

# Re-Discovering Wittgenstein

Deirdre Christine Page Smith, Bergen, Norway

The Wittgenstein Archives at the University of Bergen (WAB) opened its doors June 1st 1990 with several goals: producing a machine readable version of Wittgenstein's *Nachlass*; developing software to assist scholars in locating, viewing and analyzing *Nachlass* texts; developing registration systems and software to present, work with and analyze original textual sources; and establishing links to international Wittgenstein research and computer programming projects with similar text encoding goals. (WAB report 1991) Both the conception and realization of WAB's participation in the DISCOVERY project (Digital Semantic Corpora for Virtual Research in Philosophy) fit with WAB's initial goals. This paper originates in The Wittgenstein Archives at the University of Bergen's (WAB) participation in the eContent+ funded DISCOVERY project. Although I begin with a brief history of WAB's work, my main objective is to present the advantages of DISCOVERY's semantic approach to texts using examples from Ms. 139a, otherwise known as Wittgenstein's 'Lecture on Ethics'.<sup>1</sup>

## Compiling a machine-readable edition of Wittgenstein's *Nachlass*

Preparing a machine readable version of Wittgenstein's *Nachlass* began in Norway already in 1981 under the aegis of The Norwegian Wittgenstein Project (Det norske Wittgensteinprosjektet) which was a cooperative endeavour between the philosophy departments at Norway's four main universities in Oslo, Bergen, Trondheim and Tromsø. Unfortunately, in the late 1980's the materials it prepared could not be made publicly available since the rights to them were disputed, making both gaining permission for distribution and acquiring money to finance the project difficult.

Auspiciously, in the early 1990's, WAB attained both permission from Wittgenstein's literary trustees (G.E.M. Anscombe, Anthony Kenny, Peter Winch and Georg Henrik von Wright) and funding. Software prototypes developed by, and some 3,200 pages transcribed by, the Norwegian Wittgenstein Project formed the foundation for WAB's initial work. WAB's first goal was to transcribe 7,500 pages of the 20,000 pages of Wittgenstein's *Nachlass* and complete the most important elements of the software needed to view them. What is special about this transcription and the challenge to those developing software for the machine-readable edition, was to reproduce these pages as truly as possible. This meant capturing in a digital format the many cross outs, deletions, rewordings, cross references, etc., found in Wittgenstein's *Nachlass*. To this purpose, WAB developed a standard for registering these aspects of Wittgenstein's texts which formed the basis for software allowing the kind of versatile representation the project demanded.

WAB has cooperated closely with the Text Encoding Initiative (TEI) which works toward establishing guidelines for text encoding as well as the interchange of electronic texts. TEI was established in 1988 and its initial set of Guidelines (TEI P1) issued in 1990. Since then, WAB has

been actively involved in and followed the development of TEI guidelines. However, since TEI guidelines were based on a standard (SGML Standard Generalized Markup Language) which restricted encoding possibilities, WAB chose not to follow them. Instead, WAB further developed the MECS (Multi-Element Code System – developed by Claus Huitfeldt) coding system into MECS-WIT, which better suited its needs.

## Bergen Electronic Edition (BEE)

In 1992, WAB and Oxford University Press agreed to utilize the machine-readable version along with electronic facsimiles of the original manuscripts and typescripts toward publishing Wittgenstein's *Nachlass* on CDROM. In 2000, a 6 CD version of the BEE was released. In addition to containing complete sources and drafts (over 50 different manuscripts) for Wittgenstein's published works, the BEE includes previously unpublished or simply unknown material.

By presenting the *Nachlass* in what is termed a "combination of editions" (cf. Pichler and Haugen 2005), the BEE's comprehensiveness, however, extends beyond merely collecting Wittgenstein's body of work into one edition. This is accomplished by providing *Nachlass* texts in two separate but interlinked versions: diplomatic and normalized. The former remains true to the original manuscript and typescript versions, preserving all deletions, over writings, spelling errors and word substitutions. The latter shows editorial corrections, while deleted and overwritten texts are omitted and only the last alternative of two different readings is rendered (earlier alternatives can be viewed upon request). Having these two versions at hand gives the reader insight into Wittgenstein's writing process, and, in doing so, also an enhanced understanding of his thought's development. The same flexibility which allows for interlinked diplomatic and normalized versions enables specialized searches within manuscript sections, whole manuscripts and between manuscript groups as well as date ranges, specified languages, graphic material and mathematic notation. And it is in these latter features we get a taste of the advantages of a semantic approach. (cf. Pichler 2002)

## DISCOVERY

After a period engaged in several EU projects promoting international research and virtual infrastructures for collaborative research and e-learning, WAB co-initiated DISCOVERY in 2005 with a host of European partners. DISCOVERY's goal is to construct a *Philospace*, a virtual meeting place for philosophical collaboration and access to philosophical texts and media. These texts and media, called *Philosource*, are a collection of primary philosophical texts from the Pre-Socratics, to 16th to 18th century philosophical and scientific texts from Descartes, Bruno, Spinoza, Leibniz, Vico, Baumgarten and Kant; a variety of primary material (manuscripts, published works, etc.) from Nietzsche and Wittgenstein and is rounded out by 300 video/sound segments from leading contemporary philosophers such as Gadamer, Deleuze, Vattami et. al. (see: <http://www.discovery-project.eu/index.html>).

<sup>1</sup> For a discussion of difficulties met in its implementation see "XYZ" by Wilhelm Krüger in this volume.

WABs contribution to DISCOVERY consists of 5,000 pages covering 'The Big Typescript' (1929-1934), the Brown Book complex (1934-36), the 'Lecture on Ethics' (1929) and 'Notes on Logic' (1913). What is exciting about the range of texts which WAB is thus preparing for DISCOVERY, is that they capture the consolidation of Wittgenstein's thought between his middle and late (*Philosophical Investigations*) period. Similarly to the BEE, the above *Nachlass* texts will be available in interlinked layers with a study layer added in between (as currently defined this shows editorial interventions regarding spelling, grammar and deletions as well as substitutions and cross outs where they make sense within the context of a sentence). What is new in DISCOVERY is threefold: these texts will be available for free, text encoding has been migrated from MECS-WIT to TEI/XML, and, most interestingly for the purposes of this paper, they will be encoded with semantic tags.

Unlike general text searches, semantic labelling helps researchers locate passages where the term or concept for which they search is discussed, but not literally used. One can of course try to approximate this process by using synonyms or alternative wordings, but many occurrences will still be left out. A somewhat different case would be someone searching for examples of Wittgenstein's use of rhetorical questions. With a regular general text search, one might attempt locating these by searching for a question mark followed by a quotation mark. This would, however, work neither in most standard search functions (where both '?' and '"' are operators) nor the BEE. Even considering the BEE's increased flexibility, only an individual with specialized knowledge of the system and its parameters can achieve such a search. Even assuming one has this specialist knowledge, the search would still not help distinguish between rhetorical questions and e.g. direct quotations or dialogue. Yet another case would be someone looking for instances of metaphor, simile or other literary devices. Although one might locate some of these passages simply by searching for these terms and hoping that they are followed by actual examples, far more will remain hidden. Semantic labelling thus clearly represents an advance in WAB's goal of developing software to assist scholars in locating, viewing and analyzing *Nachlass* texts.

## Re-discovering Wittgenstein

I would like to illustrate these differences using a concrete example from Ms 139a, otherwise known as Wittgenstein's 'Lecture on Ethics'. Although all versions of 'Lecture on Ethics', Ts 207 (published in *Philosophical Occasions* and BEE) and Mss 139a-b (published in BEE), will be available in DISCOVERY's *Philospace* as a *Philosource*, it is Ms 139a which concerns us here.

One of the first problems Ms 139a offers for semantic labelling is its lack of paragraph divisions (this holds for Ts 207 and Ms 139b). Such labelling requires units of text which are restricted in length, both to make labelling more exact and to assist users in locating labels. For this reason, it was necessary to divide Ms 139a into smaller units. This was done according to thematic units (standard English paragraphs) which are well suited for semantic labelling.

Already in the second paragraph of Ms 139a we find examples where semantic tags are superior to simple word searches. In the second sentence we find a seemingly innocent word in two forms: 'communicate' and 'communicating'. This is not exactly a word which has inspired much in the way of secondary literature. However, when we look

at the way it is used, it can function as a synonym for other words such as 'language' and 'explanation', which might be of higher conceptual relevance for Wittgenstein researchers as well as other philosophers who are have a more general interest in Wittgenstein's philosophy. Yet if we take Wittgenstein's use of communication in this paragraph as a whole, we find rather that it falls under two major themes found in the paragraph: 'difficulties met communicating thoughts generally and philosophical explanations specifically' and 'What are the boundaries of communication/language?'. And with this we can begin to answer a question readers may already have asked themselves, "What is the difference between semantic labelling and making a good index?" A good index might next to the entry 'communication' write 'cf. explanation, language' and vice versa. This practice enables the user to find occurrences of words, their synonyms and phrases containing both, however it does not make sense of the use of words or phrases. What semantic labelling does share with making indexes is to identify in advance what will be of interest for a reader and to facilitate its location. Semantic labelling does not stop here. It goes further to abstract an overall meaning from each paragraph based on these individual words and phrases. E.g. 'difficulties met communicating thoughts generally and philosophical explanations specifically' is based on more literal examples found in the text: being a non-native language speaker, saying something which comes from the heart, showing the listener both the road of an explanation and the end/goal to which it leads.

The example I have just described falls under the first, Content, of six categories with which we are currently working. The other categories are: Form, Text Exegesis, History of Philosophy, Philosophical Slogans and Comments. As with my first example, even though we may find examples of all of these in a good index, they would neither be listed by category nor allow the kind of 'sense making' semantic labels do. If we look again at paragraph two of Ms 139a, we find several metaphors (talking from the heart, a hearer seeing the road a philosophical explanation goes down and the end too which it leads) and two rhetorical questions which all would be difficult to locate in a general index or word search. In our current scheme these would fall under the category Form. Other current candidates falling under Form are: definition, example, analogy and simile. Regarding our third and fourth categories, Text exegesis and History of Philosophy we find that Wittgenstein's use of 'human being' ('a human being who tries to tell other human beings something which some of them might possibly find useful') can be traced to a discussion with Maurice Drury around 1930 referring to William James as a 'human person', and Wittgenstein responding, "That is what makes him a good philosopher; he was a real human being." (Goodman, p. 37) Here we have reference to both a conversation contemporary to Ms 139a as well as to a figure in the history of philosophy whose work influenced Wittgenstein throughout his life. Although there is not much in the way of Philosophical Slogans found in this particular paragraph, we find many in the next: Ethics, Aesthetics, value and good. Perhaps more so than the Content category, this one most resembles an index. Yet here again we find the possibility of listing a slogan which, although not literally used, imbues a whole paragraph. The final category with which we are currently working, Comment, is a space where further reflections on the contents of a paragraph as well as clarifications regarding the labelling process itself can be placed. If we return to the second paragraph of Ms 139a, this category could be used to go into more explanatory detail regarding Wittgenstein's use of the road metaphor as something which returns in several guises in his later philosophy: a rule standing as a sign-

post *Philosophical Investigations* (PI) §85, perspicuous representation PI §122, method of projection PI §139, 141, 366, as well in PI §426 where he mentions God and uses other religious analogies to capture something to which we do not have access.

In the context of Ms 139a it becomes evident early on both how well the actual process of semantic labelling fits with important aspects of Wittgenstein's philosophy, but also some dangers it might pose. On p. 4 of Ms 139a Wittgenstein uses Francis Galton and composite photography as an example of the effect he would like to achieve by using synonyms to help communicate his thoughts on the lecture's theme, Ethics. There is a tension in both the example of Galton, Ms 139a as a whole as well as Wittgenstein's philosophy more generally regarding the use of examples both to point to a common thread as well as to illustrate the difficulty of showing something essential to all. (cf. "Ethics, Language and the Development of Wittgenstein's Thought in Ms 139a" in this volume) Like Galton's layering of images one on top of the other to form a composite, the application of semantic labelling to a work offers different views depending on which layers make up the composite. Although this may increase our understanding of a work, we (both encoders and users) must not mistake it for a final statement about a work, see the composite as an image of something real in the sense of a fact of the matter. Although Galton failed through composite photography to show common types of human constitutions (illness, criminality) he did succeed in showing that fingerprints are unique.<sup>†</sup>

## Literature

1991 The Wittgenstein Archives at the University of Bergen: Background, Project Plan, and Annual Report 1990, Bergen: Working Papers from the Wittgenstein Archives at the University of Bergen No. 2.

Goodman, Russell 2004 *Wittgenstein and William James*, Cambridge: Cambridge University Press.

Pichler, A. and O.E. Haugen 2005 "Fra kombinerte utgaver til dynamisk utgivelse: Erfaringer fra edisjonsfilologisk arbeid med Wittgensteins filosofiske skrifter og nordiske middelaldertekster" in *Læsemåder: Udgavetyper og målgrupper, Nordisk Netværk for Editionsfilologer, Skrifter 6* P. Dahl, J. Kondrup and K. Kynde (eds.) Copenhagen: C.A. Reitzels Forlag, pp.178-249.

Pichler, Alois 2002 "Encoding Wittgenstein. Some remarks on Wittgenstein's Nachlass, the Bergen Electronic Edition, and future electronic publishing and networking" In: *TRANS. Internet-Zeitschrift für Kulturwissenschaften 10/2001ff*, Wien: Research Institute for Austrian and International Literature and Cultural Studies (INST). (<http://www.inst.at/trans/10Nr/pichler10.htm>)

Wittgenstein, Ludwig 2000 Ms 139a: "Lecture on Ethics" in *Wittgenstein's Nachlass: Bergen Electronic Edition*, Oxford: Oxford University Press.

Email: [deirdre.cp.smith@gmail.com](mailto:deirdre.cp.smith@gmail.com)

<sup>†</sup> I would like to thank WAB director Alois Pichler and Wilhelm Krüger for discussion and comments on drafts of this paper.