Graphentheoretische Modelle zu Wittgensteins *Vorlesungen und Gespräche über Ästhetik, Psychoanalyse und religiösen Glauben*

von Joseph Wang (Innsbruck)

Was für die naturwissenschaftlichen Disziplinen bereits der Alltag ist, ist für Geisteswissenschaftlerinnen noch ein relatives Neuland; gemeint ist hier die elektronische Datenverarbeitung mittels Computer. Während die Meteorologinnen täglich unzählige Messungen als Datensätze aufnehmen und diese mit geeigneten Algorithmen bearbeiten, sodass sie das Wetter relativ verlässlich voraussagen können, müssen die Geisteswissenschaftlerinnen die Texte selbst durchlesen und interpretieren. Dabei helfen ihnen zwar Texte anderer Forscherinnen, aber letztlich müssen sie sich selbst den Inhalt aneignen, wenn sie die Texte für ihre Arbeit fruchtbar machen wollen.

Kreative Köpfe fragen sich – die Naturwissenschaften vor Augen –, ob wir die Texte nicht als „geisteswissenschaftliche Daten" ansehen können. Falls dieser Vergleich zulässig ist, dann fragen sie weiter, wie Computer bzw. Softwares eingesetzt werden könnten, um die Geisteswissenschaftlerinnen bei ihren Tätigkeiten zu unterstützen. Gibt es Algorithmen, die man zur Verarbeitung der Texte heranziehen könnte? Und könnte das Ergebnis dieser Analyse dazu verwendet werden, um den Inhalt der Texte besser zu erschließen?

Im Rahmen der beiden Projekte „Vermischte Bemerkungen Neu" und „Wittgenstein's Vorlesungen zur Ästhetik im Kontext" (FWF P19022 und FWF P21038) sind uns diese Fragen begegnet. Während der Bearbeitung beider Texte des Philosophen Ludwig Wittgenstein, der *Vermischten Bemerkungen*[1] (im Folgenden *VB* genannt) und der *Vorlesungen und Gespräche2* (*LC*) haben wir u.a. Softwares entwickelt und eingesetzt, die diese über das bloße „Digitalisieren" hinaus verarbeiten. In der Hoffnung, dass sich durch eine elektronische Verarbeitung beider Texte neue Erkenntnisse zeigen, haben wir zwei „Versuchsreihen" durchgeführt: *VB* wurde mit der GABEK-Methode[3] analysiert, und *LC* wurde mit Hilfe graphentheoretischer Überlegungen untersucht. Da die GABEK-Analyse im Projekt *Vermischte Bemerkungen Neu* bereits vorgestellt wurde[4], geht es in diesem Artikel vorwiegend um die Analyse von *LC*.

An dieser Stelle soll noch angemerkt werden, dass es andere Ansätze der elektronischen Inhaltserschließung gibt. Da die meisten Texte Wittgensteins – hier sind v.a. sein philosophischer Nachlass in der Form der *Bergen Eletronic Edition*[5] und sein Briefwechsel in der Form des *Gesamtbriefwechsels Wittgensteins*[6] zu nennen – bereits digitalisiert vorliegen, liegt es nahe, dass Texte Wittgensteins auch dort bearbeitet werden. Besonders hervorheben möchte ich das EU-Projekt DISCOVERY[7], da sich unser Projekt-Partner aus dem Wittgenstein Archives in Bergen an diesem Projekt beteiligt und wir dank des Wissenstransfers viel vom Know-how des DISCOVERY-Projektes profitiert haben. Im Rahmen dieses Projektes werden Schlagworte für eine Reihe von Bemerkungen aus Wittgensteins Nachlass ausfindig gemacht. Diese stehen dann über das

Web-Portal *Philosource*[8] als sogenannte *web ontology* den Forscherinnen zur Verfügung. Im Unterschied zur Vorgehensweise von DISCOVERY, bei der Wittgenstein-Expertinnen als Bearbeitende fungieren und der Computer als Datenverwalter dient, streben die beiden FWF-Projekte im Brenner-Archiv die (halb-)automatisierte Inhaltserschließung durch den Computer an. Hierbei soll die inhaltliche Analyse im Wesentlichen von einem Computerprogramm übernommen werden.

Vermischte Bemerkungen: Ein Sammelsurium will geordnet werden

Der philosophische Nachlass Wittgensteins umfasst ca. 20.000 Seiten.[9] Darin enthalten sind sowohl verschiedene Vorstufen seiner *Philosophischen Untersuchungen* als auch einzelne Diktate, Notizbücher, Tagebücher und Typoskripte. Vor allem in den Tagebüchern und in den Notizbüchern[10] hat Wittgenstein oft einzelne, eher persönliche Bemerkungen zwischen seinen philosophischen Gedanken niedergeschrieben. Posthum wurden viele Notizbücher unter Titeln wie *Philosophische Grammatik*, *Bemerkungen zu Grundlage der Mathematik*, oder *Über Gewissheit* veröffentlicht, dabei haben die Herausgeberinnen und Herausgeber darauf geachtet, dass die persönlichen Bemerkungen und jene Bemerkungen, die mit seinen systematisch-philosophischen Überlegungen weniger zu tun haben, nicht publiziert werden.

Erst 1977, also 26 Jahre nach Wittgensteins Tod, gab Georg Henrik von Wright eine Sammlung von Wittgensteins Bemerkungen, die in den früheren Ausgaben weggelassen worden waren, unter dem Titel *Vermischte Bemerkungen* heraus. Von Wright sammelte einzelne Textstellen aus unterschiedlichen Notizbüchern, ließ dabei die rein persönlichen Bemerkungen weg und fasste die anderen zusammen.[11] So finden wir in *VB* sowohl Überlegungen beispielsweise zu Musik und Literatur wie auch zur Religion. Da sich diese Bemerkungen sowohl thematisch als auch zeitlich weit erstrecken, könnte man leicht den Eindruck gewinnen, dass diese zusammenhang- und konzeptlos ausgewählt worden sind. Eine der Aufgaben im Projekt *Vermischte Bemerkungen Neu* bestand darin, den „roten Faden" dieser Texte – sofern ein solcher existiert – zu eruieren. Das vorläufige Ergebnis dieser Untersuchung kann in *(Re)-Constructing the Semantic Architecture*[12] nachgelesen werden.

Vorlesungen und Gespräche: Ausgangspunkt der Untersuchung

Im Jahre 1966 gibt der irische Jesuit Cyril Barrett eine Sammlung von Mitschriften dreier Schüler Wittgensteins unter dem Titel *Lectures and Conversations on Aesthetics, Psychology and Religious Belief* heraus. Dieser Text setzt sich zusammen aus den Mitschriften der Wittgenstein-Schüler James Taylor, Rush Rhees und Yorick Smithies, die den Inhalt von drei Vorlesungsreihen bzw. Gesprächen niedergeschrieben haben. Die vier Vorlesungen zur Ästhetik fanden in einem kleinen Kreis im Jahre 1938 statt, etwa zur selben Zeit wie auch die Vorlesung zur Psychoanalyse. Der Glaube war Gegenstand einer Vorlesungsreihe, die irgendwann zwischen 1940 und 1942 veranstaltet wurde.[13]

Während die Bemerkungen in *VB* heterogen sind und das „zentrale Thema", sofern es dieses überhaupt gibt, schwer erkennbar ist, ist dies bei *LC* nicht der Fall. Die

Überschriften der einzelnen Sektionen geben den Inhalt gut wieder. Zwar kommt Barrett nicht umhin, auf Unterschiede zwischen den einzelnen Manuskripten hinzuweisen – so gibt es Sprünge und Einschübe im Text[14] –, aber verglichen mit *VB* ist der Text in *LC* zusammenhängend. Eine Inhaltsanalyse mittels GABEK, wie sie bei VB vorgenommen wurde, scheint daher bei *LC* nicht angezeigt zu sein.

Datenerhebung: Zuweisung von *Keywords* an *Sinneinheiten*

Die Grundidee von GABEK, dass es ‚Sinneinheiten' gibt, in denen ‚Keywords' vorkommen, ist einleuchtend und scheint auch bei der Analyse von *LC* sinnvoll zu sein. Was ist aber eine Sinneinheit, und was ein Keyword?

Diese Methode dient dazu, einen Text zu strukturieren, und die Strukturierung ist ein notwendiger Bestandteil der Datenverarbeitung. Gängig ist die Strukturierung eines Textes in Sätze und Absätze. Da wir aber letztlich die Beziehungen zwischen den einzelnen Keywords analysieren wollen, die in einem Gedanken miteinander verbunden sind, ist die vorhandene Strukturierung des Textes in Sätzen bzw. Absätzen nur bedingt anwendbar. Denn es bedarf oft mehrerer Sätze, um einen Gedanken zum Ausdruck zu bringen, in einem solchen Fall stehen die Wörter, die durch einen Gedanken verbunden werden, in unterschiedlichen Sätzen. In einem Absatz können wiederum mehrere Gedanken wiedergegeben werden. Und so teilen wir den Text von *LC* – ganz im Sinne von GABEK – in *Sinneinheiten* auf, wobei in einer Sinneinheit ein Argument oder ein Gedanke vollständig wiedergegeben wird.[15]

Für die kodierten Wörter in einer Sinneinheit verwenden wir den Ausdruck *Keyword*. Wir fassen nämlich die Flexionen eines Wortes (beispielsweise ‚nennen' ‚nennt', ‚nannten' u.s.w.) zusammen und trennen die äquivoken Wörter voneinander. Beispielsweise gehören ‚nennen', ‚nennt' und ‚nannte' zum Keyword „nennen", aber Wörter, die vom Verb „denken" stammen, werden je nach ihrer Bedeutung in „denken" oder „denken (sich etwas vorstellen)" zugeordnet. Zu dem werden jene Wörter (wie „in", „oder" und alle Flexionen des Verbs „sein") von der Analyse ausgenommen, die nicht sinntragend sind. Da diese Keywords aus den *Wörtern* im Text stammen und keine *Schlagworte* des Textes darstellen, scheint der Ausdruck ‚Keyword' passend zu sein, um Verwirrungen zu vermeiden.

Ein Beispiel einer solchen Sinneinheit aus *LC* mit den dazugehörigen Keywords finden Sie hier:

> *Sinneinheit ‚Ex':* „Das Paradigma der Wissenschaften ist die Mechanik. Wenn die Menschen sich eine Psychologie vorstellen, ist ihr Ideal das einer Mechanik der Seele. Wenn wir betrachten, was dem tatsächlich entspricht, sehen wir, daß es physikalische Experimente und daß es psychologische Experimente gibt. Es gibt Gesetze der Physik und es gibt – wenn man höflich sein will – Gesetze der Psychologie."[16]

Keywords der Sinneinheit ‚Ex‘: Wissenschaft, psychologisch, betrachten, Gesetz, entsprechen, physikalisch, Experiment, Ideal, Psychologie, vorstellen, Paradigma, Physik, Menschen, Seele, Mechanik.

Und so wird der Text des *LC* in 316 Sinneinheiten mit 1759 Keywords getrennt. Ausgehend von diesen Daten wollen wir die Analyse durchführen. Bevor wir diese Daten analysieren können, müssen sie aber in eine Form gebracht werden, dass der Computer „weiß“, wie sie zu verarbeiten sind. Hier stellt sich die Frage nach der Verwaltung dieser Daten. Und es zeigt sich, dass sich die Datenverwaltung am besten mit der Hilfe der *semantic web technology* bewerkstelligen lässt.

Semantic Web und die Erstellung einer *ontology*

Im Rahmen des DISCOVERY-Projektes werden 5000 Seiten des Nachlasses Wittgensteins frei zugänglich im Sinne des *open access* gemacht. Hierbei liegen diese Seiten den Forscherinnen als Faksimiles, als Webseiten und als XML-Datei vor[17], zudem werden jene Schlagworte, die den einzelnen Bemerkungen Wittgensteins von Expertinnen der Universität Bergen gegeben wurden, mit Hilfe der *semantic web technology* den Forscherinnen zur Verfügung gestellt. Hierbei verwendet das Wittgenstein Archives in Bergen die Spezifikationen *Web Ontology Language* (OWL)[18] und *Resource Description Framework* (RDF)[19], um die Daten zu speichern und die Schlagworte zu veröffentlichen. Wir haben diesen Modus auch für unser Projekt übernommen, was den Austausch von Daten wesentlich erleichtert.

Ohne in die technischen Details zu gehen, kann man sich die Datenspeicherung so vorstellen: RDF wurde entwickelt, um einzelne Objekte, sogenannte Ressourcen, zu beschreiben. Dabei ist es wichtig, dass jede Ressource einen eindeutigen Namen, eine ID, bekommt. Wir betrachten nun die einzelnen Sinneinheiten wie auch die Keywords als Ressourcen mit eigener ID und nehmen diese in die Datenbank auf. Dann werden die zwei Objekt-Klassen (Sinneinheit und Keyword) in der Datenbank registriert, und es wird die Zuordnung der einzelnen Ressourcen zu den Klassen festgelegt. Sobald wir die Ressourcen, die Klassen und die Klassen-Zuordnungen in die Datenbank eingespeist haben, ‚weiß‘ der Computer, dass „Wissenschaft“ ein Keyword und „Ex“ eine Sinneinheit ist. Fortan können diese Ressourcen mit weiteren „Sätzen“ beschrieben werden. Wichtig dabei ist, dass diese „Sätze“ eine dreistellige Struktur aufweisen müssen: Ein ‚Subjekt‘ steht mit einem ‚Objekt‘ in einer bestimmten Beziehung, wobei die Beziehung ‚Prädikat‘ genannt wird. Wir legen nun das Prädikat ‚ist ein Schlagwort in‘ fest und bestimmen, dass diese Beziehung nur zwischen einer Ressource der Klasse ‚Keyword‘ und einer Ressource der Klasse ‚Sinneinheit‘ vorkommen darf, wobei das Keyword das Subjekt der Beziehung und die Sinneinheit das Objekt der Beziehung darstellt.

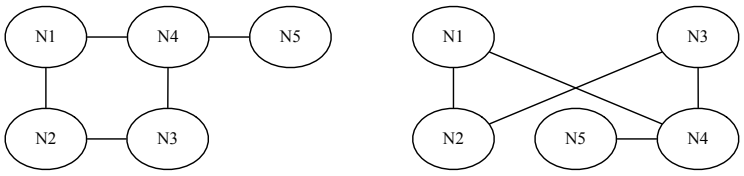
Insgesamt haben wir 4626 solche Sätze in die Datenbank eingegeben, mit denen der Computer nun „rechnen“ kann. So könnte der Computer „Fehler“ entdecken, die beispielsweise dadurch entstehen, dass in einem Satz nicht eine Ressource der Klasse ‚Keyword‘, sondern eine Ressource der Klasse ‚Sinneinheit‘ das Subjekt stellt. Diese Art

der Datensammlung von „Ressourcen“, „Klassen“, „Beziehungen“ und „Sätzen“ wird von den Fachleuten der *semantic web technology* als *ontology* bezeichnet. Wir haben also eine *ontology* erstellt, die prinzipiell in Form von OWL-Dateien auch anderen Leuten zur Verfügung steht.[20]

Die graphentheoretische Modellierung der Daten

Während das Problem der Datenverwaltung mit *semantic web technology* gelöst wird, stellt sich nun die Frage nach der Analyse dieser Daten. Wie können wir sie für uns nutzbar machen? Die naheliegende Antwort ist die Modellierung dieser Daten als Graphen, die dann mit der Hilfe graphentheoretischer Werkzeuge analysiert werden.[21]

Die moderne Graphentheorie, die auf Leonhard Euler zurückgeht[22], behandelt die mathematischen Eigenschaften von Graphen. Ein Graph *G* in diesem Sinne ist die Vereinigungsmenge aus zwei Untermengen *V* und *E* (= Verbindungslinien). Eine davon (*V*) enthält die sogenannten Knoten, und die andere (*E*) enthält die Kanten. Dabei müssen alle Knoten, die mit den Kanten in *E* verbunden werden, in der Menge *V* enthalten sein. Zusätzlich können Knoten und Kanten auch andere Eigenschaften haben, etwa einen *Namen*, eine *Länge* oder auch eine *Richtung*, die für die Analyse von Bedeutung sein können. Die Positionen der Knotenpunkte in einem visualisierten Graph spielen *keine* Rolle bei der Analyse. Beide Graphen sind demnach identisch:



Bereits mit dem Computerprogramm *Protégé*[23] ist es z.B. möglich, *ontologies* als Graphen zu visualisieren. *Protégé* wurde von der Universität Stanford als *ontology*-Editor entwickelt, mit dem man beispielsweise für eine gegebene *ontology* die Konsistenz-Analysen durchführen und Folgerungsbehauptungen aufstellen kann. Ein Ausschnitt unserer *ontology* sieht in diesem Editor so aus:

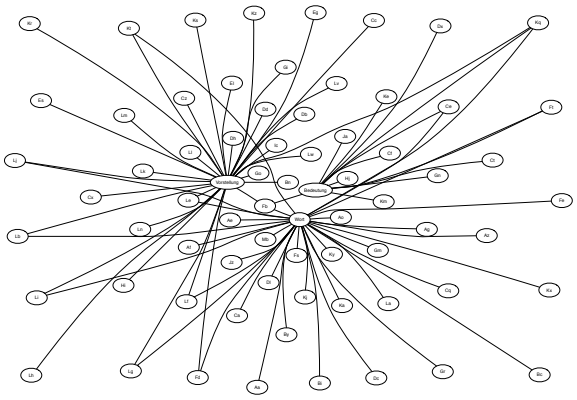
Hier werden zwei Keywords („Wörter“, „gut“) und die Sinneinheiten, in denen zumindest eines der beiden Keywords auftritt, als Knoten und die Beziehung „ist ein Schlagwort in“ als Kante dargestellt. Zusätzlich sieht man noch die Klasse Keyword (hier als Schlagwort bezeichnet) als Knoten und die Beziehung „ist ein Mitglied der Klasse“ als Kante.

Mit *Protégé* ist es zwar möglich, Graphen dieser Art darzustellen, aber das Programm ist nicht darauf angelegt, die Beziehungen tiefer gehend zu analysieren. In der *ontology* kann man eine Beziehung namens „kommt zusammen vor mit“ definieren, die zwischen zwei Keywords besteht, aber mit *Protégé* ist es relativ schwierig, diese Beziehungen, die ja bereits in der *ontology* festgehalten werden, automatisch zu berechnen.[24] Schreiben wir diese Beziehungen explizit in die *ontology* hinein, dann wäre *Protégé* zwar in der Lage, Graphen zu zeichnen, indem Keywords als Knoten und die Beziehung „kommt zusammen vor mit“ als Kanten dargestellt werden, aber eine graphentheoretische Diskussion des Graphen wäre immer noch nicht möglich.

Eine andere Möglichkeit, unsere *ontology* zu analysieren, bieten objektorientierte Programmiersprachen (wie Java). Diese warten mit der Funktion auf, spezifische Methoden für Objekte einer Klasse zu definieren. Programmiersprachen dieser Art sind so angelegt, dass man mit ihnen gut mit Klassen und Objekten umgehen kann. So lässt sich die Beziehung „kommt zusammen vor“ in solchen Programmiersprachen leicht berechnen. Wir haben uns daher dafür entschieden, ein Programm in Java zu erstellen, das aus der *OWL ontology* selbstständig Graphen verschiedener Arten erstellt. Es gibt in Java zudem verschiedene *open source* Bibliotheken, die auf die Datenverarbeitung mit Graphen spezialisiert sind.[25]

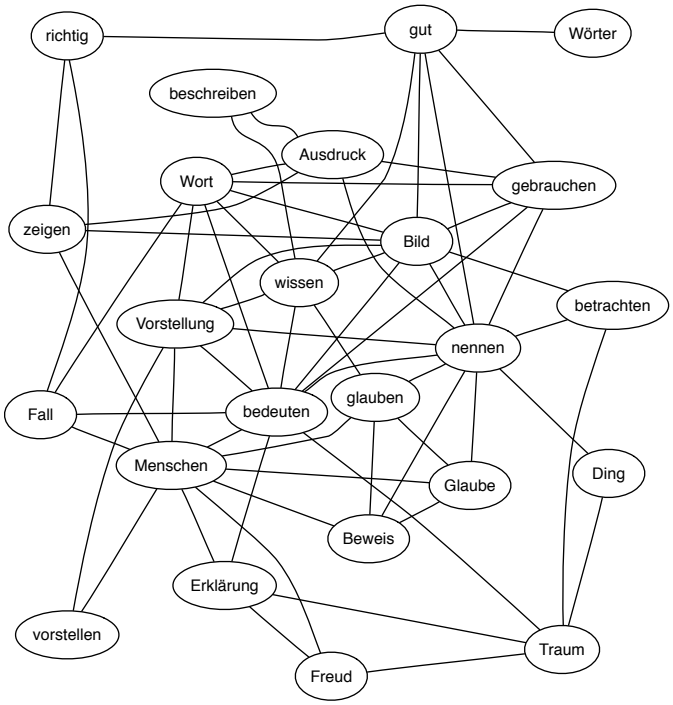
Ausgehend von der OWL-Ontologie können wir nun Graphen auf drei Arten generieren:

1. Ein Graph, in dem Sinneinheiten und Keywords als Knoten und die Beziehung „ist ein Keyword in“ als Kanten dargestellt werden, wie bei *Protégé*. Graphen dieser Art können so ausschauen:



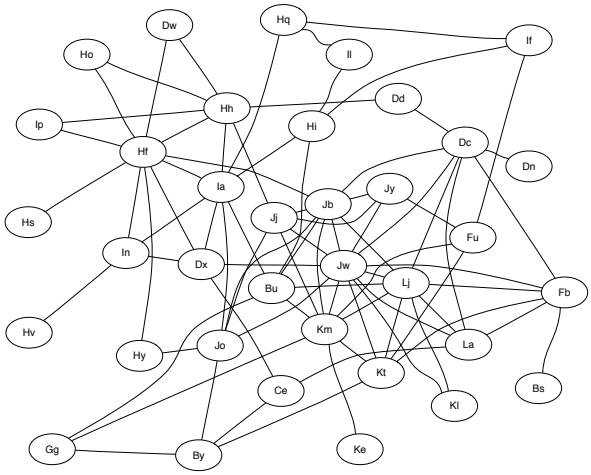
Der Graph GA1 zeigt drei Keywords „Wort“, „Vorstellung“ und „Bedeutung“ und die Sinneinheiten, in denen diese Wörter vorkommen, als Knoten. Die Kanten bilden die Beziehung „ist ein Keyword in“ ab. Im Folgenden wird ein Graph dieser Art, in dem alle Keywords und alle Sinneinheiten vorkommen, mit GA bezeichnet.[26]

2. Ein Graph, in dem Keywords als Knoten und die Beziehung „kommt zusammen vor mit“ als Kanten dargestellt werden, wobei wir hier die Keywords nach der Häufigkeit ihres Vorkommens filtern können. So können wir beispielsweise nur jene Keywords darstellen, die mindestens in zwei Sinneinheiten vorkommen, und die anderen Knoten weglassen.



Der Graph GB1 stellt alle Keywords, die mindestens in 20 Sinneinheiten auftreten, als Knoten dar. Die Beziehung „kommt zusammen vor mit“ bildet dann eine Kante, wenn die beiden miteinander verbundenen Keywords in mindestens fünf Sinneinheiten zusammen vorkommen. In unserer Analyse gehen wir aber von einem Keyword-Graphen aus, der alle Keywords und alle Kanten beinhaltet. Er besteht aus 1.759 Knoten und 33.953 Kanten. Im Folgenden werde ich mit GB auf diesen „ungefilterten“ Keyword-Graphen referieren.

3. Ein Graph, in dem die Sinneinheit als Knoten und die Beziehung „teilt sich ein Keyword mit“ als Kante dargestellt wird. Auch hier gibt es verschiedene Möglichkeiten, die Darstellung des Graphen zu beeinflussen.
Der Graph GC1 stellt alle Sinneinheiten als Knoten dar, denen mindestens 24 Keywords zugewiesen werden. Die Beziehung „teilt sich Keywords mit“ wird hier als Kante



dargestellt, wenn die zwei verbundenen Sinneinheiten mindestens drei gemeinsame Keywords haben. Ein Sinneinheit-Graph, in dem alle Sinneinheiten als Knoten dargestellt werden und der alle Kanten beinhaltet, besteht bei uns aus 316 Knoten und 14.363 Kanten. Die Bezeichnung GC referiert auf diesen Sinneinheit-Graphen.

Wir können bei GB und GC noch die „Stärke“ der Kanten angeben. Eine Kante von der Stärke 1 im GC möge heißen, dass die beiden verbundenen Sinneinheiten sich ein Keyword teilen, und eine Kantenstärke von 7 im GB soll bedeuten, dass es sieben Sinneinheiten gibt, in denen die beiden verbundenen Keywords zusammen vorkommen.

Die Durchführung der Analysen

Was wir vor uns haben, sind graphentheoretische Modelle unserer *ontology*. Die Sinneinheiten und die Keywords, die wir miteinander in Beziehung gesetzt haben, werden als ungerichtete Graphen[27] in unterschiedlichen Formen GA, GB und GC dargestellt. Diese Modelle können nun mit graphentheoretischen Mitteln beschrieben und analysiert werden. So kann man beispielsweise angeben, ob dieses Graph zusammenhängend ist. Andere graphentheoretische Analysen betreffen beispielsweise die Suche nach den kürzesten Pfaden zwischen zwei Knoten.[28] In diesem Beitrag werden wir v.a. auf die Zentralitätsanalysen von Graphen eingehen. Zuvor werden einige statistische Angaben präsentiert.

Die statistischen Daten

Ausgehend von GA können wir insgesamt 1759 Keywords und 316 Sinneinheiten ausfindig machen. Die meisten Keywords (insgesamt sind es 1055) kommen nur in einer Sinneinheit vor, hingegen sind die folgenden Keywords häufiger anzutreffen:

Keyword	Häufigkeit
Bild	41
Menschen	40
bedeuten	40
nennen	39
Wort	36
Fall	35
Ausdruck	34
Traum	34
Vorstellung	34
Freud	31
gut	30
wissen	30

Neben der Häufigkeit sind noch jene Keyword-Paare von besonderem Interesse, die öfter zusammen vorkommen. Da die Keyword-Paare als Kanten und die Stärke dieser Kanten die Häufigkeit ihres gemeinsamen Auftretens bezeichnen, können wir die Kanten in GB nach ihrer Stärke filtern. Die folgende Tabelle gibt jene Keyword-Paare wieder, die in mehr als acht Sinneinheiten anzutreffen sind:

Keyword 1	Keyword 2	Stärke
Ausdruck	Gesicht	9
Bild	nennen	10
Erklärung	Freud	9
Erklärung	Menschen	9
Freud	Traum	14
gebrauchen	Wort	9
glauben	Jüngstes Gericht	9
gut	schön	9
gut	Wörter	11
Interpretation	Traum	9
Vorstellung	Wort	10

Man kann behaupten, dass jene Keywords zu demselben Themenfeld gehören, die oft miteinander verbunden sind. Und es wird kaum überraschen, dass „Freud“ und „Traum“ oft miteinander erwähnt werden. Eine Auflistung der Keywords-Paare kann also das sichtbar machen, was wir durch das Lesen des Textes bereits erahnen, aber vielleicht noch nicht explizit sagen können.

Und wir können indirekt[29] aus statistischen Daten schließen, welche Rolle ein bestimmter Begriff in einem Text spielt und möglicherweise auch, wie stark die einzelnen Textpassagen miteinander verbunden sind. Ist es nun möglich, ein zentrales Thema bzw. die zentralen Themen eines Textes durch diese Art von Analysen ausfindig zu machen? Da wir bereits ein graphentheoretisches Modell unseres Textes haben, ist es naheliegend, die sogenannte *Zentralitätsanalyse* auf unseren Graphen anzuwenden und zu sehen, ob diese Analyse interessante Ergebnisse zu der Frage nach zentralen Themen liefern kann. Wenn angenommen wird, dass die Graphen GB und GC adäquate graphentheoretische Modelle für *LC* darstellen, scheint es naheliegend zu sein, dass jene Keywords, welche in GB zentral sind, auch das Hauptanliegen des Textes widerspiegeln. Und jene Sinneinheiten, die in GC als besonders zentral gelten, sollten wohl auch eine größere Rolle im Text spielen. Eine Zentralitätsanalyse der Graphen kann also Auskunft über die Wichtigkeit von Keywords und Sinneinheiten geben.

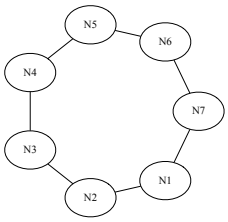
Besonders in der Soziologie findet die Zentralitätsanalyse Anwendung. Soziologinnen bedienen sich gern der Graphen, um die einzelnen Akteure und ihre Beziehungen untereinander zu modellieren. Im Zuge dessen wurden auch unterschiedliche Arten entworfen, um z.B. die zentralen Personen in einem *Sozialnetz* ausfindig machen zu können.[30] Just diese Analyseart wollen wir im nächsten Kapitel auf unseren Keyword-Graph anwenden.

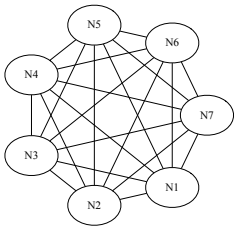
Zentralitätsbegriffe in der Graphentheorie

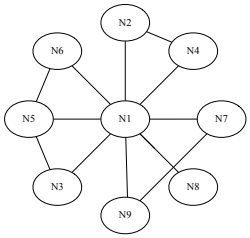
Liegt ein Graph vor, kann für jeden Knoten seine Zentralität rechnerisch bestimmt werden. In diesem Sinne ist die Zentralität eine Zahl, welche einem Knoten zukommt.[31] Um die unterschiedlichen Zentralitätsanalysen leichter verständlich machen zu können, will ich hier zunächst einige einfache Graphen zeigen, bei denen wir intuitiv den zentralen Knoten erkennen können; in einem weiteren Schritt werde ich dann drei Methoden der Berechnung erläutern. Betrachten wir zunächst die Graphen G1 bis G5.

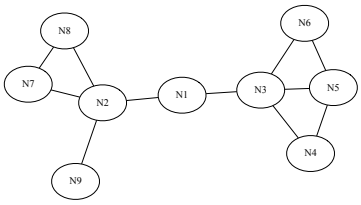
Die meisten Menschen würden sagen, dass alle Knoten in G1 und G2 „gleich zentral“ sind, denn bei diesen beiden Graphen ist es unmöglich, einen Knoten als besonders zentral von den anderen abzuheben. Hingegen können wir leicht den Knoten N1 als den zentralen Knoten für die Graphen G3 und G4 ausmachen: Bei G3 hat N1 direkte Verbindungen zu allen anderen Knoten, und bei G4 liegt N1 in der Mitte. Und zuletzt würden wir die Knoten N1, N2 und N3 als die zentralsten in dem Graphen G5 bestimmen, wobei es unmöglich ist, *den* zentralsten Knoten unter den dreien ausfindig zu machen.

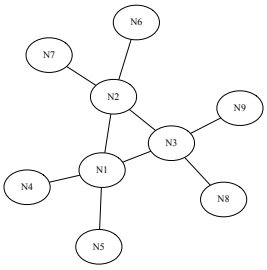
Diese einfachen Graphen zeigen uns, dass ein Graph keinen, einen oder mehrere zentrale Knoten haben kann. Bei komplizierteren Graphen können wir den zentralen


N5
N6
N4
N7
N3
N1
N2


N5
N6
N4
N7
N3
N1
N2


N2
N6
N4
N5
N1
N7
N3
N8
N9


N8
N6
N7
N2
N1
N3
N5
N4
N9


N6
N7
N2
N9
N3
N1
N4
N8
N5

Knoten meist nicht mehr intuitiv fassen. Es sind deswegen mehrere Methoden entwickelt worden, mit deren Hilfe ein Rechner diese Aufgabe übernehmen kann. Die Vielfalt der Analyse-Methoden fängt die unterschiedlichen Intuitionen ein, was einen Knoten zu dem zentralsten (bzw. wichtigsten) Knoten eines Graphen macht. Uns liegen so auch mehrere Zentralitätsbegriffe vor, die voneinander unterschieden werden sollen. In diesem Beitrag möchte ich drei unterschiedliche Zentralitätsanalysen vorstellen.[32] Es wird sich zeigen, dass bei unterschiedlicher Berechnungsweise auch unterschiedliche Knoten als zentral bestimmt werden.[33]

Degree-Zentralität

Eine relativ einfache Art, die Zentralität eines Knotens auszurechnen, ist: die Anzahl der direkten Verbindungen zu anderen Knoten zu zählen. Diese Art der Zentralität nennt man Degree-Zentralität. Das Degree-Maß eines Knotens ist 1, wenn der Knoten mit genau einem anderen Knoten verbunden ist; ein Knoten weist das *Degree* von 2 auf, wenn es zwei unterschiedliche Knoten gibt, die mit ihm verbunden sind. Im Sinne der Degree-Zentralität ist ein Knoten genau dann zentral, wenn er mit mehr Knoten durch direkte Kanten verbunden ist als andere. Der Prototyp eines Graphen mit einem derart zentralen Knoten ist sternförmig, so wie G3 in unseren Beispielen. Während unserer Intuition nach N1 der zentrale Knoten ist, weist N1 auch die meisten direkten Verbindungen auf; hier stimmt unsere Intuition mit der Degree-Zentralität überein, wir haben also gezeigt, dass es Fälle gibt, in denen die Degree-Zentralität sinnvoll angewendet werden kann.

Führen wir diese Bestimmung des Grades an jedem Knoten im GB durch, so sieht das Ergebnis aufgelistet absteigend nach dem Degree-Zentralitätsmaß so aus:

Keyword	Degree-Zentralität
nennen	452
Bild	451
Menschen	446
Fall	429
Traum	423
bedeuten	416
Wort	364
Erklärung	363
Freud	350
Vorstellung	347
Ausdruck	342
zeigen	339
betrachten	335
gut	329

wissen	322
beschreiben	303

Wir sehen, dass die degree-zentralsten Knoten „nennen“, „Bild“, „Menschen“ und „Fall“ heißen. Bevor wir aber dieses Ergebnis diskutieren, wollen wir zunächst noch andere Methoden der Zentralitätsanalyse besprechen. Betrachten wir allerdings den Graphen G4, so sehen wir Folgendes: Im G4 weisen zwar N2 und N3 die meisten direkten Verbindungen auf, aber wir würden dennoch sagen, dass N1 der zentrale Knoten ist. Obwohl N1 nur zwei direkte Verbindungen aufweist, liegt er doch eher „in der Mitte“. Um diese Intuition einzufangen, wurde der Begriff der Closeness-Zentralität eingeführt, den wir im Folgenden besprechen wollen.

Closeness-Zentralität

Eine Möglichkeit, „Zentralsein“ zu explizieren, liegt im Rekurs auf das Längenmaß. Wir bezeichnen oft einen Ort als zentral in einer Region, wenn von ihm aus die anderen Orte der Region in kurzer Zeit erreicht werden können. Diese topologische Überlegung steht hinter der Closeness-Zentralität, nach der ein Knoten dann zentral ist, wenn der Weg[34] zwischen ihm und anderen Knoten möglichst kurz ist. Das kann man anhand des Graphen G4 ersehen: Dort ist der Knoten N1 zentral, weil der Weg von ihm zu jedem anderen Knoten im Graphen kleiner ist als die Wege, die einer von N2 bzw. N3 aus gehen muss, um jeden anderen Knoten des Graphen zu erreichen. Ein Knoten ist hier zentral, wenn die Gesamtstrecke (bzw. die Durchschnittsstrecke) zwischen ihm und jedem anderen Koten im Graphen am kürzesten ist.

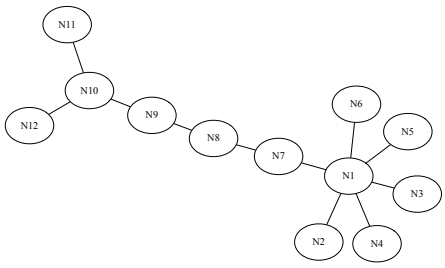
Heute existieren Algorithmen, welche diese Berechnung für jeden Knoten im Graphen übernehmen. Der Computer rechnet die Länge der kürzesten Strecken zwischen dem Knoten und jedem anderen Knoten im Graphen aus und bildet den Durchschnitt dieser Strecken. Da ein Knoten umso zentraler ist, je kleiner die Durchschnittsstrecke ist, bildet man von ihr den Umkehrbruch, um eine richtige Reihung zu bekommen. Man muss darauf achten, dass diese Methode nur für zusammenhängende Graphen anwendbar ist. Gibt es unzusammenhängende Teile in einem Graphen, so existiert mindestens ein Knoten, der für die anderen unerreichbar ist, und die kürzeste Strecke zwischen ihnen wäre unendlich groß.

Wendet man dieses Verfahren bei GB an und eliminiert man zwei Keywords, die mit dem Rest des Graphen nicht zusammenhängen, so erhalten wir die folgende Tabelle:

Keyword	Closeness-Zentralität
nennen	0.5738562091503269
Bild	0.5732941560561541
Menschen	0.5725464623410499
Fall	0.5695750891988324
Traum	0.5677335919818945

bedeuten	0.5671834625322998
Wort	0.5574603174603174
Erklärung	0.5560481317289424
Vorstellung	0.5542929292929293
Ausdruck	0.5535939470365699
zeigen	0.5534194768358021
Freud	0.5530708661417323
betrachten	0.5523749606794589
gut	0.5502977123158884
wissen	0.5490931832395247
beschreiben	0.5466998754669987

Hier zeigt sich, dass die Keywords „nennen“, „Bild“, „Mensch“ und „Fall“ am closeness-zentralsten sind. Zwar sind die ersten zehn „closeness-zentralsten“ Keywords auch die zehn degree-zentralsten, ab dem Platz elf unterscheiden sich die Tabellen aber. Während der Begriff der Closeness-Zentralität plausibel erscheint, gibt es dennoch Fälle, wo unsere Intuition und die Rechenergebnisse auseinander gehen. Betrachten wir den folgenden Graphen G6:



Intuitiv würden wir auch in diesem Graphen N1 als den zentralen Knoten bezeichnen. Eine Berechnung zeigt aber, dass N7 der closeness-zentralste Knoten ist. Viele Menschen würden intuitiv die vielen Verbindungen, die bei N1 zugegen sind, gegen seine etwas „abseits“ gelegene Position gegenrechnen. Es wäre demnach schön, wenn wir eine Möglichkeit hätten, die Degree- mit der Closeness-Zentralität zu verbinden. Eine Möglichkeit ist die Betweenness-Zentralität.

Betweenness-Zentralität

Die Betweenness-Zentralität misst, wie oft die kürzesten Pfade in einem Graphen einen Knoten beschreiten. Dabei darf dieser Knoten weder der Anfangs- noch der Endknoten sein. Die ihr zugrundeliegende Idee besagt, dass ein Knoten dann besonders wichtig ist, wenn er als „Schaltstation“ für andere Knoten gilt. Wenn ein Knoten eine Betweenness-

Zentralität von eins aufweist, dann heißt das, dass unter den vielen kürzesten Pfaden, die es in einem Graphen geben mag, ein Pfad diesen Knoten überschreitet (bzw. zwei „halbe“ Pfade diesen Knoten überschreiten). Je höher sie ist, desto zentraler ist der Knoten.

Die Berechnung dieser Zentralität ist zeitaufwendig. Um sie zu messen, müsste man die kürzesten Pfade zwischen allen Knoten ausrechnen. Da es mehrere Pfade zwischen zwei Knoten geben kann, die sich als gleich kurz erweisen, wird in diesem Falle der Anteil an Betweenness-Zentralität aufgeteilt. Wenn in einem Graph die kürzesten Pfade zwischen N1 und N2 ein Mal über N3 und N4 gehen und ein Mal über N3 und N5, dann bekommt N3 bei der Berechnung der Betweenness-Zentralität einen Punkt, während N4 und N5 jeweils einen halben Punkt bekommen.

Dieses Maß ist im Gegensatz zur Closeness-Zentralität auf alle Graphen anwendbar. Bemerkenswert ist dabei auch, dass alle Knoten, die nur eine direkte Verbindung zu einem anderen Knoten haben, automatisch den Wert 0 bekommen. Keiner der kürzesten Pfade beschreitet einen solchen Knoten.

Lässt man den Computer die Betweenness-Zentralität für alle Knoten in GB messen, so lauten die ersten sechszehn Plätze wie folgt.

Keyword	Betweenness-Zentralität
Menschen	68485.46078405058
Bild	66930.7373538551
Fall	65230.794208802086
nennen	62742.17336738318
bedeuten	59283.222909602584
Traum	54873.516273263114
Erklärung	45303.0734925954
Vorstellung	44481.19683069976
Ausdruck	40710.29030021832
Freud	40695.480267430925
zeigen	40261.2845724881
Wort	37203.41818812672
gut	36762.53851623657
betrachten	35223.071282629055
wissen	34737.27938494061
Ding	29830.747063798968

Es zeigt sich, dass das betweenness-zentralste Keyword in *LC* das Keyword „Menschen“ ist.

Zusammenfassung der Zentralitätsanalysen

Wie können wir die Ergebnisse der Zentralitätsanalysen interpretieren? Was heißt es, dass „nennen“ in GB sowohl das degree-zentralste als auch das closeness-zentralste Keyword ist? Können wir sagen, dass der Begriff „Menschen“ eine besondere Rolle spielt, nur weil der Knoten mit diesem Keyword der betweenness-zentralste ist?

In der Tat können wir *graphenimmanent* keine Antwort auf diese Fragen finden. Vergleichen wir den Graphen mit einer Funktionskurve im Koordinaten-System und die Suche nach zentralen Knoten mit der Suche nach Wendepunkten dieser Kurve, lässt sich klar machen, dass die Interpretation eines mathematischen Befundes an einem Modell immer nur im Bezug zum Modellierten möglich ist. Mathematisch lässt sich ausrechnen, ob eine bestimmte Kurve einen Wendepunkt besitzt und, sofern vorhanden, wo dieser Punkt ist. Was dieser Wendepunkt nun bedeutet, hängt wesentlich davon ab, *was* die Kurve abbilden soll. Stellt sie die Beziehung zwischen der Dosis eines Medikamentes und seiner Wirkung dar, so wirkt das Medikament (im Normalfall) im Bereich des Wendepunktes am besten. Stellt die Kurve die Beziehung zwischen Nahrungsangebot und Populationsgröße einer Tierart dar, so wächst die Population bei der gegebenen Nahrungsmenge am schnellsten. Ohne den Bezug zum Modellierten können wir den Wendepunkt nicht interpretieren.

Um die Frage nach der Interpretation der Zentren von GB beantworten zu können, ist es daher unerlässlich, sich zum Einen das Modellieren (d.h. die Trennung eines Textes in Sinneinheiten und die Zuordnung von Keywords zu diesen), und zum Anderen den ursprünglichen Text vor Augen zu halten. Nur so können wir vermeiden, „nennen“ und „Menschen“ als zentrale Themen in *LC* zu bezeichnen. Die Keywords „nennen“ und „Menschen“ sind die zentralsten Knoten in einem Keyword-Graph. Als solche spielen sie in einem Graphen eine wichtige Rolle, sie bilden die *verbindenden* Elemente, nicht aber die Haupt-Elemente. Betrachten wir einige Stellen, in denen „nennen“ auftaucht:

- „Aber im Fall einer gotischen Kathedrale spielen völlig andere Dinge eine Rolle für uns, und wir würden die Tür nicht ›richtig‹ nennen [...].“ (S. 19)
- „Wir haben es hier mit einer Art von Unbehagen zu tun, die wir ›gerichtet‹ nennen können, wenn ich z.B. Angst vor dir habe, ist mein Unbehagen auf dich gerichtet [...].“ (S. 27)
- „Wir könnten untersuchen, was wir eine Erklärung eines ästhetischen Urteils nennen würden.“ (S. 34)
- „Aber wenn ich dieses eigenartige Ding zeichne, das wir Gesicht nennen, und dann eines ein wenig anders, würdest du sofort sehen, daß es einen Unterschied gibt.“ (S. 48)
- „Wenn du sagst: »Glaubst du das Gegenteil?« — du kannst es das Gegenteil glauben nennen, aber es ist völlig verschieden von dem, was wir normalerweise das Gegenteil glauben nennen würden.“ (S. 77)

Das Keyword „nennen“ taucht immer dort auf, wo Wittgenstein auf den Sachverhalt hinweisen will, dass wir etwas einem Namen oder einer Eigenschaft zuschreiben. Und wenn wir die Stellen anschauen, wo „Menschen“ vorkommt, dann erhalten wir die folgende Liste:

- „Eine interessante Sache ist die Vorstellung, die die Menschen von einer Art von Wissenschaft der Ästhetik haben. Ich würde lieber davon reden, was mit Ästhetik gemeint sein könnte.“ (S. 24)
- „Die Menschen haben immer noch die Vorstellung, daß die Psychologie eines Tages alle unsere ästhetischen Urteile erklären wird, und sie meinen damit die experimentelle Psychologie. Das ist sehr komisch – tatsächlich sehr komisch.“ (S. 34)
- „Viele dieser Erklärungen werden akzeptiert, weil sie einen eigenartigen Reiz haben. Das Bild von Menschen mit unbewußten Gedanken ist reizvoll. Die Vorstellung einer Unterwelt, eines Geheimkellers. Etwas Verstecktes, Unheimliches. Vgl. die beiden Kinder bei Keller, die eine lebende Fliege in den Kopf einer Puppe stecken, die Puppe beerdigen und dann fortlaufen [...]. (Warum tun wir so etwas? Wir tun so etwas.) Man ist bereit, eine Menge zu glauben, weil es unheimlich ist.“ (S. 41)
- „Die gläubigen Menschen zweifeln nicht so, wie man normalerweise jede historischen [sic] Aussage, besonders aber Aussagen über eine weit zurückliegende Vergangenheit bezweifelt.“ (S. 80)

Diese Stellen belegen, dass das Keyword „Menschen“ dort kodiert wird, wo Wittgenstein auf die Meinungen der gewöhnlichen Menschen verweisen will. Ohne näher auf die Philosophie Wittgensteins eingehen zu können, erinnere ich daran, dass sie viel mit der Analyse des Sprachgebrauchs zu tun hat. Dass Menschen meinen, d.h. sich Urteile bilden und diese auch aussprechen, spielt daher eine große Rolle in seinem philosophischen Streben. Dass diese beiden Begriffe als zentral in *LC* aufgewiesen werden, zeigt, dass sie auch eine wichtige Rolle in der Ästhetik bzw. in der Religionsphilosophie Wittgensteins spielen. Jedoch können wir *nicht* behaupten, dass sich die Ästhetik oder die Religionsphilosophie Wittgensteins um das Meinen der Menschen drehen würde.

Diskussion

Das Erstellen einer *ontology* im Sinne des *semantic webs* stellt ein wichtiges Element in der Verlinkung der einzelnen Texte Wittgensteins dar. Verwendet man überall das gleiche Vokabular und indiziert man so alle Schriften Wittgensteins, ist es möglich, per Mausklick alle Stellen aufzulisten, in denen beispielsweise vom „(religiösen) Glauben“ gesprochen wird. So eine Sammlung erspart den Leserinnen das mühselige Durchforschen des Gesamtwerks und bietet gegenüber einer sogenannten Volltextsuche einen wesentlichen Vorteil: Eine Volltextsuche nach „glaube“ würde viele Stellen als Treffer aufweisen, in denen Wittgenstein seine Meinung mit „ich glaube, dass ...“

einleitet; doch nicht in allen diesen Stellen spricht Wittgenstein auch vom Glauben. Diese Stellen würden bei der korrekten Anwendung der *semantic web technology* in der Trefferliste gar nicht aufscheinen. Dank der Auflistung aller verknüpften Keywords bekommen die Wissenschaftlerinnen auch wertvolle Hinweise, welche anderen Stellen noch gesichtet werden sollen. Man kann sagen, dass die *ontology* ähnlichen Zwecken dient, die auch Register in einem Buch erfüllen.

Die *semantic web technology* kann bei der Erstellung digitaler Editionen also sinnvoll angewendet werden. Im Falle des Werkes Wittgensteins kommt ein weiterer Vorteil hinzu: Wittgenstein-Forscherinnen sind auf der ganzen Welt verteilt, sie können aber zusammenarbeiten: Während in Bergen der Nachlass indiziert wird, arbeitet man im Brenner-Archiv am Briefwechsel. Die Ergebnisse können leicht zusammengeführt werden, ohne dass sie einander in die Quere kommen. Die Verwendung dieser Technik erleichtert so auch die Arbeit der Forschenden.

Was für die *semantic web technology* gilt, gilt nicht unbedingt für das Modellieren des Textes als Graphen von Sinneinheiten und Keywords. Hier gilt es, zunächst die Kritik an unserem Modell von der Kritik am Modellieren selbst zu unterscheiden.

Man kann an der Qualität unserer Vergabe von Keywords zweifeln, wie man die Register eines Buches beanstanden kann. Unterschiedliche Bearbeiterinnen vergeben unterschiedliche Keywords. So unterliegt beispielsweise die Zusammenführung von Synonymen dem subjektiven Urteil des Bearbeitenden. Wir trennen z.B. „Mensch“ von „Menschen“, und „glauben“ von „Glaube“; würde man diese als Synonyme erachten, würde man eine andere *ontology* und damit eine andere Verlinkung bekommen.

Dem ist hinzuzufügen, dass auch die Register eines Buches der Willkür der Herausgeberinnen unterworfen sind. Ob eine bestimmte Ausgabe eines Werks ‚gut‘ genannt wird, hängt z.T. auch von der Qualität der Register ab, und in unserem Fall mögen die Leserinnen uns darauf aufmerksam machen, wie wir die Keyword-Liste verbessern könnten. In Bergen versucht man außerdem, den Benutzerinnen z.T. die Beschlagwortung der einzelnen Bemerkungen zu überlassen: Jeder Mensch darf seine eigene Keyword-Liste und somit seine eigene *ontology* von Wittgensteins Nachlass kreieren und sie auch anderen Forscherinnen zur Verfügung stellen. Wir erwarten mit Spannung, ob die Wittgenstein-Forschung durch ein solches Vorgehen profitieren wird.

Das Modellieren des Textes als ein Graph von Sinneinheiten und Keywords sehe ich nicht prinzipiell als fragwürdig an, allerdings würde ich die Zentralitätsanalysen und deren Interpretation nur mit großer Vorsicht durchführen. Während die Keywords einzelnen Einträgen in den Registern entsprechen, bilden die Sinneinheiten die Textstruktur ab, welche in Druckausgaben den Seiten, den Kapiteln bzw. den Absätzen entsprechen. Wenn der Vergleich zulässig ist, dann wären die Erstellung von Keyword-Listen und die Trennung des Textes in Sinneinheiten Pendants gängiger Techniken in einer Buch-Publikation, und diese Techniken sind ja an sich unverdächtig. Mit modernen Mitteln können diese Arbeiten nun schneller und automatisiert erfolgen. Das Modellieren eines Textes als Keyword-Graph ist methodisch demnach einwandfrei, wie ist die Zentralitätsanalyse diesbezüglich aber zu beurteilen?

Für sich betrachtet ist die Zentralitätsanalyse auch ein probates, graphentheoretisches Werkzeug, das v.a. in sozialwissenschaftlichen Studien zum Einsatz kommt. Die hier vorgestellten Zentralitätsbegriffe konkurrieren zwar untereinander, aber mit deren Hilfe können Wissenschaftlerinnen Aussagen über die Struktur eines sozialen Netzwerks treffen, die sonst nur schwer zu eruieren wären.

Wir haben aber gesehen, dass uns die Interpretation der Zentralitätsanalysen von Keyword-Graphen Schwierigkeiten bereitet. Ohne Bezug zum Text lässt sich schwer behaupten, warum „nennen" und „Menschen" die zentralen Knoten stellen, mit der Kenntnis der Wittgensteinschen Philosophie scheint dies eine Binsenweisheit zu sein, die man ohne zu rechnen auch gewusst hätte. Es hat den Anschein, als ob die Zentralitätsanalyse nur wenige Vorteile, aber großen Aufwand mit sich bringt. Es stellt sich daher die Frage, ob sich diese Art von Analysen überhaupt lohnt.

Ein abschließendes Urteil kann ich mir in diesem Augenblick nicht erlauben. Es scheint, dass die Zentralitätsanalysen für einen einzelnen kurzen Text kaum Erkenntnisgewinn mit sich bringen. Zudem taucht das zentrale Thema in vielen Texten – hier sind in erster Linie Gedichte zu nennen – gar nicht auf, und somit würde es in der Keyword-Liste und im graphentheoretischen Modell des Textes fehlen. Das führt letztlich dazu, dass eine solche Zentralitätsanalyse, wie sie in diesem Beitrag geschildert wird, gar nicht erfolgreich sein kann.

Es ist aber nicht ausgeschlossen, dass wir, wenn Ergebnisse der Zentralitätsanalysen von vielen Texten vorliegen, den Computer mit Gewinn einsetzen können. Hier möchte ich ein mögliches Einsatzgebiet nennen: das automatische Beschlagworten von Texten. Die Grundidee hinter dieser Anwendung ist, dass Texte zum gleichen Thema auch meist einen ähnlichen Wortschatz in ähnlicher Weise verwenden. Wenn wir graphentheoretische Modelle zweier Texte vor uns haben und diese einander ähnlich sind, dann können wir davon ausgehen, dass diese Texte auch ähnliche Themen ansprechen. Dank der Retro-Digitalisierungsbestrebungen liegen uns viele digital verfügbare Texte vor, deren Volltext die Forscherinnen durchsuchen können, während Informationen über den Inhalt fehlen. Könnte der Computer die Modelle selbständig aus einem Text generieren und diese mit bekannten Modellen vergleichen, wäre er in der Lage, für bestimmte Textsorten die Schlagworte automatisch zu generieren. Diese betreffen v.a. wissenschaftliche Texte, in denen das Anliegen der Autorinnen relativ klar zum Ausdruck gebracht wird. Das Verständnis für einen bestimmten Text kann der Computer zwar aufbringen, wir hätten aber dann zusätzlich zu der Volltextsuche ein wertvolles Werkzeug, das Benutzerinnen helfen kann, in der Welt der unüberschaubaren Textangebote das Gesuchte auch zu finden.

Anmerkungen

1 Ludwig Wittgenstein: Vermischte Bemerkungen. Eine Auswahl aus dem Nachlaß. Herausgegeben von Georg Henrik von Wright unter Mitarbeit von Heikki Nyman. Neubearbeitung des Textes durch Alois Pichler. Frankfurt am Main: Suhrkamp, 1994.

2 Ludwig Wittgenstein: Vorlesungen und Gespräche über Ästhetik, Psychoanalyse und religiösen Glauben. Neu übersetzt von Ralph Funke. Frankfurt am Main: S. Fischer 2000. Die Abkürzung *LC* stammt aus dem

englischen Originaltitel: Ludwig Wittgenstein: Lectures and Conversations on Aesthetics, Psychology, and Religious Belief. Edited by Cyril Barrett. University of California Press 1966.

3 Zu GABEK (Ganzheitliche Bewältigung Komplexer Probleme) gibt es eine Vielzahl einschlägiger Literatur, die unter http://www.gabek.com/index.php?id=90 (2012/03/06) aufgelistet wird. Ganz aufschlussreich ist der online publizierte Artikel von Josef Zelger: „Welchen Zielen dient eine PC-unterstützte Textanalyse? oder Über Wissensdarstellung durch linguistische Netze", online unter http://www.gabek.com/uploads/media/Linguistische_Netze_01.pdf (2012/03/06).

4 Das Projekt *Vermischte Bemerkungen Neu* wurde bereits in Kerstin Mayr, Joseph Wang: „Vermischte Bemerkungen Neu", Vernetzung des Nachlasses Wittgensteins mit seinem Briefwechsel. In: Mitteilungen aus dem Brenner-Archiv 26, 2007, 195-199, beschrieben. Ein vorläufiges Ergebnis der GABEK-Analyse von *VB* kann man in Kerstin Mayr: (Re)-Constructing the Semantic Architecture of Wittgenstein's Vermischte Bemerkungen by Syntactic Analysis. In: Alois Pichler, Herbert Hrachovec (Hg.): Wittgenstein and the Philosophy of Information. Heusenstamm: Ontos 2008, 205-211 studieren.

5 Ludwig Wittgenstein: Nachlass: The Bergen Electronic Edition. Charlottesville, VA: Intelex, 2003.

6 Ludwig Wittgenstein: Gesamtbriefwechsel/Complete Correspondence. Electronic Edition. Herausgegeben von Monika Seekircher, Brian McGuinness, Anton Unterkircher. Charlottesville, VA: Intelex, 2004.

7 Siehe die Web-Präsenz des Projektes unter: http://www.discovery-project.eu/home.html (2012/02/12).

8 Das gesamte Material von *Philosource* kann online eingesehen werden: http://www.discovery-project.eu/philosource.html (2012/03/05).

9 Siehe „Introduction" in BEE, online unter: http://library.nlx.com/xtf/view?docId=witt_nl_de/witt_nl_de.00.xml;chunk.id=div.PMPreface.Page.2;toc.depth=1;toc.id=div.PMPreface.Page.2;brand=default (2012/03/06).

10 So ist z.B. die berühmte Bemerkung Wittgensteins, dass ihn „Boltzmann, Hertz, Schopenhauer, Frege, Russell, Kraus, Loos, Weininger, Spengler, Sraffa beeinflußt" haben, eingebettet in Bemerkungen über die Arithmetik (Siehe MS 154, Taschennotizbuch, 15r-19v, in BEE).

11 Siehe: VB, ix-xi.

12 Siehe Mayr (2008). Es ist in diesem Zusammenhang wichtig darauf hinzuweisen, dass die GABEK-Analyse, welche hinter der Analyse von VB steht, primär ein Werkzeug zur Analyse von gesprochenen Texten ist. Gerade bei der Auswertung von Umfragen und Interviews, in denen eine Vielzahl von Personen befragt wird, zeigt sich die Stärke der GABEK-Methode. Da bei Interviews die Textmaterialien, deren Inhalt bewältigt werden muss, zwar inhaltlich heterogen sind, aber in ihrer Struktur große Ähnlichkeiten aufweisen, ist GABEK sehr hilfreich, um „linguistische Gestalten" ausfindig zu machen.

13 Siehe LC, 9.

14 So gibt es ausgedehnte Fußnoten und Einschübe auf den Seiten 12, 19 und 30.

15 Siehe hierzu beispielsweise das dritte Kapitel von: Josef Zelger und Andreas Oberprantacher: „Processing of Verbal Data and Knowledge Representation by GABEK®-WinRelan®." In: Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, Bd. 3(2), Art. 27, 2002, online unter http://www.qualitative-research.net/index.php/fqs/article/view/866/1882 (2012/03/20).

16 Wir bezeichnen die Sinneinheiten mit zwei Buchstaben und reihen sie nach dem Alphabet. „Aa" kommt vor „Ab", und „Bz" kommt vor „Ca". Diese Passage stammt von LC, S. 45f.

17 Die Website „Wittgenstein Source" ist erreichbar über dieses URL: http://www.wittgensteinsource.org/ (2012/03/18).

18 Siehe hierzu: World Wide Web Consortium (Hg.): „OWL 2 Web Ontology Language. Document Overview", online unter: http://www.w3.org/TR/2009/REC-owl2-overview-20091027/ (2012/03/06).

19 Siehe hierzu: World Wide Web Consortium (Hg.): „Resource Description Framework (RDF): Concepts and Abstract Syntax", online unter: http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/ (2012/03/06). Wie man hier leicht erkennen kann, werden die Techniken RDF und OWL von dem W3Consortium verwaltet, das auch für die Spezifikation der Webseiten (HTML und XHTML) verantwortlich ist.

20 Durch das „Verbinden" unterschiedlicher *ontologies* liegen uns prinzipiell Rechenmöglichkeiten vor, welche die unterschiedlichen Informationen in diesen *ontologies* nutzen; dabei müssen sie nicht unbedingt das gleiche Vokabular verwenden. Erstellt jemand eine *ontology*, in der das Prädikat „enthält Keyword" vorkommt, und definiert man die Beziehung „enthält Keyword" als die „Umkehrrelation" zum Prädikat „ist ein Schlagwort in", kann der Computer beide *ontologies* zu einer großen *ontology* verbinden und mit

den Daten aus beiden *ontologies* rechnen. Das Verbinden der unterschiedlichen Informationsquellen stellt den Hauptvorteil der *semantic web ontology* dar.

21 RDF selbst basiert auf der Graphentheorie, daher ist die Anwendung derselben naheliegend. Die Analyse hätte auch mit Hilfe der Aussagen- bzw. der Prädikatenlogik oder mit Werkzeugen der Statistik erfolgen können. Wir sehen von diesen ab, da wir auf der einen Seite neben den vorhandenen Sinneinheiten und Keywords noch anderer Daten bedürfen, damit der Computer mit aussagen- bzw. prädikatenlogischen Modellen arbeiten kann. Auf der anderen Seite können die wichtigsten statistischen Analyseergebnisse auch von graphentheoretischen Modellen ausgehend berechnet werden.

22 Siehe: Ulrik Brandes: „Graphentheorie". In: Christian Stegbauer, Roger Häußling (Hg.): Handbuch Netzwerkforschung. Wiesbaden: Springer 2010, 345-353, hier 345. Dieses Kapitel ist zudem eine gute Einführung in die Graphentheorie im Allgemeinen.

23 Siehe dazu die Website des Projektes online unter: http://protege.stanford.edu/ (2012/03/12).

24 Hier sei der Hinweis angebracht, dass die Umwandlung einer XML-Datei in eine andere per XSLT eine gute Möglichkeit bietet, unsere vorhandene *ontology* in eine zweite *ontology* zu transformieren, in der diese Beziehungen als Behauptungen festgeschrieben sind.

25 In unserem Projekt verwenden wir für die Darstellung und Analyse der Graphen *Prefuse Information Visualization Toolkit* (online unter: http://prefuse.org/ (2012/03/06)), *JUNG* (online unter: http://jung.sourceforge.net/ (2012/03/06)) und *JGraphT* (online unter: http://www.jgrapht.org/ (2012/03/06)).

26 Die Graphen GA1, GB1 und GC1, die hier abgebildet werden, werden mittels des Programms GraphViz (http://www.graphviz.org/ (2012/03/06)) generiert.

27 Ein Graph ist dann ungerichtet, wenn für eine Kante die Reihenfolge der durch sie verbundenen Knoten unwichtig ist.

28 Diese Analyse wird beispielsweise von den Routenplanern öfters ausgeführt, um den kürzesten Weg zwischen zwei Orten herauszufinden.

29 Hier ist es wichtig zu betonen, dass diese Schlüsse indirekt, d.h. über die Analyse der Modelle erfolgen. Ähnliche Probleme haben die Naturwissenschaftlerinnen, wenn sie aus Daten zu einem Tiermodell Rückschlüsse auf die Menschen erzielen wollen.

30 Es ist bemerkenswert, dass dieses Maß v.a. für die Soziologinnen von Bedeutung ist. So nehmen die Zentralitätsanalysen in allen einführenden Werken zur sozialen Netzwerkforschung einen wichtigen Platz ein. Siehe hierzu: Stanley Wasserman, Katherine Faust (Hg.): Social Network Analysis, Methods and Applications. Cambridge: Cambridge University Press 1994, 169-219; und Peter Mutschke: „Zentralitäts- und Prestigemaße" in: Christian Stegbauer, Roger Häußling (2011), 365-378.

31 Oft werden diese Zahlen so normiert, dass sie zwischen 0 und 1 liegen; je größer die Zentralität eines Knotens ist, umso zentraler ist auch der besagte Knoten. In diesem Beitrag haben wir von dieser Normierung abgesehen.

32 Der Vollständigkeit halber muss gesagt werden, dass die Zentralitäts-Maße, die hier vorgestellt werden, alle auf ungerichtete Graphen Anwendung finden, wobei die Kanten nicht gewichtet, d.h. als gleich lang, angenommen werden.

33 Eine Gesamtschau über nicht weniger als acht Zentralitätsbegriffe in der Graphentheorie liefert Wasserman/Faust (Anm. 30). Mutschke erläutert in seinem Beitrag elf unterschiedliche Arten, wie man die Zentralität bestimmen kann. Während die englischen Bezeichnungen der einzelnen Zentralitätsbegriffe einheitlich ausfallen, gibt es Unterschiede in den deutschen Benennungen. Ich folge in diesem Artikel den Benennungen von Mutschke.

34 In der Graphentheorie ist der Pfad zwischen zwei Knoten A und B definiert als eine Folge von Knoten, die mit A anfängt und mit B aufhört, wobei zwischen zwei aufeinander folgenden Knoten eine Kante herrscht und kein Knoten mehr als ein Mal in der Folge vorkommt. In einem Graphen, bei dem alle Kanten als gleich lang angesehen werden, ist der kürzeste Weg zwischen A und B jene Folge, die am wenigsten Knoten aufweist. Es ist wichtig zu beachten, dass es mehrere kürzeste Wege zwischen zwei Knoten geben kann.