner-Archiv at the University of Innsbruck and the Oxford Text Archive at Oxford University Computing Services.

5. Conclusion, future developments

We believe that by employing the principles described above we have succeeded in satisfying the requirements set forth above:

1. The code system provides a uniform, well-documented and consistent set of editorial principles representing the manuscripts at a level of detail which goes beyond any existing critical edition.
2. Not only does the transcription document the exact *Nachlaß* sources for each and every section and page, the entire transcription will contain links to facsimile reproductions of the relevant pages of the *Nachlaß*.
3. Similarly, each and every section will contain references to the exact locations in the already published editions of the same material wherever relevant.
4. We are transcribing the entire *Nachlaß* and thus do not have to

¹¹ Cf Pichler 1995 for a discussion of this problem

involve ourselves in editorial interventions to reduce thematic redundancy. Since the material is kept in machine-readable form this does not place heavy practical or economical burdens on the user. Each section contains links to earlier and later versions of the same text, thus making it easier for the user to orient himself.

5. Since it is kept in machine-readable form the edition can easily be extended in the case of new *Nachlaß* findings, or modified in the light of new evidence concerning their interpretation.

It should be clear, however, that this is not a traditional edition, but first and foremost a way of providing access to the *Nachlaß* itself. It is for this reason that it is so important to provide an electronic facsimile. In practice such a facsimile could never have been published in print. First of all, with its 20,000 pages it would have been too bulky to handle conveniently. Second, a printed facsimile of this size would not have been able to match the reproduction quality of the electronic facsimile, with full colour, high-resolution images.

Let us finally look at some possible sequels, which lay beyond the current scope of the Wittgenstein Archives but which could easily be implemented once the Wittgenstein Archives' complete version of the *Nachlaß* has become available.

One might include machine-readable versions of the existing printed editions of Wittgenstein's works and link them up to their *Nachlaß* sources. One might also include translations of the published editions to other languages, linking them to their respective original language versions. Furthermore, the electronic edition might include a comprehensive bibliography of secondary sources on Wittgenstein, as well as pointers and links to other digital libraries containing material of relevance for the study of Wittgenstein's writings.

Around such an electronic edition, a networked community of users will soon arise, and it is natural to envisage that they will build up electronic fora for the discussion of Wittgenstein's philosophy, organize events such as seminars and mutual visits, etc.

The users will then serve their own interests as well as the interests of the entire user community by informing the maintainers of the source material about errors and shortcomings, thus making the constant improvement of the electronic sources on Wittgenstein a truly collective effort.

References

Biggs, M. and Pichler, A. 1994 Wittgenstein: Two Source Catalogues and a Bibliography, Catalogues of the Published Texts and of the Published Dia-

grams, each Related to its Sources, Working Papers from the Wittgenstein Archives at the University of Bergen, no 7

Hilmy, S.S, 1987 *The Later Wittgenstein. The Emergence of a New Philosophical Method*, Basil Blackwell, Oxford

Huitfeldt, C. 1991 "Wittgenstein-Nachlaß: Nothing is hidden", *Mitteilungen aus dem Brenner-Archiv* no 10/1991, Innsbruck, pp 93-106.

Huitfeldt, C. 1994a "Computerizing Wittgenstein - The Wittgenstein Archives at the University of Bergen", Johannessen, Larsen and Åmås (eds): *Wittgenstein and Norway*, Solum, Oslo, pp 275-294.

Huitfeldt, C. 1994b "Toward a Machine-Readable Version of Wittgenstein's Nachlaß: Some Editorial Problems", *Beihefte zu editio* Band 6, Max Niemeyer Verlag, Tübingen, pp 37-43.

Huitfeldt, C. 1995a MECS - A Multi-Element Code System, forthcoming in *Working Papers from the Wittgenstein Archives at the University of Bergen*

Huitfeldt, C. 1995b MECS-WIT - A Registration Standard for the Wittgenstein Archives at the University of Bergen, forthcoming in *Working Papers from the Wittgenstein Archives at the University of Bergen*

Kenny, A. 1984 "From the Big Typescript to the Philosophical Grammar", *The Legacy of Wittgenstein*, Basil Blackwell, Oxford, pp 24-37.

Pichler, A. 1994 *Untersuchungen zu Wittgensteins Nachlaß*,

Working Papers from the Wittgenstein Archives at the University of Bergen, no 8

Pichler, A. 1995 "Transcriptions, Texts and Interpretation" - published elsewhere in this volume

von Wright, G.H 1982 "The Wittgenstein Papers", *Wittgenstein*, Basil Blackwell, Oxford (first published in *Philosophical Review* 78, 1969)

von Wright, G.H 1992 "The Troubled History of Part II of the Investigations", *Grazer Philosophische Studien*, 42, Amsterdam, pp 181-192.

WAB 1990-1993 *The Wittgenstein Archives at the University of Bergen: Project Report 1990-1993 and Critical Evaluation*, *Working Papers from the Wittgenstein Archives at the University of Bergen*, no 9

Claus Huitfeldt,
The Wittgenstein Archives at the
University of Bergen
Harald Haarfagresgt 31
N-5007 Bergen Norway

Email: Claus.Huitfeldt@hd.uib.no

Advantages of a Machine-Readable Version of Wittgenstein's Nachlaß

The advantages of a machine-readable version of Wittgenstein's Nachlass can be divided into two subsections: the advantages of the availability of texts in electronic form in general, and the advantages of a machine-readable version of Wittgenstein's Nachlass as opposed to a book edition. I will discuss both these issues, but will lay most emphasis on the latter.

The advantages of machine-readable texts in general can be summarized as follows:

1. Machine-readable texts have the great advantage of enabling easy, cheap, space saving and fast production, reproduction and distribution. Having produced a text in electronic form in my office in Bergen I can send it across the network to my colleague in Japan, and he receives it in a few seconds instead of the time taken by ordinary mail.

2. Machine-readable texts are open for revision, including corrections and additions, change of format, font and style etc. Anyone who works with electronic text processing is familiar with the benefits of the PC when it comes to first the production and later the revision of texts.

3. Machine-readable texts are open for all types of computer assisted analyses, be they statistical, grammatical, stylistic etc., or content analyses. Not only are you able to find within seconds a certain word or string of words in a large text corpus, but more importantly, the computer is able to recognize similar formulations (similar of course according to the definition of similarity you give the computer) and produces indices and concordances for you.

4. Machine-readable texts can be converted into paper printouts and book editions, while books cannot be converted as easily into texts in electronic form. The way from a machine-readable version to the book is always shorter, cheaper and less complicated than the transition from a book edition to a machine-readable version. In addition, it always leaves you with more choices and possibilities than the other way round.¹

Concerning a machine-readable version of Wittgenstein's Nachlass in particular, I would like to emphasize the following:

¹ For further reading see e.g. Butler 1992.

Wittgenstein's Nachlass as a whole confronts us with a series of problems which can only be dealt with by a machine-readable version. Let me give you three examples:

1. On his return to Cambridge in 1929 Wittgenstein started writing on the right hand pages of von Wright's Nachlass-catalogue (von Wright 1982) no. 105. Then Wittgenstein continued on the right hand pages of no. 106. The entire sequence of the texts in nos. 105 and 106 is roughly speaking as follows:

Right hand pages of no. 105

Right hand pages of no. 106

Left hand pages of no. 106

Left hand pages of no. 105

Editing this in book form, one has to decide: Do I want to keep the physical sequence of the text, i.e. edit the text in the order of the von Wright numbers - this would result in an edition where left and right hand pages have little to do with each other and the reader has to skip over one page in order to follow the text order. Or do I want to stick to the text order, which means printing first the right hand pages of no. 105, then the right hand pages of no. 106 etc. One might bring in a third criteria and care only about the content and therefore split up what in the chronological or physical orders are unities.

All of these interests are justified; only, producing a book edition, you have to decide which one you choose - unless you produce three books - while, producing a machine-readable version, you can encode the transcription in such a way, that it allows you to extract from the transcription all the three possibilities.

2. Think of a much discussed case, the edition of PHILOSOPHICAL GRAMMAR, for which the editor Rush Rhees has been - to a large extent unfairly - much criticised. If you were to edit Wittgenstein's philosophical ideas of the early 30s in book form, you would be confronted with at least two problems: 1. Which of the many formulations of an idea are you going to choose? 2. Which one of the arrangements of these formulations are you going to choose? Always the latest ones? If, in the later arrangement, remarks have been omitted, will you include the earlier versions?

In the case of the manuscript sources of PHILOSOPHICAL GRAMMAR, Part 1, you are confronted with a problem similar to the former case of nos. 105 and 106, but still more complicated: Shall one publish the remarks in their chronological order (let's make the erroneous claim that this can be uniquely identified), or shall one follow Wittgenstein's instructions to reorganize the whole (as Rush Rhees did)? If you choose the latter, then you reorganize three

manuscripts (nos. 114, 115, 140) and one typescript (no. 213) into one text, by omitting, selecting, rearranging and replacing remarks with other remarks. What if you had a medium which allowed you to get both a text which corresponds to the physical structure of the sources and a text which represents the result of the intended reorganization? In fact, Rush Rhees might have warmly welcomed the notion of a machine-readable version of what has become PHILOSOPHICAL GRAMMAR.

But the greatest advantage of a machine-readable version is in this case, that it allows you to make so-called hypertext-links which guide you - at a keystroke - from a certain remark to its variants in the same or a different manuscript, instead of - as in the case of a book edition - 1. having to resort to printed concordances or a critical apparatus and 2. having to follow up the reference in the same or a different book volume. This advantage becomes even clearer in the next example.

3. Wittgenstein's Nachlass contains "several 'layers' or stages of basically similar pieces of text"². One example are the many versions of PHILOSOPHICAL INVESTIGATIONS §§ 1-4, whose history is illustrated in the following two graphs.

Graph 1 shows the development in time and relates the dates of the individual versions to the dates of the manuscripts/typescripts in which they are located. Graph 2 marks the positions of the individual versions within the number of pages in the manuscripts/typescripts. MS and TS numbers refer to von Wright's Nachlass-catalogue.

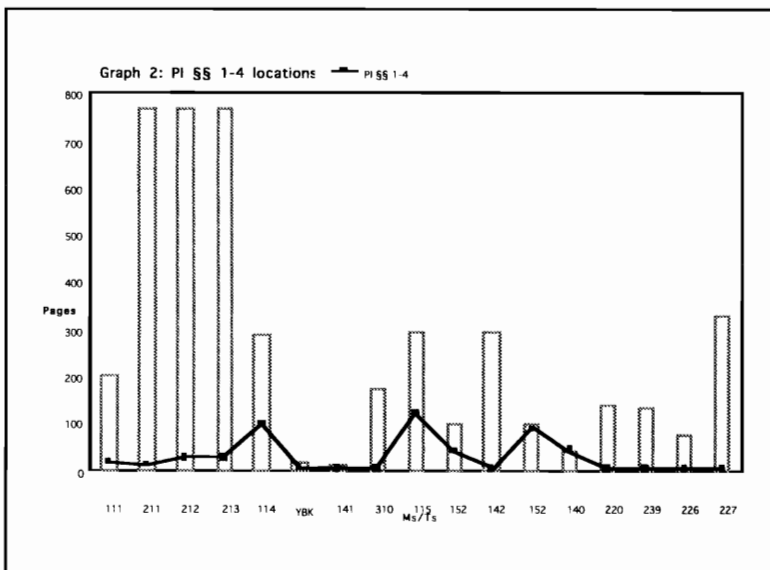
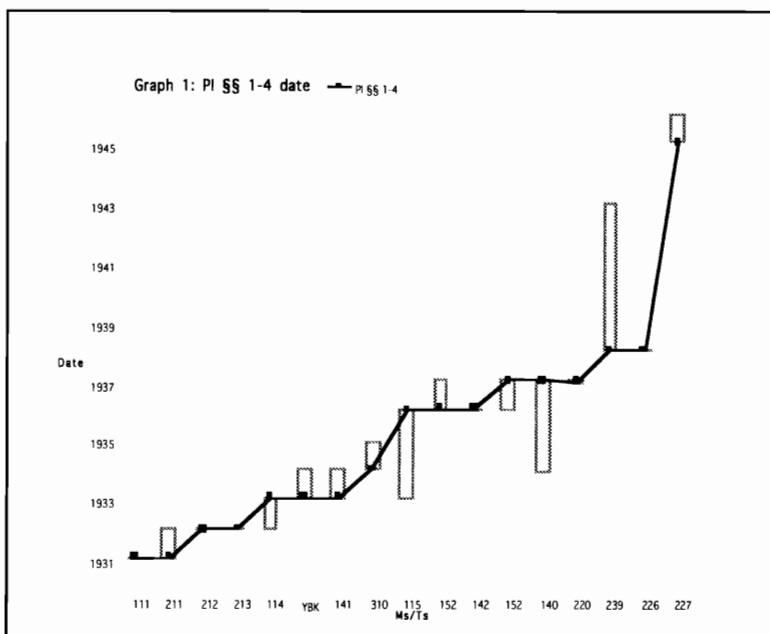
Dictation 311 is referred to by its publication in "The Yellow Book" (YBK) in *Wittgenstein's Lectures: Cambridge, 1932-1935* (since access to the original dictation was not possible). MS 142 was only recently discovered³, and it has therefore not yet been possible to investigate it. However, one can assume that it contains another version of the paragraphs in question.

Not all the indicated versions of the paragraphs contain the complete text of PI §§ 1-4. PI § 1e e.g. stems from MS 115: p. 79f (ca. 1933) and is - in the manuscripts which were available for investigation - first intended as a part of the whole from MS 140 (1937) onwards.

² Pichler 1993: p. 55, Huitfeldt 1994a: p. 38.

³ See Koder 1993: p. 52f.

Advantages of a Machine-Readable Version of Wittgenstein's Nachlaß



It would clearly be a big advantage to be able to jump over from one version to another without having to use concordance volumes and without having to look through different book volumes. Or, what about getting all these versions by just letting the computer search for the word "Augustinus"?

Let me focus here on the following question: What are the texts in this case? Are they the physical units nos. 111, 211 etc.? Or parts of them? Or do we again have to adopt an intertextual, - or "hypertextual" - view regarding what text is, as in the case with the sources of PHILOSOPHICAL GRAMMAR? What about seeing as texts the lines which the single remarks draw throughout the Nachlass?

With a machine-readable version you can obtain both printouts which correspond to the physical division of the Nachlass and printouts which run counter to the physical criteria, but fulfil other criteria such as chronological, thematic, or, as I would wish in this case, evolutionary criteria.

I do not want to say that book editions of Wittgenstein's Nachlass make no sense. But editing the entire Nachlass in book form is an undertaking which would not only cost too much in relation to what it would achieve, but would also be confronted with many difficult editorial problems, for which today's computer technology can provide simple solutions.

Producing a machine-readable version might eventually also result in book editions of the Nachlass. However, one thing is clear from the point of view of the Wittgenstein Archives: A book edition of the whole, without a more comprehensive and open "mother" machine-readable-version to which you can refer, is of little value. But, having this machine-readable version, you are free to choose your stylesheet for desired prints on paper, i.e.: text, which has been deleted in the original, can be printed in the main text (marked as deleted), put in a footnote or entirely omitted; words underlined in the original can be printed as underlined, in italics or with any other emphasis marker; given several alternative readings you have the opportunity to choose exactly how you want to present them. You may only want to display one of the readings; you may want to display them all in the main text; or you may want to display one in the main text, the rest in a footnote, etc.etc.⁴

Let me close with a remark about the impact which I think a machine-readable version of the Wittgenstein Nachlass has on the notion of text and text editing. Machine-readable versions make it more clear to us what texts are and what text editing means: Texts are not objectively existing entities which just need to be discovered and presented, but entities which have to be constructed.

⁴ For further reading see: Huitfeldt 1991, Pichler 1993, Huitfeldt/Letnes/Pichler 1993, Huitfeldt 1994a, Huitfeldt 1994b, Stern 1994.

They are products of both the author and the reader. All that exists in the case of Wittgenstein's Nachlass are scripts which first of all need to be identified, interpreted and organized. Having a machine-readable version of Wittgenstein's Nachlass provides a multiplicity of ways to organize and construct texts, it makes this easy - and it makes it obvious that there is an element of construction.⁵

References

- Butler, Christopher S. (ed.) 1992 *Computers and Written Texts* Oxford: Blackwell.
- Huitfeldt, Claus 1991 "Das Wittgenstein-Archiv der Universität Bergen", *Mitteilungen aus dem Brenner-Archiv* 10, 93-104.
- Huitfeldt, Claus 1994a "Toward a Machine-Readable Version of Wittgenstein's Nachlaß", *Philosophische Editionen. Erwartungen an sie - Wirkungen durch sie. Beiträge zur VI. Internationalen Fachtagung der Arbeitsgemeinschaft philosophischer Editionen (11.-13. Juni 1992 Berlin)* 37-43. Tübingen: Max Niemeyer Verlag.
- Huitfeldt, Claus 1994b "Computerizing Wittgenstein", *Wittgenstein and Norway* 275-294. Edited by Kjell S. Johannessen, Rolf Larsen, Knut Olav Åmås. Oslo: Solum.
- Huitfeldt, Claus/Letnes, Ole/Pichler, Alois 1993 "The Wittgenstein Archives at the University of Bergen", *Conference Abstracts. The 1993 Joint International Conference: The Association for Computers and the Humanities, The Association for Literary and Linguistic Computing. June 16-19, 1993, Georgetown University Washington, DC* 82-91.
- Koder, Johannes 1993 "Verzeichnis der Schriften Ludwig Wittgensteins im Nachlaß Rudolf und Elisabeth Koder", *Mitteilungen aus dem Brenner-Archiv* 12, 52-54.
- Pichler, Alois 1993 "Wittgensteins Nachlaß", *Information Philosophie* 4, 54-60.
- Stern, David G. 1994 "The Wittgenstein Papers as Text and Hypertext: Cambridge, Bergen, and Beyond", *Wittgenstein and Norway* 251-273. Edited by Kjell S. Johannessen, Rolf Larsen, Knut Olav Åmås. Oslo: Solum.
- von Wright, Georg H. 1982 "The Wittgenstein Papers", *Wittgenstein* 35-62. Ox-

⁵ On this point see further Alois Pichler, "Transcriptions, Texts and Interpretation" in this volume.

ford: Basil Blackwell. The catalogue was first published 1969 in *The Philosophical Review* 78, 483-503.

Alois Pichler
The Wittgenstein Archives
University of Bergen
Harald Hårfagresgt. 31
N-5007 Bergen
Norway

Allen Renear

Practical Ontology: The Case of Written Communication

*Denn ich möchte mit dem der Philosophie gegebenen, den geschriebenen und gesprochenen Sätzen, quasi den Büchern, anfangen.
Und hier begegnet man der Schwierigkeit des "Alles fließt".
– Ludwig Wittgenstein (CV 1930)*

1. Introduction

It is not just philosophers, or other persons engaged in discourse manifestly philosophical, who utter the characteristic questions of ontology and metaphysics: "What is that thing? What are the parts of that thing? What sorts of things are there?" Such questions are also posed every day by chemists, curators, mechanics, gardeners, children, and tourists. Or is this a pointless comparison? Is there really anything, apart from the interrogative mood, in common between questions like "What is a carburetor?" and "What is a thought?"

This line of enquiry eventually exercises familiar philosophical issues about the separation of science and metaphysics, the nature and kinds of knowledge in general, and, more particularly, the nature of classification, natural kinds, and explanation, scientific and otherwise. However, the present essay addresses these things only indirectly. Instead what I undertake below is to present a sketch for a preliminary empirical description of certain practices of questioning, classification, reasoning, and theorizing in a particular domain, one where it is impossible to avoid the sense that, however ostensibly practical the agenda of the practitioners, their enquiries are at least partly ontological in character.

The domain is computer text encoding, which is the representation of textual information in order to facilitate word processing, printing, typesetting, linguistic analysis, or other computer-assisted processing. For my purposes its practitioners include both the designers and the reflective users of encoding systems – that is, not only engineers and software designers but also clerks, authors, editors, and scholars. Originally discussions of techniques and approaches in this domain were dominated by a practical concern: using the computer to more efficiently create office documents and printed books. But all along there has

been an interest in the development and accumulation of general theoretical knowledge as well (Engelbart 1963) and in the 1980s this interest grew to include substantial efforts to organize and systematize the practical knowledge and strategies of the field (Reid 1980, Goldfarb 1981, Coombs et al 1987). Now, in the 1990s, there are a number of researchers attempting to more fully theorize this knowledge and connect it explicitly to related bodies of theoretical work in other areas (Sperberg-McQueen 1991, Rohr 1991, Renear et al forthcoming, Huitfeldt 1995).

As questions of how to solve practical problems give rise to theoretical enquiry the views of textuality offered by practitioners recapitulate an interesting, if familiar, evolution from a kind of platonistic essentialism, to a less platonistic and less essentialist (but still realist) pluralism, to positions that seem more pragmatic, constructivist, and even antirealist. In the discourse embodying this evolution many argumentative strategies familiar to philosophers are deployed. These include arguments from hypothetical variation (to discover essential properties), existential instantiation (to display ontological commitment), conceptual involvement (to detect conceptual priority) and others.

The following will be superficial. My only objective at the moment is to recommend computer text encoding to philosophers as a rich field for understanding the interplay between technological practice, disciplinary methodology, scientific theorizing, and philosophical reflection.¹

2. The Domain of Computer Text Encoding

Computer text encoding is the representation of textual information on the computer, typically in order to facilitate the use of computer hardware and software in creating and processing documents such as letters, reports, scientific papers, novels, etc. The sorts of processing made possible by text encoding includes editing, formatting, printing, typesetting, indexing, retrieval, analysis, and so on. This representation proceeds by recording, in the computer's memory, both the linguistic content of a document (such as alphabetic characters and punctuation) and additional information related to this content (such as the identification of titles, paragraphs, and footnotes, or the indication of desired

¹ This essay owes much to conversations and collaborations, spanning many years, with Geoffrey Bilder, Lou Burnard, James H. Coombs, Steven J. DeRose, David Durand, Claus Huitfeldt, Elli Mylonas, W. Richard Ristow, Michael Sperberg-McQueen, and the members of CHUG, the Brown University Computing in the Humanities Users' Group. It also, like so much similar work being done today, owes an enormous debt to the Text Encoding Initiative.

formatting effects). A "text encoding system" is the system of codes which implement this representation.²

It was in the 1960s that text encoding systems began to be used commercially to support computer text processing and computer typesetting. At that time a compositor or author would prepare a data file consisting of "markup" (computer codes specifying formatting information) and "content" (codes specifying the linguistic items of the text, such as alphabetic characters and punctuation). This file would then be processed by a formatting software, creating another data file which would then be transferred to a printer and to printed pages of text.

With the advent of microcomputers in the 1980s interactive word processing quickly came to dominate commercial and academic document development in the industrial countries. These systems also were based on encoding systems, but unlike the "batch formatters" the computer codes were typically hidden from users even during document creation. Naive users were quick to assume that there were no codes involved at all and that they were somehow given direct and "transparent" access to "text itself". But all users of computer text processors are unavoidably involved in text encoding whether they know it or not.

3. The Emergence of Content-Based Text Encoding

3.a Content-Oriented Text Processing

As described above early computer text processing and typesetting systems were "batch" systems. A data file would be prepared and then processed into formatted page images. This data file consisted of (i) the linguistic content of the text to be rendered and (ii) interspersed codes indicating (typically, at least in the earliest systems) formatting effects. The formatting codes would usually be mnemonic expressions such as ".skip 3;" for "skip three lines" or ".indent 5;" for indent five columns. Delimiters such as "." and ";" would allow the computer to distinguish computer markup from text.

It was natural for software designers to simplify the creation and maintenance of the data file described by allowing abbreviations for regularly occurring strings of formatting codes. The formatting codes could then be automatically substituted for the mnemonics during processing. This not only made it easier to input the text, but it also made it easier to maintain and revise the formatting.

² We speak loosely and indifferently, as do the practitioners themselves, of "recording information" and "representing documents" without any attempt to reflect at this point a philosophically coherent account of information and representation.

In some cases these abbreviations, or macros, simply received their identity from the particular combination of formatting effects achieved, for instance "Format 17" might mean "left adjust; Times 14; skip 2". and would be used wherever those effects were called for, whether in a title, a list item, or a verse quotation. But there was a tendency to eventually identify a macro directly with a type of text element rather than with a particular intended formatting effect. For instance, ".prose extract;" would be used to identify a portion of text as a prose extract. Unlike "Format 17", which was directly connected with a particular formatting effect and had no other identity, ".prose extract" was only indirectly associated with its formatting effects via a style sheet which controlled how "prose extracts" were to be formatted.

This approach to text encoding and text processing turned out to have many practical advantages. In the 1970s a number of software designers and computer scientists reflected on these advantages and came independently to the conclusion that the best way to design efficient and functional text processing systems was to base them on the view that there are certain features of texts -- such as titles, chapters, paragraphs, lists and so on -- which are fundamental and salient, and that all processing of texts (such as authoring, editing, formatting, browsing and analysis, and analysis) should be implemented indirectly through the identification and processing of these features rather than by directly inserting formatting codes, or other processing codes, directly into computer files.³ These features have been called "content objects" and this approach to text processing "content-oriented text processing".⁴

3.b Advantages of Content-Oriented Text Processing

It is the wide-ranging nature of the practical effectiveness and efficiencies of the content object approach which initially motivated conclusions about its significance for our understanding of the nature of text -- and which continues to be seen as either explaining or being explained by theories about "the nature of text". In the interests of brevity I present only an outline of advantages followed

³ For early discussions of the application of this approach to text processing see Brian Reid (Reid 1980) and Charles Goldfarb (Goldfarb 1981). For a later overview see Coombs et al, 1987.

⁴ While adjectives like "content," "editorial," "logical," "literary," and the like are frequently used synonymously, they may also be used to mark fine distinctions. The present essay is using "content" to suggest the broadest possible connection with the sense or meaning of the document as opposed to its material form. We continue to use the noun "object" despite the currency of "element" because the latter word has become, in the nomenclature of text processing standards development, the technical term for objects of the sort we are describing -- and we are exploring an intuitive "pre-theoretical" notion, not expounding a stipulated technical term.

by, as an example, an explanation of how composition is simplified.

Advantages for Authoring, Editing, and Transcribing

- Composition is simplified.
- Writing tools are supported.
- Alternative document views and links are facilitated.

Advantages for Publishing

- Formatting can be systematically specified and modified.
- Apparatus construction can be automated.
- Output device support is rationalized.
- Data Portability is maximized.

Advantages for Archiving, Retrieval, and Analysis -- Text as a Database

- Information retrieval is supported.
- Analytical procedures are supported.

For example, composition simplified because formatting considerations make no claims on the attention of the author, editor, or transcriber during composition: rather than needing to remember both (i) the required style conventions relevant to the text being produced and (ii) the formatting commands used by the software in order to format the text according to those conventions, an author or editor instead simply identifies each text element and the appropriate formatting takes place automatically. In the jargon of software engineering content objects let the author or transcriber deal with the document at the "level of abstraction" appropriate to their roles – identifying a text object as a quotation or verse line is an authorial task, making decisions to italicize or center a title is the task of a typesetter's or designer's task.

3.c Markup

Reflecting the perceived theoretical centrality of markup in general, and descriptive markup in particular, the 1980s saw research on the types of markup systems and the classification of markup as well as several projects to standardize markup systems. These efforts eventually resulted in an international standard for defining descriptive markup systems; it is ISO 8879: *Information Processing -- Text and Office Systems -- Standard Generalized Markup Language (SGML)*.⁵

⁵ *Information Processing -- Text and Office Systems -- Standard Generalized Markup Language (SGML)*, ISO8879-1986, International Organization for Standardization (ISO) 1986. See also *The SGML Handbook*, Charles Goldfarb, (Oxford University Press, 1990). SGML has been endorsed by many European and North American

SGML specifies a machine-readable format for defining encoding systems, and for specifying the grammar of document classes. SGML is not itself a set of markup tags for specific content objects, but rather it is a *meta-grammar* for defining sets of markup tags. The technique for specifying these syntactical constraints is similar to the production rule meta-grammar invented by Noam Chomsky to describe natural languages.

A particularly important SGML markup system is that of the Text Encoding Initiative. This markup system was designed specifically for literary and linguistic text and the research and discussion generated by this project has been responsible for much of the analysis and theorizing described below.⁶

4. Textual Ontology

4.a Phase I: Platonism

Although not initially articulated baldly as an ontological thesis, the view a text is an *ordered hierarchy of content objects* is implicit in the early efforts to theorize about the development of text processing and typesetting software (Goldfarb 1981, Reid 1980, Coombs et al 1987) and is clearly implied as a pre-supposition in the standards (SGML, TEI, and others) for markup systems.

The early arguments of the content-object approach to text representation were not intended to make a theoretical point, but rather to promote a particular practice as being more efficient and functional than the competing alternatives. But in the course of debate the partisans of content-oriented text processing inevitably claimed that the alternative representational practices were inefficient because they were based on a false model of text and that their many disadvantages and inadequacies ultimately flowed directly from that flawed conception. Text processing systems based on the OHCO model on the other hand were efficient and functional – because they were based on an accurate view of the real nature of text. (Coombs et al 1987). In "What is Text, Really?" (DeRose et al 1990) all of these arguments are rehearsed and the thesis is made explicit: "Text is an ordered hierarchy of content objects".

Five broad categories of arguments that text is an ordered hierarchy of content objects may be discerned.

governmental agencies, professional organizations, and scholarly societies, and it has been achieving steady acceptance by software developers as well.

⁶ The TEI is sponsored by the Association for Literary and Linguistic Computing, the Association for Computing in the Humanities, and the Association for Computational Linguistics. The first version of the Draft TEI guidelines were issued in July 1990; the final version was released in the Spring of 1994.

Pragmatic. These arguments begin, as described above, with the observation that there are many practical advantages to modeling a text as an OHCO rather than using one of the alternative models. Texts modeled as OHCOs are easier to create, modify, print according to varying specifications, transfer from one application to another, analyze, reorganize, retrieve data from, and so on. These phenomena of the comparative efficiency and functionality of texts represented as OHCOs are best explained, according to this argument, by the hypothesis that texts *are* ordered hierarchies of content objects.

Empirical. Closely related to the pragmatic arguments these begin by observing that content objects and their relations figure very prominently in our talk about texts, and specifically in our descriptions, explanations, theories, hypotheses, and generalizations about texts. For instance, our theories and conjectures about literature make use of terms for chapters, titles, sections, paragraphs, sentences, footnotes, stanzas, lines, acts, scenes, speeches, etc. These have prominent explanatory roles in our talk about texts and in our theorizing about texts and related subjects such as authorship, literary history, criticism, poetics, and so on.. If we resolve ontological questions by looking to the nominal phrases in our theoretical assertions, then we will conclude that such things – chapters, verses, lines, etc. – are indeed the stuff of which literature is made.

Metaphysical. For instance the classic argument from hypothetical variation as used to distinguish essential from accidental properties in scholastic metaphysics, or, in a more contemporary philosophical idiom, to establish 'identity conditions' for objects. Here one notes that if a layout feature 'of a text' (such as leading or typeface) changes, the 'text 'itself' still remains essentially the same, but if the number or structure of the text's content objects changes – say the number of chapters varies or one paragraph is replaced by another – then it is no longer, strictly speaking, 'the same text'.

Productive Power. These are also arguments to the effect that an OHCO representation can generate the other proposed representations (e.g. can be formatted into a bitmap image plain text) but not vice versa, and (ii) this suggests that the target representations are defective.

Conceptual Priority. These are arguments to the effect that understanding and creating text (e.g. reading and writing) necessarily requires grasping the OHCO structure of a text, but does not essentially involve grasping any other structure and therefore the OHCO structure, but no other structure, is essential.

4.b Phase 2: Pluralistic Realism

When researchers from the literary and linguistic communities began using SGML in their work, the tendency of SGML to assume that documents could be

represented as a single logical hierarchical structure quickly created real practical problems for text encoding projects (Barnard et al. 1987). Briefly the difficulty is that while the SGML world seemed to assume that text encoders would always represent the logical (as opposed to physical) structure of a text as a single hierarchical structure, there in fact turned out to be many hierarchical structures that had reasonable claims to be 'logical'. A verse drama for instance contains dialogue lines (speeches), metrical lines, and sentences. But these do not fit in a single hierarchy of non-overlapping objects: sentences and metrical lines obviously overlap (enjambment) and when a character finishes another character's sentence or metrical line then dialogue lines overlap with sentences and metrical lines.

Part of the problem seems to be that there is no univocal sense of 'text', 'book', or 'document' and that consequently these words do not, without further qualification, designate genuine 'natural kinds' that play useful roles in explanations and descriptions of the world. Instead, they have many different senses that play various very diverse theoretical roles.

One response to this situation is a doctrine of "analytical perspectives" where an analytical perspective is, roughly, a natural family of methodology, theory, and analytical practice. Each analytical perspective on a text – e.g. prosodic, linguistic, dramatic – does seem to typically determine a hierarchy of elements.

The principle would be:

AP-1: An analytical perspective on a text determines an ordered hierarchy of content objects.

AP-1 does seem to reflect actual text encoding practices in the literary and linguistic text encoding communities. Usually when text encoders find overlapping objects they assume that they pertain to different aspects of the text.

But does every perspective really determine a hierarchy of content objects? There is a quick way of casting doubt upon AP-1. Discussions of many sorts about texts are filled with characterizations, descriptions, and hypotheses that explicitly relate text objects from different perspectives – chapters and themes, speaker and meter, narrative and paragraphing. Moreover there are even technical terms, such as enjambment and caesura, that specifically refer to relationships between objects from overlapping families. Because a technical vocabulary can be plausibly considered a sign of an analytical perspective the existence of this terminology suggests that some analytical perspectives that contain overlapping objects.

This suggests:

AP-2: For every distinct pair of objects x and y that overlap in the structure determined by some perspective $P(1)$, there exists diverse perspectives $P(2)$ and $P(3)$ such that $P(2)$ and $P(3)$ are sub-perspectives of $P(1)$ and x is an object in $P(2)$ and not in $P(3)$ and y is an object in $P(3)$ and not in $P(2)$.

where: x is a sub-perspective of y if and only if x is a perspective and y is a perspective and the rules, theories, methods, and practices of x are all included in the rules, theories, methods, and practices of y , but not vice versa.

Our simple Platonic model of text – an ordered hierarchy of content objects – has given way to a *system* of concurrent perspectives which decompose indefinitely into concurrent sub-perspectives – with suspiciously baroque complexity.

Moreover, Claus Huitfeldt has pointed out that despite the apparent hierarchical tendency within analytical perspectives, not only is there no assurance that decomposition into ultimately sub-perspectives without overlaps is possible, but we can easily demonstrate that it is not possible: possible element tokens in some perspectives clearly overlap with other element tokens of the same type. Examples of this are strikeouts, versions, and phrases (in textual criticism), narrative themes in narratology, and many many others.

4.c Phase 3: Pragmatism

The realist of course has more important claims to defend than hierarchical structures. It is tempting to say that we have discovered that text structure is determined not by genre fixed in advance, but by the methodologies with which we approach the text. If those methodologies in turn are not simply reliable techniques for eliciting an objective structure, but are rather merely instruments our interests and purposes, then once again we find ourselves moving along a familiar path towards a pragmatic constructivism about the nature of things.⁷ I note this progression without any suggestion that it is normative– subtler and more sophisticated versions of each approach recur at every point and in fact in the coming exchanges on these issues I myself expect to be defending realism.

⁷ The existence and importance of non-hierarchical text-descriptions have been noted by Paul Rohr, in an as yet unpublished paper (Rohr 1991). Rohr, starting without the hierarchical bias imparted by traditional text encoding schemes, and basing his work on deconstruction and modern literary theory, proposes a completely non-hierarchical notion of textual markup.

5. Conclusion

This essay has given a very brief indication of how practitioners and researchers in the area of computer text processing have been thinking about the nature of text. Although we are still at the very beginning of a principled understanding text, documents, markup systems, and computer text processing, I have tried to suggest that practical inquiry in general, and text encoding in particular, provides a rich domain for studying the interplay between technology, methodology, and theory.

In the course of scientific investigations we say all kinds of things; we make many utterances whose role in the investigation we do not understand. ... Our thoughts run in established routines, we pass automatically from one thought to another according to the techniques we have learned. And now it comes time for us to survey what we have said ... now we have to clarify our thought processes philosophically. (CV 1947)

Bibliography

- Barnard, David T., C. A. Fraser, and George M. Logan. "Generalized Markup for Literary Texts." *Literary and Linguistic Computing*, 3 (1988): 26-31.
- Barnard, David T., Ron Hayter, Maria Karababa, George Logan, and John McFadden. "SGML-Based Markup for Literary Texts: Two Problems and Some Solutions." *Computers and the Humanities*, 22 (1988): 265-276.
- Coombs, James S., Allen H. Renear and Steven J. DeRose. "Markup Systems and the Future of Scholarly Text Processing" *Communications of the Association for Computing Machinery*. 30 (1987): 933-47.
- Engelbart, D. C. "A Conceptual Framework for the Augmentation of Man's Intellect." *Vistas In Information Handling*, Vol. 1, P. Howerton, Ed., Spartan Books, Washinton DC. (1963).
- Engelbart, D. C., R. W. Watson, and J. C. Norton. "The Augmented Knowledge Workshop." In *Proceedings of the National Computer Conference* (New York, June 4-8), AFIPS Press, Reston Va., 1973, 9-21.
- DeRose, Steven, J., David Durand, Elli Mylonas and Allen H. Renear. "What is Text, Really?," *Journal of Computing in Higher Education*. 1:2 (1990).
- Fraser, C. A. *An Encoding Standard for Literary Documents*. M.Sc. Thesis, Department of Computing and Information Science, Queen's University, 1986.

- Goldfarb, Charles. "A Generalized Approach to Document Markup." in *Proceedings of the ACM SIGPLAN--SIGOA Symposium on Text Manipulation*, New York: ACM, 1981.
- Goldfarb, Charles. *The SGML Handbook*, Charles Goldfarb. Oxford: Oxford University Press, 1990.
- Huitfeldt, Claus, and Viggo Rossvær. *The Norwegian Wittgenstein Project Report 1988*. Bergen: The Norwegian Computing Centre for the Humanities, 1989.
- Huitfeldt, Claus. *MECS -- A Multi-Element Code System*. Forthcoming in Working Papers from the Wittgenstein Archives at the University of Bergen (seen in draft in October 1992).
- Huitfeldt, Claus. "Multi-Dimensional Texts in a One- Dimensional Medium.," *Computers and the Humanities* 28, 1995: 235-241
- International Organization for Standardization (ISO). *Information Processing -- Text and Office Systems -- Standard Generalized Markup Language (SGML)*, ISO8879-1986, International Organization for Standardization (ISO), 1986.
- Quine, Willard van Orman. "On What There Is" in *From a Logical Point of View* Cambridge Mass: Harvard, 1953.
- Reid, Brian. "A High-Level Approach to Computer Document Formatting." in *Proceedings of the 7th Annual ACM Symposium on Programming Languages*. New York: ACM, 1980.
- Renear, Allen. "Representing Text on the Computer: Lessons for and from Philosophy." *Bulletin of the John Rylands University Library of Manchester*, 74 (1992): 221-48.
- Renear, Allen. David Durand, and Elli Mylonas. "Refining Our Notion of What Text Really Is." *Research in Humanities Computing*. Oxford: Oxford University Press, forthcoming.
- Rohr, P., "The TextBase Paradigm: Architectural Considerations for a Second-Generation Scholar's Workstation" Senior Thesis, University of Chicago, (seen in draft, 1991).
- Smith, Joan. "The Standard Generalized Markup Language (SGML) for Humanities Publishing." *Literary and Linguistic Computing*, 2:3 (1987): 1971-75.
- Sperberg-McQueen, C. Michael. "Text in the Electronic Age: Textual Study and Text Encoding, with Examples from Medieval Texts." *Literary and Linguistic Computing*, 6:1 (1991).
- Sperberg-McQueen, C. Michael. and Burnard, Lou, eds. *Guidelines for the En-*

Allen Renear

coding and Interchange of Machine-Readable Texts. Chicago and Oxford: TTI, 1990, 1993, 1994.

Dr. Allen Renear,
Scholarly Technology Group
Brown University
Box 1885
Providence
RI 02912 Rhode Island
USA

David G. Stern

**New Evidence Concerning the Construction
// the Troubled History //
of Part I of the *Investigations***

My intention was that *some* day all this should be one book.
to make a book (out) of the
My intention was one day to make a book out of all this *these* thoughts
to make a book out of them all
to make out of all of them a book¹

The copy-edited typescripts that were used in the publication of the *Philosophical Investigations* are not to be found in the surviving Wittgenstein papers. In the case of Part II, no copy of the typescript that was used to print it has survived. However, the catalogue of the Wittgenstein papers does include a typescript of Part I, under the name of typescript 227. A second copy of typescript 227 came to light in 1993, and was given to the Wren Library, in Cambridge, where most of the Wittgenstein papers are kept. In the summer of 1993 I compared the two typescripts with each other and the published text of Part I. Both typescripts are carbon copies of the same typed text; both typescripts contain extensive manuscript additions, variants, and deletions in several different hands; neither had been copy-edited. There are significant differences between all three texts. The main aim of this paper is to clarify the relationship between them, and to explain how the second typescript casts light on Wittgenstein's role in the production of Part I. However, we must first consider what was already known about the relationship between Part I and typescript 227.

To the best of my knowledge, there has been very little published discussion of the question of the relationship between typescript 227 and the published text of Part I. The principal sources of information on this topic are to be found in the work of von Wright. However, within von Wright's publications on the Wittgenstein papers, one can discern two very different answers to this question. His catalogue of the Wittgenstein papers and his paper on the origin and composition of the *Investigations* suggest that typescript 227 is the text of Part I, or that any differences are negligible. While the catalogue does not explicitly

¹ Wittgenstein *Nachlass*, MS 160, final page.

discuss the relationship between typescript 227 and the published text of Part I, the entries it contains for the typescripts of each part of the *Investigations* imply that only Part II has been lost:

227. Typescript of Part I of the final version of the *Investigations*. (1944)-45-(46). 324 pp.

234. Typescript of Part II of the final version of the *Investigations*. Probably dictated in 1949. (Missing.)²

Similarly, in von Wright's invaluable paper on "The Origin and Composition of the *Investigations*," typescript 227 is summarily described as "essentially the final version of Part I of the printed work" and the paper does not further discuss the relationship between typescript 227 and the published text.³

However, von Wright's research on the sources of the *Investigations*, much of it unpublished, has led him to very different conclusions. In a footnote that was added to the most recent version of his catalogue, he briefly describes his editorial work with Heikki Nyman on "a rather more critical edition of the *Investigations*"; the reader seeking further information is referred to the introduction to his *Wittgenstein* for further information.⁴ The meticulously edited "Helsinki edition" of the principal sources of the *Philosophical Investigations* reconstructs five successive stages in the construction of the *Philosophical Investigations*. Although the Helsinki edition, completed fourteen years ago, offers an invaluable overview of some of the principal stages in the composition of the *Investigations*, it remains unpublished. However, it has been deposited in research libraries in Bergen, Cambridge, Cornell, Helsinki, and Oxford, and has thus been made available to a number of researchers working on the Wittgenstein papers.

The texts Wittgenstein's literary executors have published do not usually show variant readings, changes in wording, deletions and the like. By contrast, the Helsinki edition aimed to show every significant difference between the text and the printed "final" text. The Helsinki edition of Typescript 227 is followed by 60 pages of closely typed apparatus, listing hundreds of differences between the typescript, the manuscript additions and deletions, and the published text. Most of them are minor differences in spelling and punctuation, but there are also many places where the typescript contains several variant readings and

² von Wright, 1982, p. 48.

³ von Wright, 1982, p. 115.

⁴ von Wright, 1993, p. 504, n. 19. The footnote refers the reader to von Wright, 1982, pp. 7-10 and pp. 111-136 (the introduction and the essay on the origin and composition of the *Investigations*.)

none of them is deleted, or where different words or phrases are chosen in each text. In the preface to their edition of typescript 227, von Wright and Nyman observe that they found no manuscript alterations or corrections by Wittgenstein in the final quarter of the text. These pages, a later addition to a typescript which Wittgenstein had revised in his own hand, contain numerous revisions in other hands. At the end of this painstaking editorial work, von Wright was in no doubt that the typescript which had been used in printing the book had been lost. The preface to *Wittgenstein* contains a compressed summary of the four stages in the revision of Part I that are included in the Helsinki edition. The description of Typescript 227 reads as follows:

(4) TS 227. Later in the year 1945 Wittgenstein expanded the then existing intermediate version with some 400 remarks the manuscript sources of which are in his writings from almost the entire period after his return in 1929 to Cambridge. In a letter to Rush Rhees of June 1945 he still refers to this as "my first volume." As already said, this is the typescript of what the editors (Anscombe and Rhees) called "Part I" of the book as printed. The copy of the typescript from which the printing took place is, regrettably, lost. The edited typescript with comments produced by us in Helsinki is made from the only surviving copy of 227. In editing we encountered some problems arising from the fact that many of the corrections and additions to it were not inserted into this copy by Wittgenstein himself but later transferred from the printer's copy.⁵

In 1992, in an article on "The Troubled History of Part II of the *Investigations*", von Wright returned to the question of the relationship between Typescript 227 and the published text. This time he presented the view that the printer's typescript had been lost in somewhat more qualified terms, implying that his construal of the evidence had been challenged but that he remained convinced of its correctness:

The preserved TS 227 is, for all I can understand, *not* identical with the copy from which the printing of the book took place. It has numerous insertions which are not in Wittgenstein's handwriting. They are presumably copied from the printing typescript which seems to have been the only complete copy at the time of Wittgenstein's death. If I am right, the "original" typescript of Part I thus is lost.

⁵ von Wright, 1982, p. 9.

This regrettable fact also remains something of a mystery. I have been told that checking matters sometime in the 1970s with the publisher, Basil Blackwell, had revealed the entire typescript of the book had been returned by registered mail to one of the editors. Professor Anscombe, however, cannot remember having received it. I myself have not been able to confirm the information about its return after printing.⁶

While I share von Wright's conviction that the copy from which the printing of the book took place has been lost, the only textual reason he gives is not convincing. The book could just as easily have been printed from a copy that had insertions that were not in Wittgenstein's hand. My principal reason for agreeing with von Wright can be stated very simply: there are far too many many discrepancies, both large and small, between the text of typescript 227 and the published text. While typescript 227 has been extensively revised, in several different hands and inks, there are no signs of copy-editing in the typescript, and there are hundreds of differences, both large and small, not only in spelling and punctuation, but also in choice of words, between the text of the published book and the text of typescript 227. Most of the changes are minor, but there are many places where a choice of wording is left open in the typescript, or where one variant is deleted in the typescript yet used in the published text. Another text, in which those changes had been made, presumably another heavily annotated carbon copy of the typed part of typescript 227, must have been used by the printer. However, as we shall see, the second copy of typescript 227 cannot be that text either, for it, like the first copy, frequently diverges from the published text.

At first sight, the second copy of typescript 227 is strikingly similar to the first. Both consist of a carbon copy of the same typescript, which has been substantially modified in a variety of hands. Each of them contains many additions and corrections in Wittgenstein's hand which are usually copied over into the other typescript in another's hand; the changes in Wittgenstein's hand are almost always followed in the published text. Wittgenstein's handwriting, mainly in pencil or black ink, is usually recognizable in the manuscript, but as many of the changes are a matter of deleting variants in the typescript by putting a line through them, one cannot tell by inspection whether they are in his hand or not; this could probably be resolved by an analysis of the ink. However, it is clear that the second typescript greatly reduces the number of cases where there is a

⁶ von Wright, 1992, p. 182.

significant discrepancy between the published text and the revisions in Wittgenstein's hand.

While most of the manuscript corrections in each copy of typescript 227 correspond, both with each other and the published text of Part I, there are also many differences between the published text and one or both copies of typescript 227. Thanks to the the Helsinki edition of the first copy of typescript 227, which lists each departure from the published text, it is not so difficult to identify each point at which the two texts disagree, either with each other, or with the published text. The differences between the two typescripts and the published text can be divided into two main categories. The first consists of those cases where the published text agrees with neither of the surviving typescripts. These differences can be attributed to changes that were introduced at a later stage in the editorial process: the vast majority are very minor changes in spelling and punctuation, presumably made by the copy-editor, that do not affect the sense.

Once one puts the editorial differences between the published text and both typescripts to one side, one finds that the published text almost always follows the text of at least one of the two surviving typescripts. However, there are many places where the revisions to the first and the second copy of TS 227 do not agree, and there does not seem to be any clear pattern as to *which* of the surviving typescripts the published text follows. Given that Rhees and Anscombe jointly worked on editing the printed text, it seems likely that the two surviving typescripts of 227 are the two working copies that they used, and that they are responsible for most of the copying of Wittgenstein's alterations from one text to another. It is very likely that they used a third copy of typescript 227, and that the corrections which are in a hand other than Wittgenstein's in both surviving copies were copied over from his work on the third copy. It is possible that Wittgenstein himself collated the two surviving typescripts, writing up the result in a third copy of the typescript. It is also possible that the editors were responsible for all of the work of producing the printing text. Without further information from Professor Anscombe, or the publishers, some of the final stages of the production of the book will remain a mystery. But the new evidence provided by the second copy of typescript 227 makes it clear that the vast majority of editorial decisions as to the wording of the text were Wittgenstein's. In particular, it turns out that Wittgenstein did edit the text of the last quarter of typescript 227, but only in the second copy of the typescript.

The most striking discrepancy between the second copy of typescript 227 and the published text, is one that it shares with the first copy of typescript 227:

neither contains the published title of "Part I" at any point.⁷ Another significant divergence between the typescripts and the published text concerns the so called "*Randbemerkungen*," or marginal remarks. These slips contained additional remarks that were attached to particular places in the typescript; most of them include instructions as to where they should be inserted in the main text. In most cases, those instructions were not followed, and the material in question was published at the bottom of the page, separated from the main text by a line. For instance, both copies of the *Randbemerkungen* that are printed on pages 11 and 14 explicitly state that they should be inserted at the end of sections 22 and 28 respectively. Other *Randbemerkungen* may have been footnotes: in the second copy of typescript 227, the *Randbemerkung* on page 33 of the published text has a "1" next to the words "On page 60", and one finds a corresponding "1", in the same hand, at the end of section 70, on page 60 of the typescript. In both copies of the typescript, the end of the penultimate sentence in section 142 has a superscripted "1" connecting it with the *Randbemerkung* printed at the bottom of that page. Several *Randbemerkungen* are attached to section 138; one is printed at the bottom of that page, two more on the next page, and two others on page 147; there is no evidence in either typescript to explain why this decision was made. Without the printer's copy of the text, the reason for these discrepancies will remain a mystery; but in view of the very small number of other significant differences between the two typescripts and the printed text, one cannot rule out the possibility that last-minute mistakes were made.

Perhaps the most significant information provided by the second copy of the typescript is that there are very few places where Wittgenstein did not choose between the variant wordings he had written down. Wittgenstein often included alternate wordings in his work: he repeatedly copied variant wordings from his manuscripts, and the broken underlining he used to indicate that he was not quite happy with his choice of words, into subsequent typescripts, continually rewriting and revising. The mass of variants recorded in the two copies of typescript 227 is actually much more characteristic of his writing as a whole than the published text.⁸

References

Stern, D. G. 1995 *Wittgenstein on mind and language*, New York: Oxford University Press.

⁷ Cf. von Wright (1992).

⁸ For further discussion of these issues, see Stern (1995) and (1996).

New Evidence Concerning the Construction...

Stern, D. G.: 1996 "The availability of Wittgenstein's writing," in Sluga, H. and Stern, D. G.: *The Cambridge Companion to Wittgenstein*, New York: Cambridge University Press.

von Wright, G. H. 1982 *Wittgenstein*, Oxford: Blackwell.

von Wright, G. H. 1992, "The Troubled History of Part II of the *Investigations*", *Grazer Philosophische Studien* 42 181-192.

von Wright, G. H. 1993 "The Wittgenstein Papers", in Wittgenstein, L. 1993 *Philosophical Occasions, 1912-1951* 480-510, J. Klagge & A. Nordmann (eds.) Indianapolis, IN: Hackett. Originally published in the *Philosophical Review* 78 (1969) 483-503.

David G. Stern
Department of Rhetoric
University of California, Berkeley
Berkeley CA 94720-2670
USA