

The voice of Wittgenstein? An exploration of a corpus of Kirchberg contributions 2001-2010

Neuman, Yrsa, Turku, Finland
yneuman @ abo.fi

Falch, Rune J., Bergen, Norway
rune.falch @ fof.uib.no

Abstract

At the time of Wittgenstein's death in 1951 only the *Tractatus Logico-Philosophicus* had been published. Since then, a vast amount of material has been made available and put to use by scholars in a growing mass of commentary and secondary literature. In this article, we introduce and take some first steps in exploring a digital corpus which may be used to investigate the use of Wittgenstein's work and other aspects of the writing of Wittgenstein scholars.

Which sources do scholars cite, and how? On what themes are Wittgenstein's own words mostly used? What do citation practices of Wittgenstein scholars reveal about the implicit conventions of our research community?

The corpus of Kirchberg Wittgenstein-related conference pre-proceedings papers, published by the Austrian Ludwig Wittgenstein Society 2001-2010, were digitized, marked up in XML format, and re-published Open Access by the Wittgenstein Archives in Bergen in 2013. Digital methods of structuring and searching the corpus bring new perspectives on Wittgenstein scholarship within reach.

I. Introduction

This paper explores some ways to investigate the practices behind the use of citations of and bibliographic references to Wittgenstein of a digital corpus of Kirchberg pre-proceeding papers. Apart from bibliographic referencing practices, the corpus already contains or can be developed to provide interesting possibilities of researching quotation practices by Wittgenstein researchers in detail.

'Citations' include quotes (block quotes and quotes marked by quotation marks) and other textual references to the work or words of other authors (Wittgenstein and others).

Citing other authors is not merely aimed at accurate attribution, but is also an aspect of constructing an authorial self. The corpus may provide empirical material for systematic study of the ways in which Wittgenstein enters in Wittgenstein studies, for writers in dialogue with Wittgenstein: as a more or less valued discussion partner, as an adversary or supporting theorist, as a theme or object of study.

II. The Corpus: Kirchberg Wittgenstein Papers 2001-2010

The digitally published corpus of Kirchberg Papers consists of 520 articles in English and German, altogether around 5000 pages of material. The papers were digitized and encoded at the Wittgenstein Archives in Bergen (WAB) under the auspices of the EU-funded project AGORA: Scholarly Open Access Research in European Philosophy in 2011-14. They include most of the papers pertaining to Wittgenstein or to associated themes which were published in the pre-proceedings during the 10-year period 2001-2010. Although ALWS granted WAB permission to republish, WAB contacted all authors about each of the papers, which were only marked up and published upon the explicit permission from the author. For only 12 out of 905 papers, authors declined republication. A further 89 papers were excluded as not Wittgenstein-related. That leaves 284 papers unpublished and unmarked, either because the WAB team weren't able reach the author or because of missing replies. Altogether, the digital corpus includes 70% of the printed Wittgenstein-related material.

The body of material reflects Wittgenstein studies generally to some extent. The length limit of the papers in the corpus differs from normal research articles in that they are shorter, and in particular on that count, results drawn from the corpus should not be expected to be representative for a wider set of scholarly practices.

The articles in the corpus were encoded using XML (Extensible Markup Language), which is a system for tagging and annotating documents which functions widely across the Internet. The annotations are included in the files, distinguished from the text by brackets <> (thus differentiating markup and content) and follow a syntactic structure which can be expanded according to the specific needs for the particular document. The central citation-related encoding elements used in this material are <quote, <ref and <bibl. <quote is used only for direct quotations. <ref indicates a reference to a work in cases of paraphrase or mentioned works. And <bibl is used in the bibliography section of the papers, in which the authors list the various works quoted or referenced (directly or indirectly) in the text. All three elements can be further specified by adding attributes and attribute values; e.g. a quote may be tagged with the attribute "block" (in contrast to inline quotations which are distinguished from body text using quotation marks).

Such encoding makes references in the body text automatically retrievable. During the digitisation process, bibliographic references to references in the text which had been left out from the bibliography by the author were included and marked up as completions (<bibl type="plusPrimary">) by WAB's markup team. The bibliographic references were tagged with acronyms (W-TLP, W-PI, W-OC etc.).

During the first mark-up process, the attributions in the body text were not annotated using the acronym system due to time constraints. All quotes are marked up as quotes, but the accompanying references were tagged only occasionally. This means that in a paper, citations which were marked for instance "*ibid.*" or "(Wittgenstein 1967)" by the author are not automatically retrievable, sortable and categorizable. Although it would be possible to sift through all the material manually, this is a demonstration of one limitation of the digital material: it is a discovery which suggests the need for additional markup in both the existing and future material in order to alleviate searching and navigation.

III. The corpus in context

Ken Hyland (2004) has conducted corpus studies on articles from a range of disciplines in order to elucidate the differences in academic culture between disciplines. As a discourse analyst, he understands academic attribution as a mode of interaction. In a citation study on philosophy journals, he found that in philosophy papers, when cited works were presented, 89% of citations were incorporated into the article in the form of a summary, 8% as generalizations (from several sources), 2% as quotes (direct quotes with quotation marks) and only 1 % in block quotes (marking an extended use of original wording) (Hyland 2004: 26). In Hyland's philosophy corpus, consisting of merely 10 research articles from journals published in 1997, the prevalence of citations were 10,8 per 1000 words.

Assuming that Kirchberg conference visitors follow their usual citation practices, Hyland's corpus suggests that most citations in the Kirchberg papers will occur as indirect discourse (summaries) and only around 3 % in direct quotations (block quotes and quotes marked using quotation marks).

In philosophy, "knowledge is constructed through a dialogue with peers in which perennial problems are recycled through personal engagement" (Hyland 2004: 36), and "author visibility" is high compared to other disciplines: authors cited are explicitly and repeatedly mentioned. This suggests that investigation of citations of Wittgenstein should be possible on a corpus like the Kirchberg Wittgenstein Corpus.

IV. Citation practices in Wittgenstein studies

Our first analyses show that there is a wide variety in practical citation practice. To some extent, Wittgenstein's writings are referred to using a system of abbreviations (TLP, PI, PU etc.). In parallel, Wittgenstein's works are often referred to in the text by publication year (Wittgenstein 1997). In several of these cases, the reference in the body text is not complete (author, year, page).

The corpus also shows that in 146 of 520 papers (28%), one or more references to Wittgenstein's works or the Nachlass are left out of the bibliography (annotated <bibl type="plusPrimary">), leaving the reader to him- or herself to find the source. Rather than widespread sloppiness, this citation practice may be - in line with Hyland's perspective - a display of what is not explicitly said, that some Wittgenstein sources are taken for granted as the backdrop of discussion and therefore not perceived as standing in need of explicit and detailed referencing. It is possible to investigate this conjecture, however the references added would need to be explored in detail in their context.

The practices of citation discussed above are interesting, but more importantly, they may be of crucial importance for research practices in the future. Furthering the use of specific references may contribute to a better discussion, especially of Wittgenstein in studies with exegetic elements: a unified practice, i.e. the use of a standard referencing model, would make citations available in searches to a much greater extent than they are today, and in this way, research dialogue between readers of specific sections, works, or even paragraphs could be improved.

By subtracting the bibliographic details in the quotes for each of the papers from the bibliographic lists and comparing this with the remaining bibliographic items not paired with a "reference" (a

reference which is not a quote), one would arrive at a list of citations which are not yet localised. This set of citations, if marked in an unclear manner or not at all in the body text, would be material evidence produced by corpus analysis of a direct imperative for improvement in citation practices.

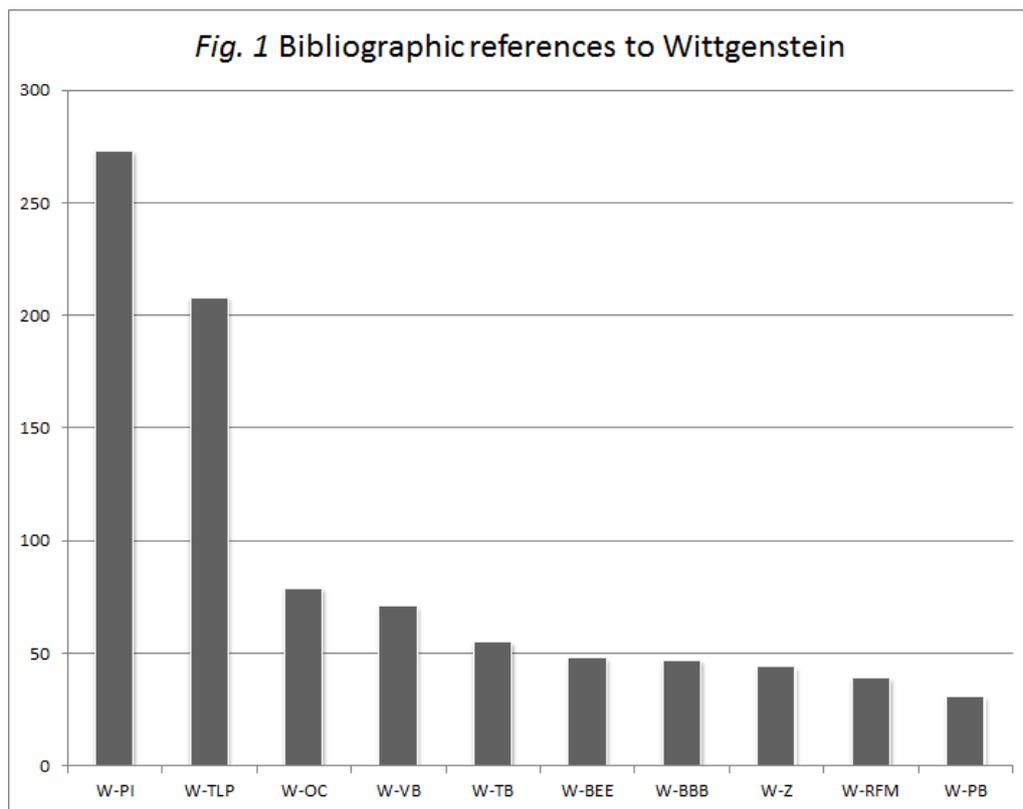
Not only would author application of a standard set of abbreviations be needed, but completing the XML markup of the corpus would improve not only inquiries into citation practices or other quantitative corpus analyses, but also searches in the digital material by readers: a uniform and thorough markup practice would make searches less vulnerable to each reader's "searching style", since references could be categorized as instances of the same "general concept" of a work: e.g. "PI", "PU", "Philosophische Untersuchungen", "Wittgenstein 1951" and even "ibid." could all be marked up as instances or tokens of the "class" or "type" Philosophical Investigations. Cited works in parts or whole, or even particular remarks would be distinguishable.

V. Wittgenstein quoted

Picking out all those quotations which are marked with quotation marks and as block quotes automatically gives a result of 5348 items (including Wittgenstein and other authors). Of these, quotes of Wittgenstein which appear more than 10 times altogether are scarce. The most common one we were able to identify is §7 in the TLP, which appears 14 times, in different variations: "What we cannot speak about we must pass over in silence" (3 instances), "Whereof one cannot speak thereof one must be silent" (8 instances), "Wovon man nicht sprechen kann..." (3 instances). Our conclusion from this exercise is that no particular quotation or set of quotations stand out as statistically significantly the most common. This is an interesting finding as an expectation we had was that there would indeed be some clear trends. At the same time, this is a demonstration of the diversity of research themes which the corpus makes available.

A difficulty in retrieving quotes automatically lies in identifying the different variations due to the number of translations, differences in interpunctuation, the scope of material selected in the quotes (a few words or a part of a remark (which may of course also overlap in part)) etc. The variations of the quote above were extracted by retrieving the marked quotes from the XML files, assigning each a code, stripping them of punctuation etc., and ordering them in an Excel file, then skimming the quote column ocularly and finally searching by a string of the German counterpart. Furthermore, we used a concordance tool for corpus linguistics, AntConc, which confirmed the result that any exceptionally commonly quoted paragraphs do not seem to exist.

The two most common bibliographic items consisting of Wittgenstein's works refer to *Tractatus* and *Philosophical Investigations* (fig. 1 below). Note that W-BEE, Bergen Electronic Edition (48 references) includes the other works and that also some other sources may be overlapping.



VI. Citation recommendations

A unified and consistent citation practice amongst authors would not only make citation research easier, but more importantly, it would give fellow researchers now and in the future better tools to search for and find and enter dialogue with authors using the same material. Consistent use of standardized modes of referencing would also aid digital material producers and the markup community in producing this easily searchable material. There is already to some extent a standard citation practice, which originates in von Wright's catalogue of Wittgenstein's works (von Wright 1982/1993), but there is room for a more consistent practice and renewed recommendations.

The explorations of the corpus, in combination with other sources, suggest the following citation practice regarding Wittgenstein's works.

- Using abbreviations ("citation keys") following the Pichler, Biggs & Szeltner 2011 bibliography <http://www.ilwg.eu/files/Wittgenstein_Bibliographie.pdf>

This practice is recommended by the publications *Wittgenstein Studien* and *Nordic Wittgenstein Review* (although not consistently applied in the latter case).

- Including abbreviation references with every quote or other form of attribution in the body text, for instance "(TLP 2.1)", hence avoiding 'ibid.' and "Wittgenstein [version publication year]" altogether.
- When possible, enter abbreviation references also for summaries of reasoning in the body text.

- Including in the bibliographic items all references to Wittgenstein's work (avoiding mere BEE reference).
- Avoid own translations when not necessary.

Further recommendations may include the following:

- For digital primary resources, for instance to Wittgenstein Source, provide direct links: Wittgenstein, L., Ms-141, <http://wittgensteinsource.org/Ms-141_n>. It would be advisable to refer to Wittgenstein Source manuscripts rather than providing one bibliographic reference to BEE.
- For republished digital secondary resources, follow the Wittgenstein Repository recommendation: [author]: "[title]". In: [Resource title], <link>. [Republication body, year]. Original publication in: [complete traditional reference].

VII. Concluding words

This preliminary exploration into a recently created digital corpus and its possibilities of development shows how it may open routes for both recommendations for technicalities of scholarship but also to detailed studies of Wittgenstein reception. With our presentation we hope to have inspired to more research and a continued dialogue on citation practices within our field of research.

Acknowledgements

Øyvind Liland Gjesdal, Senior executive officer, University Library at the University of Bergen, Norway (Analyses and XML expertise)

Alois Pichler, Wittgenstein Archives, University of Bergen, Norway

Christian Erbacher, University of Siegen, Germany

The research behind this article has been funded by the Academy of Finland, and by the Department of Philosophy and the University Library at the University of Bergen.

References

From the ALWS archives: A selection of papers from the International Wittgenstein Symposia in Kirchberg am Wechsel <<http://wittgensteinrepository.org/agora-alws/>>. Republication by the Wittgenstein Archives at the University of Bergen, 2013.

Hyland, Ken 2004 *Disciplinary Discourses*, Ann Arbor: Michigan University Press.

Nordic Wittgenstein Review, Author Guidelines

<<http://www.nordicwittgensteinreview.com/about/submissions#authorGuidelines>> (accessed May 14, 2016).

Pichler, Alois, Biggs, Michael A. R., and Szeltner, Sarah Anna 2011, "Bibliographie der deutsch- und englischsprachigen Wittgenstein-Ausgaben", *Wittgenstein-Studien* 2 (1).

<http://www.ilwg.eu/files/Wittgenstein_Bibliographie.pdf>

Wittgenstein, Ludwig 2000 *Wittgenstein's Nachlass, Text and Facsimile Version, The Bergen Electronic Edition*, Oxford: OUP. (BEE)

Wittgenstein Repository, ed. Wittgenstein Archives at the University of Bergen 2014

<www.wittgensteinrepository.org> (accessed May 15, 2016)

Wittgenstein Source, ed. Wittgenstein Archives at the University of Bergen 2015

<wittgensteinsource.org> (accessed May 15, 2016)

Wittgenstein Studien, Notes for Contributors. International Wittgenstein Society, ILWG

<<http://www.ilwg.eu/files/Notes%20for%20Contributors.pdf>> (accessed May 14, 2016)

von Wright, G. H. 1993 (1982) "The Wittgenstein Papers - A Catalogue", in: Klagge, James C. and Nordmann, Alfred (eds.) *Ludwig Wittgenstein: Philosophical Occasions*, Indianapolis and Cambridge, MA: Hackett.