

# **(Re)-Constructing the Semantic Architecture of Wittgenstein's *Vermischte Bemerkungen* by Syntactic Analysis**

KERSTIN MAYR, INNSBRUCK

## **1. Introductory remarks**

In the context of a cooperative project between the Brenner-Archive at the University of Innsbruck (FIBA) and the Wittgenstein Archives at the University of Bergen (WAB), a computer supported qualitative analysis of Wittgenstein's *Vermischte Bemerkungen/Culture and Value* is being carried out.<sup>1</sup> This is done with GABEK (*Ganzheitliche Bewältigung von Komplexität / Holistic Processing of Complexity*, see <http://www.gabek.com/>), a method based on the theory of linguistic *gestalten* (Zelger 1999), and its computer implementation WinRelan (*Windows Relationen Analyse*). By a content/semantic analysis of the material an integrated view of individual aspects of Wittgenstein's originally scattered and often private notes and remarks on various topics, which were assembled, edited and published by von Wright in *Culture and Value* (1994), could be obtained. Wittgenstein is in no way the editor but yet the author of this compilation which is nevertheless treated to a holistic analytical approach. At best this investigation might reveal the intentions of von Wright, were it not for the fact that he did confess to be uncertain about the desirable outcome.

It is the analysis's basic intention to look and investigate philosophically relevant semantic fields (patterns) within the remarks from which we could then gain semantic knots acting as thematic anchors for further investigations in Wittgenstein's *Briefwechsel/Complete Correspondence* (BW) and the *Bergen Electronic Edition* (BEE).<sup>2</sup>

## **2. What a text analysis can do**

Georg Henrik von Wright saw himself faced with the problem of the arrangement of these numerous notes scattered among the philosophical

and biographical texts Wittgenstein had left. In his foreword to the first edition of *Culture and Value* (1977) von Wright wrote:

It was a decidedly difficult task; at various times I had different ideas about how best to accomplish [the selection and arrangement of these remarks]. To begin with, for example, I imagined that the remarks could be arranged according to the topics of which they treated – such as “music”, “architecture”, “Shakespeare”, “aphorisms of practical wisdom”, “philosophy”, and the like. Sometimes the remarks can be arranged into such groupings without strain, but by and large, splitting up the material in this way would probably give an impression of artificiality. (von Wright 1977, ix)

In some cases it seems difficult to decide what Wittgenstein was referring to and therefore any kind of classification or attribution to certain topics only by reading through these notes would lack any rule- or criteria-based investigation. This is now where computer based text analysis comes into play. A text analysis tool could be used to identify the context and importance of text units. Thus, we try to investigate any inherent semantic and topical structure of this seemingly loose collection applying clear and transparent criteria. We are not primarily interested in analyzing the circumstances under which the *Vermischte Bemerkungen* were written and later combined. The texts themselves will be our first and only fields of investigation – at least at this stage. Again, this is to say that even though Wittgenstein was not involved in this compilation it deserves all the care due to a coherent product.

It seems a strange thing to concede that all these remarks are taken out of context and proceeding to treat their aggregation as a significant whole. However, despite being a loose collection, the textual analysis of these remarks assembled in *Culture and Value* could result in something like topical signposts hinting at recurrent themes in Wittgenstein’s corpus. In this way we could gain access to clusters in the corpus which may be indicative of philosophical topoi hitherto uninvestigated as such. Thus, once a first analysis will have been completed, framing and re-framing into the larger context of text genesis as well as Wittgenstein’s writings and letters should follow.

With Wittgenstein's works in general and with the *Vermischte Bemerkungen* in particular the question is again one of textuality. The question what constitutes a text (by Wittgenstein), is becoming even more virulent with the *Vermischte Bemerkungen* since the text itself was not arranged by Wittgenstein but edited posthumously. The problem, now, is to locate this text's (or rather these text units') central cores holding the essentials of its meaning(s). Before any attempt at an interpretation of this text can be made, the semantic "hot spots" have to be identified. Once uncovered, what we would get are various semantic fields and meaning-structure(s). Frequency as well as the degree of cross-references between different semantic fields may indicate probable semantic and thematic "centers of gravity". Thus, what a semantic text analysis can do, is looking for a "textual architecture" and trying to hint at crucial text criteria such as cohesion, coherence, intratextuality and – to some extent – intertextuality within *Vermischte Bemerkungen*. So we could finally reveal one or more thematic "red threads" and an arrangement of the remarks according to various topics would be less artificial or at random.

Any interpretation of the text arises in that the topical building blocks (semantic fields) are understood as the meaning-structure(s) of the text. Metaphorically speaking, every city (text) consists of various houses (semantic fields) and its inhabitants (keywords), performing with inhabitants (keywords) of other houses (semantic fields) all contained within one and the same city (text). Each of them is of different importance in the structure of the text. However, content analysis applies a set of techniques to a given text to determine the following:

- the identity of the main houses and inhabitants (semantic keywords and fields),
- the relations in which they stand to each other (constituting semantic networks),
- the hierarchy of these relations and how they evolve (forming the textual framework).

Content analysis consists in revealing the foci within a certain text, i.e. its meaning. This necessarily implies two things. First, there must be a theoretical conception of the text describing both the textual organization of the things said as well as the structural organization of the thought-processes of the author. Secondly, this implies the use of a tool which rigorously tries to exclude the subjectivity of the investigator to a maximum extent. In the case

of the *Vermischte Bemerkungen* both can best be done by rule-based text-coding. Since the actual version we have is a mere construct, the question is if the various text fragments may hint at a larger underlying textual (and philosophical) conception or “hypertext”, which would finally legitimate the appliance of the concept of “text” to the *Vermischte Bemerkungen*.

### **3. Applying GABEK/ WinRelan to the *Vermischte Bemerkungen***

The advantage in using the GABEK/ WinRelan method lies within the fact that it allows a hierarchically structured presentation of a highly complex text and its network layers. The main objective of this analysis is to clarify and highlight content-related (semantic) interdependencies and intervening variables – hypotheses on interdependencies can be generated in a further step. Whereas other semantic text analysis tools are designed to help the researcher identifying particular components of natural language (morphemes, words, syntax, semantics etc) and call upon a number of pre-defined rules, GABEK is a method in which themes (or classes of concepts) as well as causal inter-relations among themes are encoded. The method involves a three step encoding process.

#### **3.1 The encoding process**

##### **3.1.1 Coding of Keywords**

When using WinRelan the first step is to divide the text up into chunks, which are then transferred onto so called index cards (see Fig. 1). Each card should include a semantically closed statement<sup>3</sup> whereby the length of text units represented on these cards is determined by the number of keywords. Keywords are words that constitute the semantic content of a text and are – in general – easily identified.

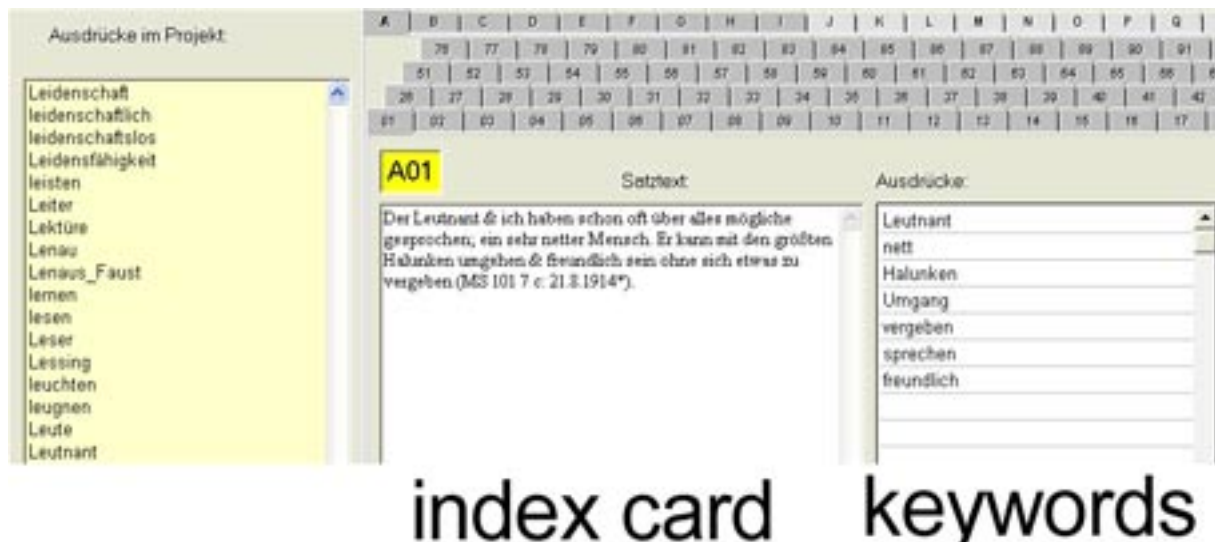


Fig. 1: Index card and corresponding keywords

What we finally get is a kind of concordance, so we can, for instance, list all words in alphabetical order (see Fig. 2) which are repeated in the text two or more times, or create a chart showing the words in the text ranked in order of their frequency of occurrence (see Fig. 3).

Gesamtliste	Reduzierte Liste	Synonymliste				
1	ergründen	F20				
3	erhalten	C77	F63	G13		
1	Erholung	G86				
5	erinnern	B32	B83	B87	F25	G77
4	Erinnerung	C52	E28	G64	G65	
1	erkaufen	D45				
3	erkennen	A02	F80	I02		
3	Erkenntnis	C29	C77	E71		
13	erklären	A61	E11	E63	F67	F92
14	Erklärung	B59	C45	E27	E63	F90
1	erlauben	A65				
5	erleben	C15	E26	E27	E30	F93
4	Erlebnis	E24	E26	F91	F96	
1	Erlebnisinhalt	F96				
1	erlebt	H46				
1	erleichtern	E67				
1	erlernen	G37				
1	erleuchten	C19				
1	erliegen	A27				
1	erlösen	C67				
2	Erlösung	C29	C69			

Fig. 2: Keyword list in alphabetical order

Gesamtliste		Reduzierte Liste	Synonymliste				
141	ich	A02	A06	A13	A17	A19	
64	Menschen	A24	A32	A33	A40	A62	
52	glauben	A21	A93	B54	B60	B85	
51	Mensch	A02	A08	A17	A39	A41	
51	sagen	A08	A25	A28	A32	A56	
45	verstehen	A02	A21	A59	A60	A61	
44	Leben	A34	A35	A95	B07	B08	
43	sehen	A20	A25	A33	A34	A35	
39	denken	A09	A15	A19	A20	A31	
37	schreiben	A55	A59	A60	A75	B06	
33	gut	A06	A07	A17	A21	A22	
32	Gott	A35	B50	B76	C40	C50	
32	Musik	A26	A46	A71	B37	B42	
31	andere	A09	A37	A63	A81	A86	
30	Sinn	A61	A83	B14	B31	B41	
29	Zeit	A19	A47	A49	A50	A70	
28	Gedanken	A57	B23	B40	B45	B62	
28	Sprache	A02	A75	A82	A88	A89	
26	Ausdruck	A03	A47	A49	B98	C58	
25	zeigen	A59	A82	A85	A86	A95	
25	scheinen	A37	A68	A90	B23	B49	

Fig. 3: Keyword list according to frequency

Both lists derive their power for analysis from the fact that they allow us to see every place in a text where a particular word is used and therefore help the researcher to anticipate relevant semantic fields for a subsequent detailed analysis.



Fig. 4: Analysis result: keyword list

As a rule one would have between three to nine keywords on each index card<sup>4</sup>, which would mean approximately three sentences<sup>5</sup>. As GABEK/WinRelan is mainly used for analyzing spoken text data, the keywording and coding of Wittgenstein's dense and highly complex remarks turns out to be quite a challenge. Where one would normally have several sentences on one index card, with Wittgenstein it is often necessary to have only one or two sentences on one card. As long as we are merely aiming at an identification of keywords in order to compile a keyword list (e.g. for a concordance or register), showing the frequency in usage of specific terms, this is fine. One could now argue that in order to generate a keyword list ordered by the frequency of occurrence, we would not have to use such a time consuming method; we could just run a simple word search program to create a concordance. The problem with this kind of analysis is that such a tool would only look for the occurrence of a specific term, e.g. "Geist", but would not distinguish between its different meanings. Thus, such a list of keywords out of context would not meet our needs.

Although we do not fully stick to the GABEK rules with the chunking of text units, we have to follow the rules in regard to further data processing. Now this is where WinRelan meets its limits. Especially when it later comes to building linguistic *gestalten*, i.e. doing a strictly rule-based summary of the contents of those index cards sharing again five to nine keywords, index cards with too many sentences and equal or different key-

words respectively will turn out to be useless. Why? This has to do with the algorithm used for the virtual grouping of semantically fitting index cards. We would either get too many “virtual” piles of index cards which would not help us to cluster interrelated terms or topics or we would get too large piles which would mean that we were too imprecise in coding. The latter would only show that all terms or topics are interrelated in one way or the other but that would not help us to find patterns of relevant<sup>6</sup> inter-relations and key concepts.

### 3.1.2 Coding of Evaluations

After all keywords have been coded, two more codings – based on the first one – have to be done. In a second step of the coding process an evaluation coding procedure has to be executed. This means that we are looking for keywords which are either positively, negatively or neutrally<sup>7</sup> connotated. Therefore we will create two lists; one in which we mark how keywords referring to real phenomena, states etc. are evaluated and another in which we mark evaluations of hypothetical states.

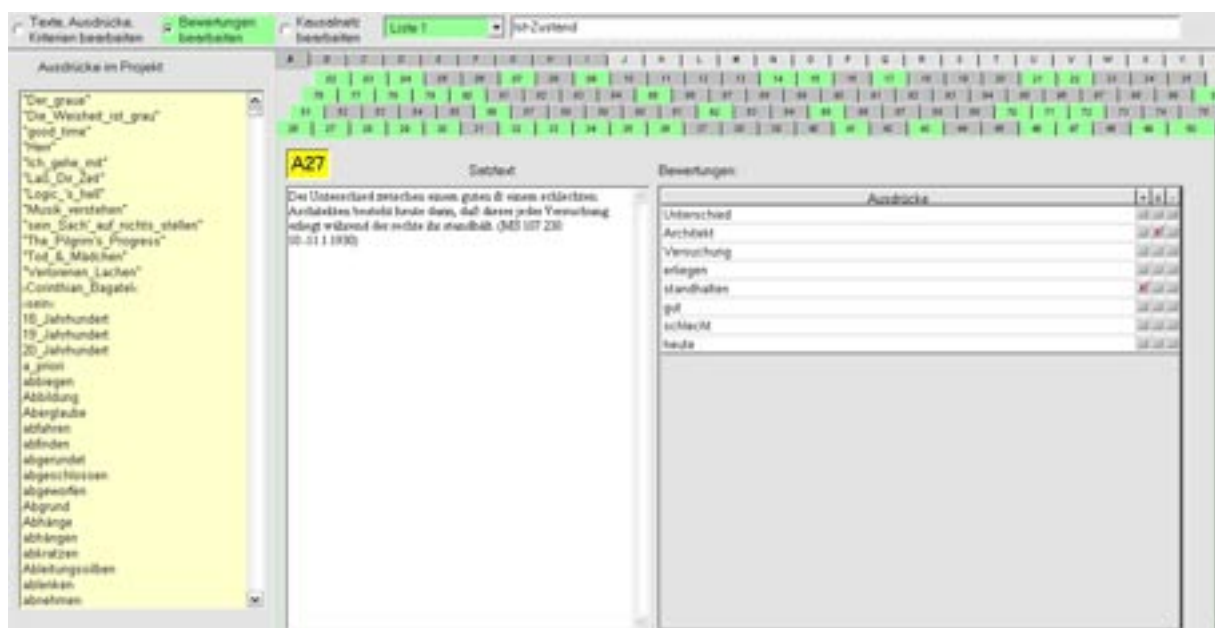


Fig. 5: The evaluation analysis of the real situation according to A27. The current situation is judged positively in the text regarding “standhalten”.

Now, why should we like to have these lists of evaluations? If these operations are executed on all the texts available, then we could receive an infor-



mative and differentiated estimate of the current or desired status of certain states and phenomena. So we could argue that Wittgenstein held rather a positive attitude towards e.g. “Glaube” (religious belief), “Dummheit” (stupidity, inanity) or “Natur” (nature) and a predominantly negative attitude towards “Wissenschaft” (science), “Mahler” and “Philosoph” (philosopher). Although such an analysis of evaluations might have been of interest at first sight – since the title holds the term “value” –, the results of the coding of evaluations for the keywords identified in *Vermischte Bemerkungen* are of no significant (quantitative) relevancy. Thus, although one might think that the title could allude to “value” in the sense of setting the amount of value of a certain topic or idea, it is rather that the word indicates value in regard to moral principles and beliefs or accepted standards of a social group. The latter assumption could be backed up by looking at the context of the word “value” within *Vermischte Bemerkungen* themselves.

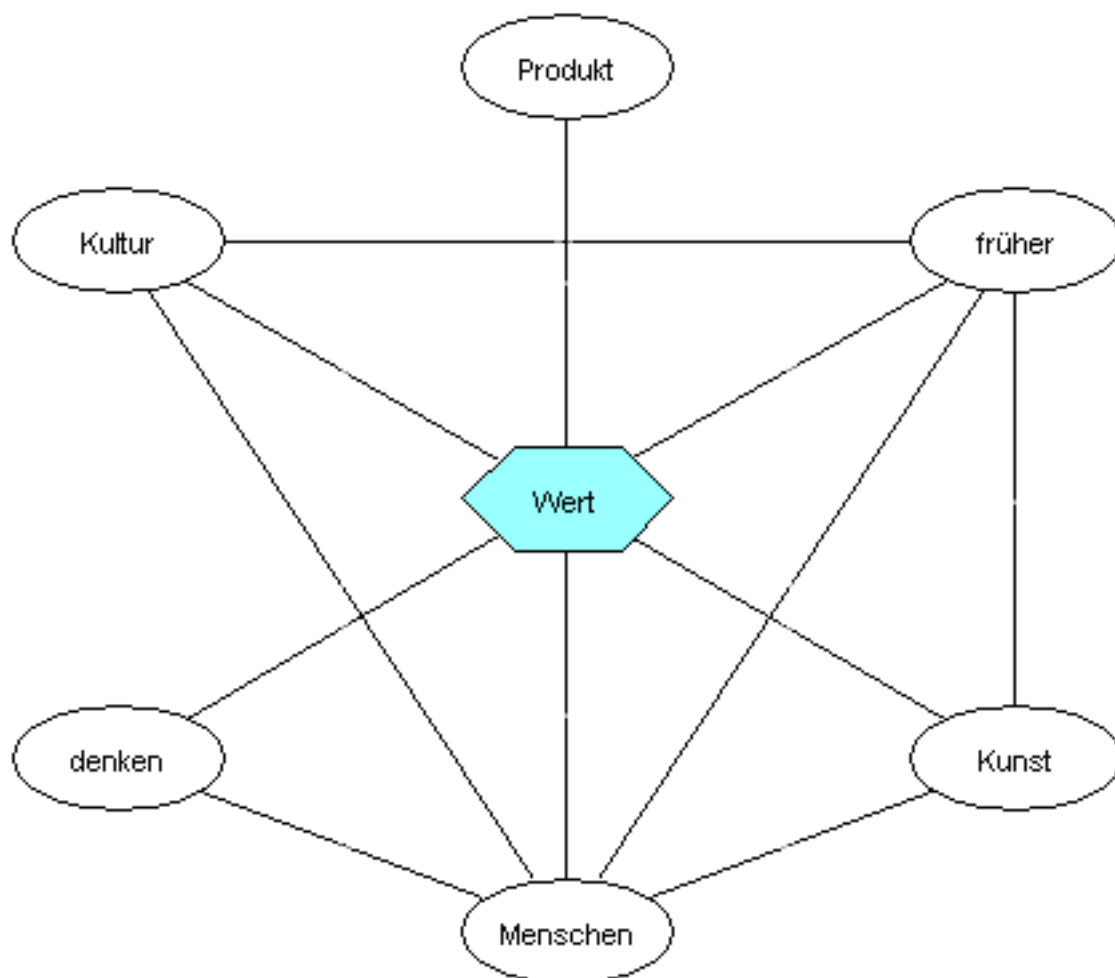


Fig. 6: Association graph to “Wert”

As can be seen from Fig. 6, in which we see a network of terms co-occurring with “Wert”, the context suggests that “value” relates to social terms and fields. The graph displays with which other words the word “Wert” (“value”) co-occurs several times. Although the scheme (Fig. 6) does not yet shed light on the quality (i.e. in which way these words – and the concepts they allude to – refer to one another) of their co-occurrence, it gives first information on the closer context of a word in use – what is of special importance when dealing with Wittgenstein, who emphasised that the meaning of a word is given by its use in a language game in a concrete environment.

### 3.1.3 Coding of Inter-relations and Inter-dependencies

Apart from keywording and the coding of evaluations, one final coding has to be done if we are interested in the nature of co-occurrence of our coded keywords. Usually this last step of the coding process aims to mark “relations” between keywords. Now, what is meant by the “relation”, “inter-relation” or “inter-dependency” between keywords? These terms refer to the syntactic-grammatical function and order of keywords within a text unit (on one and the same index card). Take the following example: Index card A07 contains the sentence “Ein gutes Gleichnis erfrischt den Verstand (MS 105 73 c: 1929)”. What we would do now is marking those keywords in a square matrix that on a syntactical level show “influences” on or “relations” to other keywords. Mostly we would code agent-patient-relations. These keywords contained in a record sheet serve as line and column designations. If in a text unit an “influence” (thus a thematic “relation”) is assumed, then the assumed influence is entered in the line of the influence variable (keyword) and the column of the influenced variable as “+” or as “-,” according to whether the influence is a positive or a negative one. In our example, mentioned above, we would then code<sup>8</sup>:

	A (Gleichnis)	B (gut)	C (erfrischen)	D (Verstand)
A (Gleichnis)				
B (gut)				
C (erfrischen)				
D (Verstand)	←↓+			

Fig. 7: Matrix recording the “influence-on”-relationship between the keywords “Gleichnis” and “Verstand” on index card A07.

This chart shows that within a certain text unit (contained in one index card) the relation between the keywords “Gleichnis” (“simile”, “analogy”) and “Verstand” (“intellect”) does exist and that it is a positive one.

Normally these codings in GABEK would be referred to as causal relations. However, with the *Vermischte Bemerkungen* mere causal coding would not suffice. If we had only marked causal dependencies, it might not be illuminating when it comes to analyzing relations between keywords occurring within *Vermischte Bemerkungen*. So we decided to mark all relations that could be defined as: I “keyword 1”, “keyword 2”. In order to isolate inter-dependencies from the textual data, the electronically captured text units on the index cards must be read again and coded in a new way. The questions we ask to the text in order to find out about the nature of the relations between keywords are: Does the text provide hints that I “keyword 1”, “keyword 2”? GABEK also allows for questions to find out whether the text shows that one keyword A leads to the increase or decrease of another keyword B. Applicable to the coding of inter-dependencies are among others (as mentioned in Oberprantacher/Zelger (2002)):

- “quantitative relations (in the form of ‘the more A, the more B’),
- real relations between keywords (‘A has an effect on/ influences B’), qualitative relations specification (‘if A improved, then the quality from B takes too’),
- statistical generalizations (‘if A, then mostly also B’ applies) and many others.”

Additionally, we could distinguish between influences that have either a favorable or an unfavorable effect on another variable. The coding of inter-

dependencies finally leads to the description of complex relations. Two lists are generated: the “inter-dependency list” and the “list of inter-relations”. Whereas, the inter-dependency list provides information on the amount of effects between keywords, the list of inter-relations shows more about the nature of these inter-dependencies themselves. Although there are other features relevant to a comprehensive data analysis, we will only go into one more important detail for reasons of comprehensibility. The final step important for our investigations is the generating of inter-dependency network graphics, which are based on the coding of inter-relations. Researchers may, for instance, choose any keyword from the keyword list and create a network by expanding it with keywords showing at least two inter-relations with the starting keyword. Let's take the following example, starting with the keyword “Kultur” (“culture”): Previously executed projects have shown that inter-dependency networks representing inter-relations contribute substantially to comprehensive textual analysis.

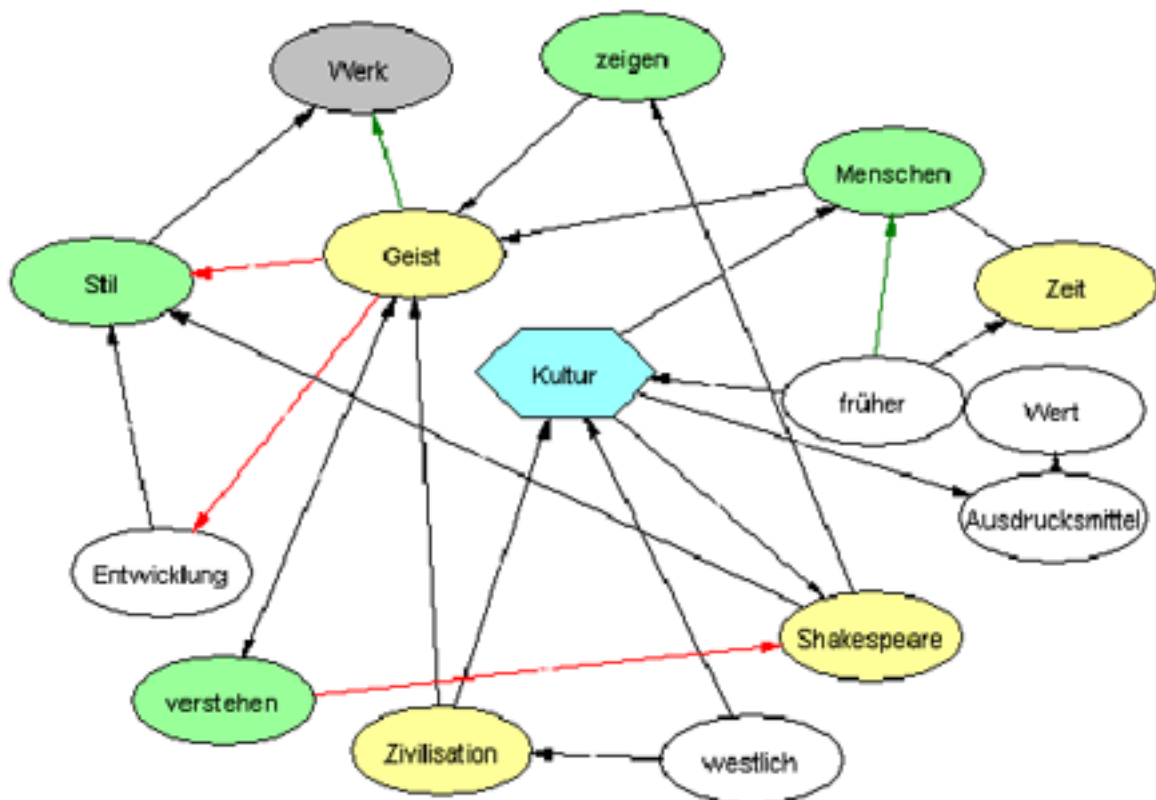


Fig. 8: Inter-dependency network graphic starting with “Kultur” (“culture”)

Now, what could this graphic show? Use of colors in the graphic can illustrate the position of a keyword in the context of other keywords (I “keyword

1”, “keyword 2”). One color could e.g. be used to show that a keyword repeatedly co-occurs with other keywords by showing influence on them. From this graphic researchers could always trace back the original contexts of keywords by going back to the index cards onto which a text unit was copied (cf. Fig. 9 below).



Fig. 9: Inter-dependency network graphic of “Kultur” (“culture”) with index card references

What this and similar graphs can indicate, is the semantic structure of associated keywords (representing conceptual fields) in that it shows the frequency of I “keyword 1”, “keyword 2”; this is signified by the amount of arrows pointing at a certain keyword or pointing away from it. If more arrows target at an item than lead away from it, it would be defined as an “aim”. However, if a keyword shows more influence on other items than it shows influences on itself, we would speak of it as a “measure”. Now, if the number of arrows depicting “influence-on-relations” and arrows illustrating “being-influenced-relations” is equal, the affected keyword is characterized as a so called “intervening variable”. Intervening variables are keywords signifying a linkage between two (or more) other key concepts.

The network above (Fig. 9) shows that keywords like “Menschen”, “zeigen” and “verstehen” are not addressed as “aims” or clear “measures”, but display mediate impact on other variables (mostly not sufficiently recognized), such as “Shakespeare”, “Leidenschaft” or “Geist”. Some other keywords, such as “Geist”, “Zivilisation” or “Zeit”, stand for “measures” and are thus affecting other concepts either positively or negatively. Again a different group of keywords stands for concepts that Wittgenstein in his remarks repeatedly addresses as desirable; they are therefore definable as “aims” (e.g. “Werk”, “staunen” et.al.).

Such an analysis and identification of inter-relations of items (keywords), conceptual fields and topics as well as semantic inter-dependencies and networks – which are achieved through the development of a rule-based network of data (units) – are needed to generate both a more profound knowledge and an understanding of the semantic structures of this (re-)constructed Wittgenstein text. This knowledge expresses itself in the unique character of its organisation and structure and can help to build the basis for further in-depth investigations and analysis concerning specific topics related to *Vermischte Bemerkungen*. In coherence with the core objectives of the analysis of the German text version (*Vermischte Bemerkungen*), an encoding of the 1st and 2nd English edition (*Culture and Value*) might be interesting for the purpose of comparison and exploration in terms of textual semantic similarity and deviation<sup>10</sup>.

Using GABEK/Winrelan we can do more than clarifying terms, showing frequencies or representing inter-dependencies. We can achieve a clear, holistic overview of the textual data. In order to do so we need to put our data into a hierarchically ordered “*gestalten* tree”.

### 3.2 Creating linguistic *gestalten*

This is done by running a cluster analysis on all keywords identified at least twice on at best five to nine index cards.<sup>11</sup> The cluster analysis is a built-in feature (in WinRelan) and helps the researcher to generate virtual piles of index cards sharing again five to nine keywords. However, if there are too many index cards with too many different keywords (cf. Zelger 1996, 11), one would get too many groups, i.e. too many topical threads so that an identification of more and less prominent themes would be impossible. On the other hand, if the index cards share too many keywords, we would get too few piles and it would seem as if all topics were equally prominent; either is problematic. Because when it comes to summarizing the content

represented on these grouped cards according to specific syntactic and semantic rules, we would either get a too comprehensive summary or only a superficial one. The summaries (*gestalten*) are semantic implications of the grouped cards and build the basis for further grouping and summarizing on the next higher level. What we get are so called *hyper-gestalten*. This process is repeated until we have no more groups to summarize. The final product is a *gestalten-tree*.

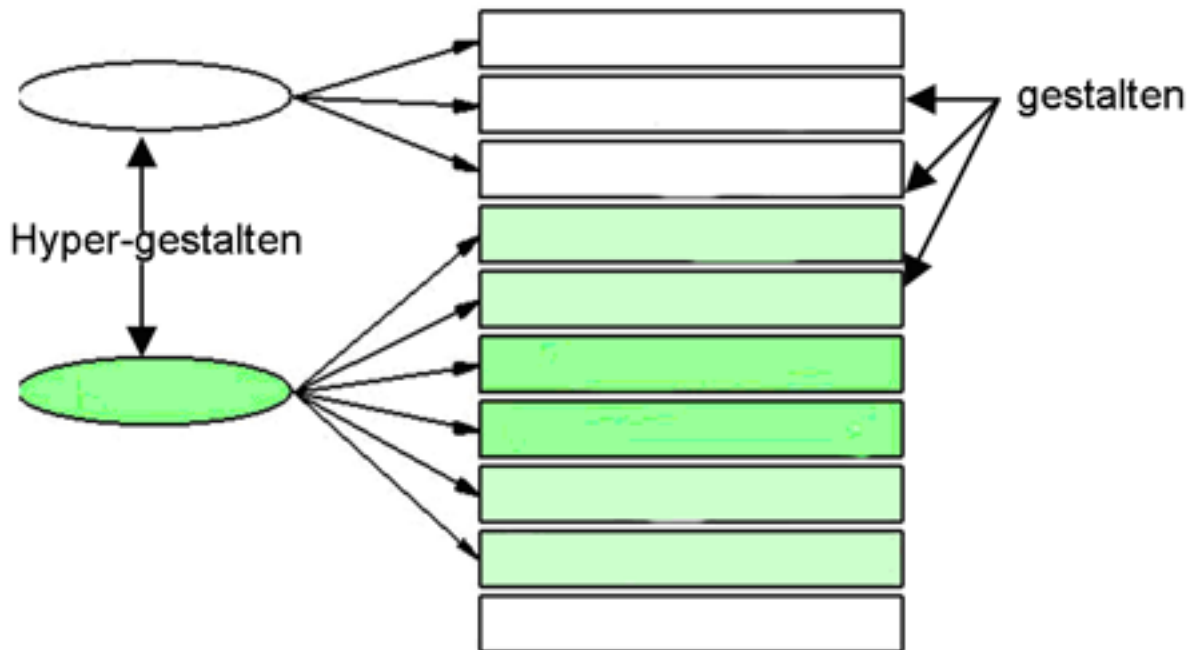


Fig.10: *gestalten-tree*

Any careless or deviant coding at an earlier stage affects the quality of the later analysis. Thus, the decision on how many sentences are to be coded on one index card is a crucial one.

So finally, these findings can provide the basis for further investigations concerning such questions as the following:

1. Can thematic/semantic structures that have been identified in the course of this analysis process be useful for further investigations? (And if so in what respect?)
2. What can be captured (is it useful?) and what cannot be captured (blind spots)? Is this method of analyzing texts in order to “prepare” them for further research balanced in terms of the amount of time needed and the actual output?

3. Is there the chance to discover semantic inter-relations topical patterns that would otherwise be overlooked or neglected?
4. Could one find patterns when e.g. analyzing text under a certain repeat? E.g., if one wanted to analyze all occurrences of a certain keyword (representing a topic) in its various contexts within Wittgenstein's writings, would one get useful structures etc. which could provide access to an understanding of the topic from a new perspective? (Would that be helpful for understanding a certain topic in Wittgenstein?)

#### 4. Conclusion

Since the GABEK-analysis of *Vermischte Bemerkungen* has not yet been finished, the questions just mentioned cannot be fully answered at the present state. What can though be said, is that first tentative assumptions regarding thematic patterns or semantic structures can be made. These could become relevant if we wanted to investigate whether there are topic patterns that would justify to put those remarks into one volume. All we have tried so far is looking at *Vermischte Bemerkungen* from a different angle by applying GABEK/ WinRelan. Within this context, we, of course, have taken account of the fact that all remarks were originally written by Wittgenstein himself, but also of the fact that these remarks do not at all constitute a complete piece of Wittgensteinian text. It is neither our task nor our aim to solve the problem of "what is a Wittgenstein text". So, we try to (re-)construct it by first chunking it, analyzing its components and rebuilding it anew. At best that will show us that there is a point to those remarks and that they are maybe not that incoherent as they may seem at first sight.

#### References

- Miller, George A. (1956): "The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information" in *Psychological Review*, 63. London: Harvard University Press, 81-97.
- Oberprantacher, Andreas/ Zelger, Josef (2002): *Processing of Verbal Data and Knowledge Representation by GABEK® -WinRelan®* in FQS – Forum Qualitative Social Research, Vol.3, No.2. [Source: <http://>



[www.qualitative-research.net/fqs-texte/2-02/2-02zelgeroberprantacher-e.htm](http://www.qualitative-research.net/fqs-texte/2-02/2-02zelgeroberprantacher-e.htm)].

Wittgenstein, Ludwig (1998): *Culture and Value, A Selection from the Posthumous Remains*, edited by Georg Henrik von Wright in Collaboration with Heikki Nyman, revised edition of the text by Alois Pichler, translated by Peter Winch. Oxford: Blackwell.

Zelger, Josef (1996): *Linguistic Knowledge Processing by GABEK. The Selection of Relevant Information from Unordered Verbal Data*. Paper at the International Conference on Conceptual Knowledge Processing. TH Darmstadt, Faculty of Mathematics, February. Preprint Nr. 42. Universität Innsbruck: Institut für Philosophie.

## Notes

1. See the project Web sites at [http://www.uibk.ac.at/brenner-archiv/projekte/wittg\\_vermishtes/index.html](http://www.uibk.ac.at/brenner-archiv/projekte/wittg_vermishtes/index.html) and [http://wab.aksis.uib.no/wab\\_fwf-cv.page](http://wab.aksis.uib.no/wab_fwf-cv.page).
2. *Wittgenstein's Nachlass: The Bergen Electronic Edition*. Edited by the Wittgenstein Archives at the University of Bergen. Oxford: Oxford University Press 2000. *Ludwig Wittgenstein: Gesamtbriefwechsel/Complete Correspondence: The Innsbruck Electronic Edition*. Edited by Monika Seekircher, Anton Unterkircher and Brian McGuinness. Charlottesville: IntelLex Corporation 2005.
3. Hereafter referred to as “sentence”.
4. This GABEK rule (cf. Buber/ Zelger (2000), p.117) is based on the findings of George A. Miller (1956) according to which a person can remember 7 (up to 9) terms.
5. Here, the term “sentence” does not refer to a linguistic sentence but to a meaning unit.
6. The relevancy of terms and topics is determined by their frequency (quantity) as well as by their interrelations (quality).
7. I.e. a term is connotated positively as well as negatively.
8. Compare Oberprantacher/ Zelger (2002).
9.  $I_{xy}$ : x shows influence on y.
10. However, if such an analysis were undertaken, researchers would have to pay special attention to the difficulties arising with the translation of texts, especially philosophical texts.
11. A cluster analysis is used to classify objects (in this case “index cards”) and groups them according to the (number of) keywords they share. Clustering is a common technique in data mining using mathematical algorithms for distance measures by which the proximity (similarity) of the objects in question is calculated.

