

3. Sinn- und Bedeutungszusammenhänge²

Zusammenfassung

Im dritten Kapitel werden erste theoretische Grundlagen für eine qualitative Textanalyse durch GABEK[®] präsentiert. Normalsprachliche Äußerungen werden als Sätze ausgesprochen, die man kurz im Bewusstsein behalten kann. Sätze sind nach Wittgenstein (1971, S. 43) die minimalen Elemente der Bedeutung. Darum werden einzelne Sätze oder kurze Textabschnitte als ungerichtete Graphen abgebildet, in denen die Schlüsselbegriffe des Satzes vollständig miteinander verbundene Knoten sind. Es sind jene lexikalischen Begriffe, die man erfassen muß, um den Sinn des Satzes zu verstehen. Die Kanten repräsentieren subjektive Sinnzusammenhänge des Sprechers. Der Satz als ganzer ist das, was für den Sprecher "Sinn macht", den er mitteilen will. Alle Sätze der gesamten verfügbaren verbalen Datenbasis bilden zusammen ein komplexes Gesamtnetz, das sowohl subjektive Sinnzusammenhänge als auch intersubjektiv verwendete Bedeutungszusammenhänge enthält.

Die Kanten des Begriffsnetzes werden durch die Anzahl der Sätze, in denen die entsprechenden Begriffspaare vorkommen, gewichtet. Wenn man nur die hoch gewichteten Kanten herausgreift, das heißt die Begriffspaare, die häufig verwendet werden, können allgemeine Bedeutungszusammenhänge selektiert werden, die häufige linguistische Interaktionen der Community charakterisieren. Das gewichtete Begriffsnetz, wird als Assoziationsnetz der Community aufgefasst, da dessen Kanten die Assoziationsbahnen anzeigen, die in den Texten der Community am häufigsten vorkommen. Begriffsnetze dienen für alle Operationen der Textanalyse durch GABEK[®] als Grundlage.

In den Unterabschnitten 2.2 und 2.3 werden formale Eigenschaften von Begriffsnetzen besprochen und die Frage beantwortet, wie linguistische Netze durch Hinzufügung von neuen Texten wachsen. Diese theoretisch-formal orientierten Abschnitte dienen dazu, GABEK[®]-Netze mit anderen linguistischen Netzen, die in der Literatur beschrieben werden, zu vergleichen. Sie sind für Leser gedacht, die sich für Themen der formalen Netzwerkanalyse interessieren.

²Dieser Abschnitt ist eine gekürzte und überarbeitete Fassung von Zelger, J. (2013)

Wenn wir Texte als linguistische Netze darstellen wollen, stellen wir fest, dass normalsprachliche Texte strukturiert sind, wenn auch recht unterschiedlich. Sie sind gegliedert in Absätze und Sätze. Dies ist wegen der Begrenzung des menschlichen Bewusstseins notwendig. Der Grund dafür ist, dass die menschliche Fähigkeit der bewussten Speicherung und Verarbeitung stark eingeschränkt ist (vgl. Sergin, 1992). Laut Miller (1956) ist es nicht möglich, mehr als drei bis neun lexikalische Ausdrücke gleichzeitig im Bewusstsein zu behalten. Da es aber darum geht, dass ein Satz 10 bis 30 Sekunden als gedankliche Einheit bewusst bleibt, werden Sätze begrenzt in Länge und Komplexität. Wenn Texte aber manchmal doch in sehr langen Sätzen vorgegeben werden, erfordert dies vom Hörer oder Leser, dass er selbst zunächst Absätze in kleinere Einheiten zerlegt, die nacheinander erfasst werden können. Dies erfordert von ihm eine Zusatzleistung der Strukturierung und Speicherung.

Im Verfahren GABEK[®] beginnen wir deswegen auch mit einer Aufteilung von Texten in kurze Texteinheiten. Dabei folgen wir meistens der vorgegebenen Strukturierung des normalsprachlichen Textes. Die minimalen sprachlichen Einheiten sollen natürlich nicht willkürlich abgeschnittene Textteile sein, etwa bloße Zeilen oder Zeilengruppen. Vielmehr soll jeder Text eine sinnvolle gedankliche Einheit bilden. Wenn jedoch auf der bloß formalen Ebene von "Sätzen" die Rede ist, die in der Software WinRelan[®] verarbeitet werden, so geht es uns nicht um bloße Zeichensequenzen sondern um deren Bedeutung. Es kann sich um deskriptive Aussagen handeln, aber auch um Normen und Werturteile. Die minimale Einheit von "Bedeutung" ist dabei immer ein Satz und nicht ein Begriff. Ich schließe mich hier Wittgenstein an (1971, S. 43), der "Bedeutung" pragmatisch definiert als eine Funktion dessen, was Personen mit Sprache tun (Alston, 1967, S.238). Wir lernen auch die Bedeutung von Begriffen meistens dadurch, dass wir sie in bestimmten Lebenssituationen auf eine bestimmte Art und Weise verwenden. Doch bleibt ein Satz die kleinste sprachliche Einheit, mit der eine vollständige Handlung durchgeführt werden kann³³.

Wenn wir uns mit der qualitativen Analyse von Texten befassen, so beginnen wir damit, Texte in "Sinneinheiten" zu zerlegen. Texte, die mit GABEK[®] analysiert werden sollen, werden zuerst in kurze Texteinheiten aufgespaltet, so dass jede Einheit zwischen 3 und 9 lexikalische Begriffe enthält. Es sollen jene inhaltlich selbständigen Begriffe sein, die nötig sind, den Sinn des Satzes nachzuvollziehen. Logische und grammatikalische Begriffe sowie Zahlbegriffe werden dabei außer Acht gelassen. Die Texteinheiten sollen selbständige gedankliche Einheiten sein. Deswegen ist es auch am besten, wenn die Texteinheiten in einem eigenen Arbeitsdurchgang händisch bestimmt werden. Doch gibt es in WinRelan[®] vor allem für große Datenmengen auch automatische Verfahren zur Ein-

³³„As a first step, we can say that for a sentence to have a certain meaning is for that sentence to be used to perform a certain illocutionary act.“ (Alston, 1967, S. 238)

teilung in Texteinheiten, für die man die durchschnittliche Länge vorher festlegen kann.

Nach der Aufteilung des Textes in kurze Einheiten erfolgt die Grundkodierung: Es werden händisch oder auch automatisch die lexikalischen Begriffe markiert, die den normalsprachlichen Text am besten repräsentieren. Synonyme und Homonyme werden in einem eigenen Arbeitsgang mit Softwareunterstützung eliminiert. Bloße Mengen von Begriffen sehen wir dann als Repräsentanten der Texteinheiten. Damit wird es möglich, eine Texteinheit nicht nur als normalen Text zu speichern sondern zusätzlich auch als eine Begriffsmenge.

Die Idee dazu habe ich von Pask (1976) bzw. Pask & De Zeeuw (1992) übernommen, die eine "Protologic" darauf aufzubauen versuchen, sowie von Bickerton (1990), der die pragmatische Anwendbarkeit einer ähnlich konzipierten "Protolanguage" nachweist. Sie verzichten für die Protolanguage auf alle grammatikalischen und logischen Ausdrücke. Sätze und Satzgruppen werden damit formal bloß als Mengen oder Mengen von Mengen lexikalischer Begriffe aufgefasst. Dies erlaubt es, im Programm WinRelan[®] sehr vielfältige Operationen über Mengen lexikalischer Ausdrücke automatisch durchzuführen.

3.1 Darstellung von Sinn- und Bedeutungszusammenhängen durch Begriffsnetze

Um die Entwicklung von Bedeutung anhand eines Beispiels zu erklären, werde ich im Folgenden Texte von Mitarbeiterneiner Forschungsabteilung der Daimler AG verwenden. Die versuchsweise Einführung eines neuen Arbeitszeitmodells in der Niederlassung in Moabit (Berlin) hatte zu Konflikten geführt zwischen Personalführung und Betriebsrat einerseits sowie zwischen drei Niederlassungen andererseits. Intendiert war ein einheitliches Arbeitszeitmodell in allen drei Niederlassungen. Um die Mitarbeiter einzubeziehen, haben wir ihnen schriftlich drei Fragen gestellt über Erfahrungen und Wünsche in Bezug auf die Arbeitszeitregelung. Im Ergebnis der GABEK[®]-Auswertung wurde ein neues Arbeitszeitmodell in Übereinstimmung mit Personalleitung und Betriebsrat für alle drei Niederlassungen beschlossen, das auch den Zielen und Grundwerten der Mitarbeiter Rechnung trug. Die hohe Arbeitszufriedenheit konnte damit erhalten werden. Vorerst verwende ich aus dem Projekt jedoch nur Beispieltex te, um zu zeigen, wie Sinn- und Bedeutungszusammenhänge in Form semantischer Netze dargestellt werden und wie linguistische Netze analysiert werden. Im Kapitel 4 werde ich dann eine Begriffsanalyse des Begriffes "Arbeitszeitmodell B" durchführen, die als methodisches Muster für Begriffsanalysen auf der Basis normalsprachliche Texte dienen kann.

Wir finden in den schriftlichen Antworten der Daimler-Mitarbeiter auf die drei Fragen zunächst einmal den Satz:

E10

*Das **Arbeitszeitmodell B** fordert vom **Mitarbeiter** mehr **Verantwortung** in Bezug auf die **Arbeit**, gibt ihm aber auch mehr **Freiheit** und fördert damit die **Motivation**.*

In der GABEK®-Protolanguage werden Aussagen in der Software WinRelan® zunächst als ungerichtete Graphen dargestellt, die wenigstens drei oder mehr lexikalische Begriffe vollständig miteinander verbinden. E10 kann also auch als Graph wie in Abbildung 3-1 repräsentiert werden. Die Kanten des Graphen geben Zusammenhänge wieder, die der Sprecher von E10 als sinnvoll erachtet. Der subjektive Sinn des Satzes ist damit das, was der Sprecher sagen will, auf jeden Fall etwas, was für ihn "Sinn macht". Wenn man ihn fragen würde, was er genauer damit meint, so könnte der Sprecher von E10 vermutlich für jede Kante des Graphen ein Beispiel oder eine Erklärung finden. Ein Satz kann sprachliche Zusammenhänge darstellen oder Meinungen des Sprechers über empirische, evaluative oder normative Zusammenhänge.

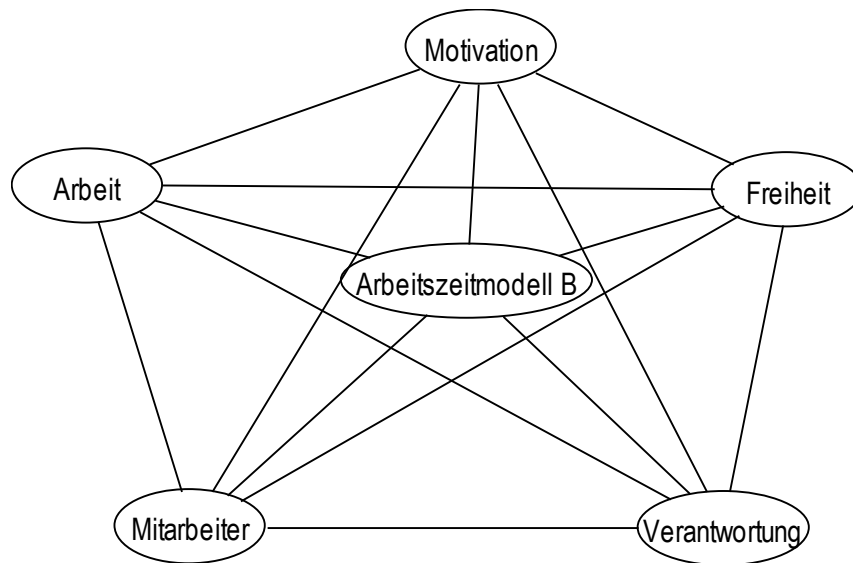


Abbildung 3-1: Darstellung der subjektiven Sinnzusammenhänge des Satzes E10 für einen Sprecher

Nun gibt es in der verbalen Datenbasis auch die Texteinheit D92, die in einem Gespräch auf E10 folgen könnte.

D92

*Der **Erfolg** des **Arbeitszeitmodells B** hängt ab von den **Arbeitsergebnissen** und von der **Zufriedenheit** der **Mitarbeiter**. Er hängt mit der Frage zusammen, wie*

weit sich Mitarbeiter **verantwortlich** fühlen und inwieweit ihnen der **Vorgesetzte** **vertraut**.

Wir sehen in Abbildung 3-2, dass sich drei Begriffe wiederholen und dass weitere Begriffe neu dazukommen. Die Verbindungen zwischen den Begriffen, die sich wiederholen, werden fett hervorgehoben.

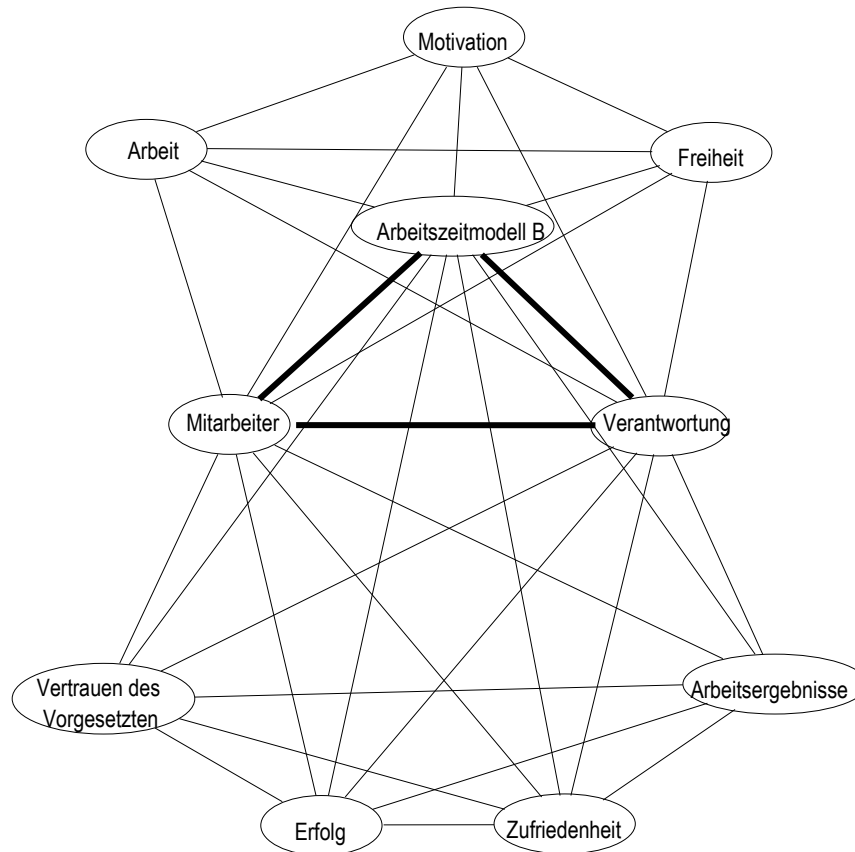


Abbildung 3-2: Darstellung der Sätze E10 und D92 als Begriffsnetz

Wir sehen, dass die Verbindungen zwischen “Arbeitszeitmodell B”, “Verantwortung” und “Mitarbeiter” als subjektive Sinnzusammenhänge in beiden Sätzen vorkommen. Damit werden sie in gemeinsame Bedeutungszusammenhänge transformiert, wenigstens für zwei Personen im Kontext des Dialoges.

Nun kommt in der verbalen Datenbasis auch ein dritter Satz vor:

D99

*Das **Arbeitszeitmodell B** fördert die **Verantwortung** und **Zufriedenheit** der **Mitarbeiter** und reduziert den **Koordinationsaufwand**. Es erhöht die **Effizienz** des **Unternehmens**, wenn das **Vertrauen des Vorgesetzten** gegeben ist.*

Damit wächst die Graphik zur Abbildung 3-3:

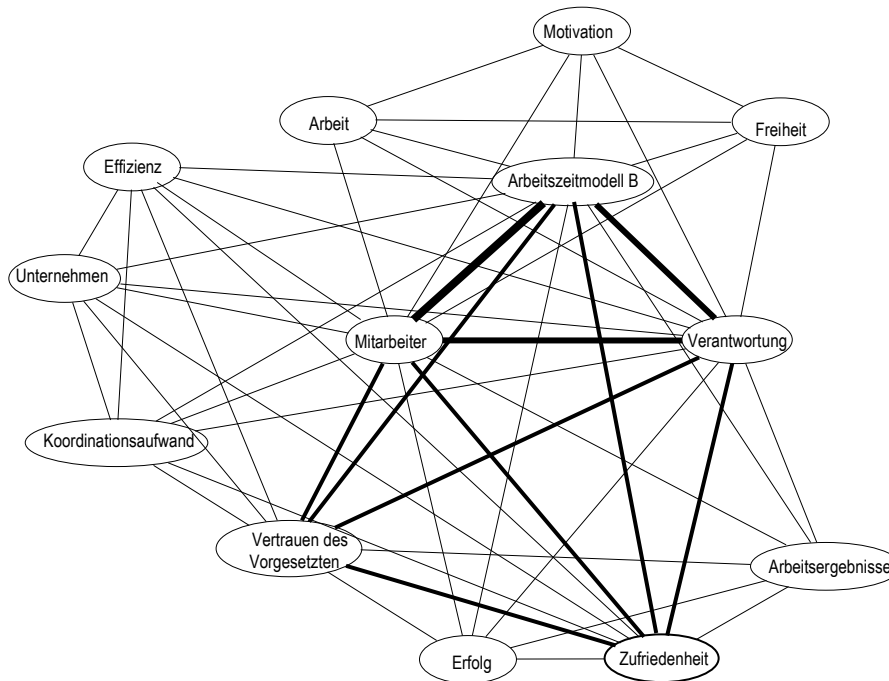


Abbildung 3-3: Darstellung des Dialogergebnisses von E10, D 92, D99, wobei wiederkehrende Sinnzusammenhänge durch fette Linien gekennzeichnet sind.

Die Abbildung 3-3 zeigt die Bedeutungszusammenhänge, die sich durch die drei Sätze ergeben haben. Die Beziehungen zwischen “Arbeitszeitmodell B”, “Verantwortung” und “Mitarbeiter” wurden in allen drei Sätzen verwendet. Deren Beziehungen zu “Zufriedenheit” und “Vertrauen des Vorgesetzten” kommen nur in zwei Sätzen vor. Von “Bedeutungszusammenhang” spreche ich dann, wenn wenigstens zwei Personen mit der Beziehung zwischen zwei Begriffen einen subjektiven Sinn verbinden.

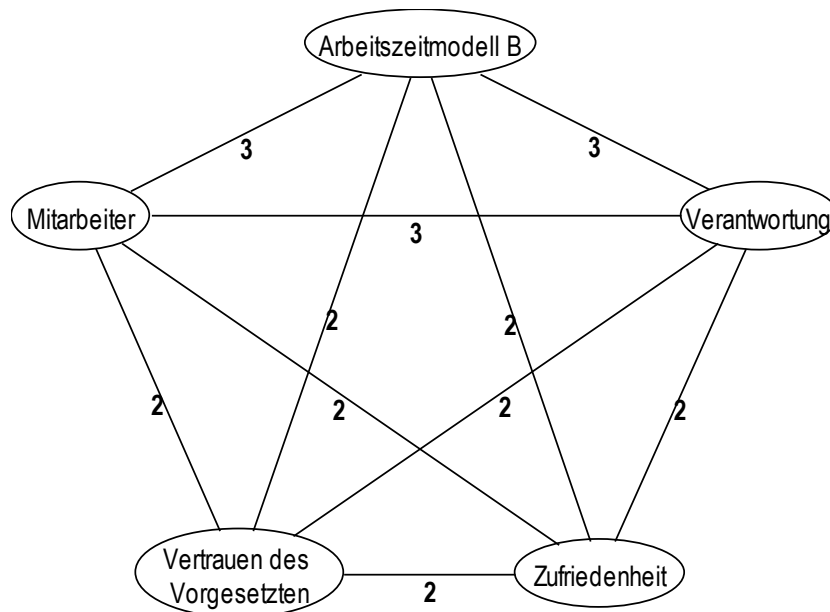


Abbildung 3-4: Mehrfach verbundene Begriffe im Dialog E10, D92, D99.

Wenn in den Texten einer Organisation Beziehungen zwischen Begriffen wiederholt verwendet werden, dann wird man die gemeinsame formale Beziehungsstruktur durch einen neuen Text interpretieren. Die Abbildung 3-4 kann wie folgt interpretiert werden: *“Das Arbeitszeitmodell B erfordert Verantwortung der Mitarbeiter. Verantwortung führt zu mehr Zufriedenheit der Mitarbeiter wenn Vorgesetzte den Mitarbeitern vertrauen”*.

Wenn der neue Satz durch ein Beispiel demonstriert werden kann, das von wenigstens zwei Personen akzeptiert wird, oder wenn er zum gemeinsamen Handeln führt, dann hat das Gespräch zu einer neuen intersubjektiven Bedeutung geführt. Das heißt, es kommt zu einer intersubjektiven Übereinstimmung darüber, *“was mit dem neuen Satz gesagt wurde”*. Das Beispiel repräsentiert dann als empirisches Modell einige Bedeutungszusammenhänge durch relevante Beziehungen zwischen Entitäten oder Attributen in der Lebenswelt der Gemeinschaft. Bedeutung in diesem Sinne ist eine Leistung der Gemeinschaft, bestimmte Aspekte der Erfahrung in Zusammenhang zu bringen und entsprechend zu benennen. Die Abbildung 3-4 stellt mit der Interpretation und mit dem Hinweis auf ein Beispiel dar, was im Gespräch gelernt wurde.

Der Prozess der Entwicklung von Bedeutung durch Gespräche wird damit als Lernprozess verstanden, bei dem die Beteiligten jene Beziehungen zwischen Begriffen anzuwenden lernen, die in der Gemeinschaft häufig verwendet werden und die in ihrer Lebenswelt zum erfolgreichen gemeinsamen Handeln führen. Die in Gesprächen zusätzlich geäußerten subjektiven Sinnzusammenhänge kann man als Angebote auffassen, die einladen, das Gespräch in neue Richtungen weiterzuführen.

ren. Sie liefern das kreative Potential, das die Offenheit der Gespräche garantiert. Doch erfordert der Lernprozess, dass bloß subjektive Sinnzusammenhänge immer wieder fallen gelassen werden, wenn sie von anderen Personen der Gemeinschaft nicht aufgegriffen wurden oder wenn keine Beispiele aus der Lebenspraxis der Beteiligten gefunden wurden, durch die sie intersubjektiv demonstriert werden können. Damit ist also auch ein Vergessen subjektiver Sinnzusammenhänge mit dem Lernprozess verbunden.

Wenn wir alle Sätze, die den Begriff "Arbeitszeitmodell B" enthalten, übereinanderlegen, entsteht ein recht komplexes Begriffsnetz. Im Zentrum des Netzes der Abbildung 3-5 steht der Begriff "Arbeitszeitmodell B". Er ist über 382 Texteinheiten mit 317 Begriffen verbunden, die ihrerseits auch wieder miteinander in Beziehung stehen.

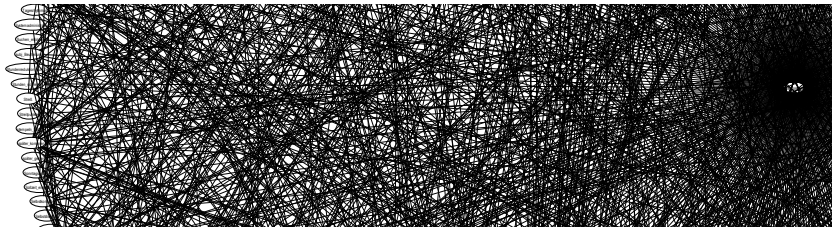


Abbildung 3-5: Ausschnitt aus dem Begriffsnetz, das die 317 Nachbarbegriffe um den Begriff "Arbeitszeitmodell B" (helles Feld im Bild rechts) enthält, mit Querverbindungen, die durch 382 Texteinheiten belegt werden.

Das Begriffsnetz der Abbildung 3-5 enthält alle subjektiven Sinnzusammenhänge und alle intersubjektiven Bedeutungszusammenhänge, die den Begriff "Arbeitszeitmodell B" einbeziehen. Ich nehme mit Steyvers und Tenenbaum an, dass die Bedeutung eines Begriffes durch das Beziehungsgefüge des Begriffes gegeben ist, das zeigt, wie der Begriff mit anderen Begriffen verbunden ist. Steyvers & Tenenbaum (2005, 73) nehmen an: "Under the semantic net view, meaning is inseparable from structure: The meaning of a concept is, at least in part, constituted by its connections to other concepts." (Steyvers and Tenenbaum 2005, 43).

3.2 Formale Merkmale von Begriffsnetzen

In den folgenden Unterabschnitten 3.2 und 3.3 werde ich mich mit formalen Merkmalen linguistischer Netze befassen und mit deren Wachstum bei der Hinzufügung neuer Texte. Leser, die sich weniger für formale Überlegungen interessieren, können diese Abschnitte übergehen, um gleich zur wissenschaftstheoretischen Methode der Begriffsanalyse durch GABEK® im Kapitel 4 überzugehen.

Um unser Netz mit anderen linguistischen Netzen zu vergleichen, können wir statistische Merkmale verwenden. Zur Angabe der Größe von Netzwerken wird normalerweise die Anzahl n der Knoten verwendet. Wenn es im Netz für jeden Knoten einen Pfad zu allen anderen Knoten gibt, dann kann die durchschnittliche Länge L des kürzesten Pfades zwischen Begriffspaaren angegeben werden. Es kann natürlich vorkommen, dass nur ein Teil des Gesamtnetzes vollständig verbunden ist. Das obige Netz der Abbildung 3-5 ist vollständig verbunden, da nur die Nachbarbegriffe zu "Arbeitszeitmodell B" wiedergegeben wurden und deren Beziehungen untereinander. Es enthält aber nicht alle Daten des Projekts. Wenn wir das vollständige Netz über die gesamte Datenbasis darstellen, dann zeigt sich, dass nur 412 der insgesamt 451 Texteinheiten miteinander verbunden sind. Das Netz der folgenden Abbildung 3-6 enthält 696 der insgesamt 712 Knoten. Die restlichen 39 Texteinheiten mit 16 Begriffen bilden Inseln außerhalb des Netzwerkes, das wir hier formal analysieren. Interessant ist auch der Durchmesser D , der die Anzahl der Schritte angibt, die für den kürzesten Pfad zwischen den Knoten mit der größten Distanz nötig sind. Als besonders interessant wird sich für den weiteren Verlauf unserer Abhandlung der Clusterkoeffizient C erweisen. Wenn wir für einen beliebigen Knoten i alle unmittelbar damit verbundenen Knoten "Nachbarn" von i nennen, so können zwei Nachbarn des zufällig ausgewählten Knotens i auch miteinander zusammenhängen. Der Clusterkoeffizient C gibt dann für das Gesamtnetz den durchschnittlichen Bruchteil der Beziehungen zwischen Nachbarn eines beliebigen Knotens i zu allen möglichen Beziehungen zwischen den Nachbarn von i an. Wenn alle Nachbarn von i verbunden sind, dann beträgt C für i 1. Wenn keine Nachbarn verbunden sind 0. Außerdem spielt der durchschnittliche Knotengrad k eine Rolle. Dies ist die durchschnittliche Anzahl der Verbindungen, die Knoten im Gesamtnetz haben. Desweiteren kann die Verteilung der Knotengrade $P(k)$ angegeben werden. $P(k)$ ist die Wahrscheinlichkeit, im Netz einen Knoten mit dem Knotengrad k zu finden. Schließlich wird oft auch noch der Exponent γ des Potenzgesetzes angegeben, dem die Verteilung der Knotengrade folgt. γ gibt dann die negative Steigung der Linie an, die in einem log-log-Diagramm am besten zur Verteilung des Knotengrades $P(k)$ passt.

Um die formale Struktur semantischer Netze empirisch zu erforschen, haben Steyvers und Tenenbaum (2005) drei Arten semantischer Netze untersucht. Sie analysierten erstens die assoziativen Verbindungen von 5018 Worten, die von Nelson, D.L.L. & Schreiber, T.A. (1999) protokolliert wurden.⁴ Zweitens untersuchten sie ein semantisches Netz mit 29381 Knoten, das den Thesaurus der englischen Wörter und Phrasen von Roget (1911) abbildet, die in 1000 seman-

⁴Den Teilnehmern wurde ein Reizwort vorgegeben. Sie sollten das Wort niederschreiben, das ihnen unmittelbar darauf einfiel. Zwei Worte wurden dann im Wortnetz durch eine Kante verbunden, wenn sie bei wenigstens zwei Versuchspersonen vorkamen. (Steyvers und Tenenbaum 2005, 49f)

tische Kategorien eingeteilt waren.⁵Drittens bildeten sie das WordNET von Miller (1995) und Fellbaum (1998) als linguistisches Netz ab, das 120.000 Wordformen mit 99.642 Wordbedeutungen in Beziehung setzt⁶.

Obwohl die untersuchten linguistischen Netze sehr unterschiedlich sind, kamen Steyvers & Tenenbaum zum empirischen Ergebnis, dass alle drei Netze die Bedingungen einer "kleinen Welt" mit "skalenfreien Strukturen" erfüllen. "We expect that the general principles of small-world and scale-free structures would be universal, but perhaps we would find quantitative variations in the clustering coefficients or power-law exponents resulting from different language histories." (Steyvers & Tenenbaum 2005, p. 55)

Mit "kleiner Welt" ist gemeint, dass zwischen beliebigen Paaren von Knoten ein Pfad existiert, der über sehr wenige Knoten und Kanten führt. Außerdem haben "kleine-Welt-Netze" einen hohen Clusterkoeffizienten. Das heißt, dass es wahrscheinlich ist, dass Nachbarn eines Knotens auch untereinander verbunden sind.

Eine "skalenfreie Struktur" ist gegeben, wenn die Verteilung der Knotengrade sehr heterogen ist. Dabei ist der Knotengrad durch die Anzahl der Kanten gegeben. Knoten mit vielen Kanten werden "hubs" genannt. Skalenfreie Netze folgen einem Potenzgesetz: Sie enthalten wenige Knoten mit vielen Verbindungen und viele nur schwach verbundene Knoten.

Um zu untersuchen, ob auch GABEK[®]-Netze die Bedingungen für "kleine Welten" und "skalenfreie Strukturen" erfüllen, haben wir (Lösch & Zelger 2013, S. 80-85) mehrere GABEK[®]-Netze analysiert und mit Ergebnissen von Steyvers & Tenenbaum 2005 formal verglichen⁷. Dabei konnte die obige Annahme von

⁵Es handelt sich um einen bipartiten Graph, in dem Wörter und semantische Kategorien zwei unterschiedliche Arten von Knoten bilden. Kanten gibt es nur zwischen Knoten verschiedener Art. Eine Verbindung zwischen einem Wort und einer semantischen Kategorie ist dann gegeben, wenn das Wort in die entsprechende Kategorie fällt. (Steyvers und Tenenbaum 2005, 50). Das bipartite Netz wurde dann umgewandelt in ein Netz, in dem nur Wörter die Knoten bilden. Dabei werden Kanten zwischen zwei Knoten gezeichnet, wenn diese Elemente wenigstens einer gemeinsamen Kategorie sind.

⁶Auch das WordNET wurde zuerst als bipartites Netz konstruiert, in dem Wortformen mit Wortbedeutungen durch Kanten verbunden waren. Später transformierten es die Autoren in ein Netz, das nur eine Art von Knoten enthielt, nämlich Wörter. (nach Steyvers & Tenenbaum, 2005, S. 49-54)

⁷In Lösch & Zelger (2013, S. 83-91) werden GABEK-Netze auch mit dem Ausdrucksnetz von Ferrer I Cancho & Solè (2001) verglichen. Es werden sehr unterschiedliche GABEK-Netze analysiert (Netze, die durch automatische oder händische Einteilung in Texteinheiten entstanden sind, durch automatische oder händische Grundkodierung; Netze mit ausschließlich lexikalischen Begriffen, mit ausschließlich grammatikalischen Knoten, mit lexikalischen und grammatikalischen Knoten; Es wurden Texte mit vorwiegend empirischem Gehalt untersucht, aber auch ein Text, der fast ausschließlich begriffliche Analysen enthält).

Steyvers & Tenenbaum sowohl für kleine als auch für große GABEK®-Netze bestätigt werden.

Um GABEK®-Netze mit dem Assoziations-Netz von Steyvers & Tenenbaumeingehender zu vergleichenverwende ich im folgenden zunächst die verbalen Daten, in denen der Konfliktüber die zwei Arbeitszeitmodelle in drei Niederlassungen der Daimler AG zum Ausdruck kommt. Drei Aussagen wurden in derAbbildung 3-3wiedergegeben. Durch Überlagerung aller verfügbaren verbalenDaten entsteht das Netz der Abbildung 3-6, das nicht nur die Nachbarknoten des Begriffes “Arbeitszeitmodell B” enthält und deren Beziehungen zueinander – wie in Abbildung 3-5– sondern zusätzlich alle weiteren Knoten der verbalen Datenbasis, die den Begriff “Arbeitszeitmodell B” nicht enthalten. Wir analysieren das Netz der Abbildung 3-6 indem wir die formalen Eigenschaften angeben: Die Größe n , die durchschnittliche Anzahl der Verbindungen k , die Anzahl der Kanten E , den Durchmesser des Netzes D , die durchschnittlich Pfadlänge L , den Clusterkoeffizienten C und den Exponenten γ des Potenzgesetzes.

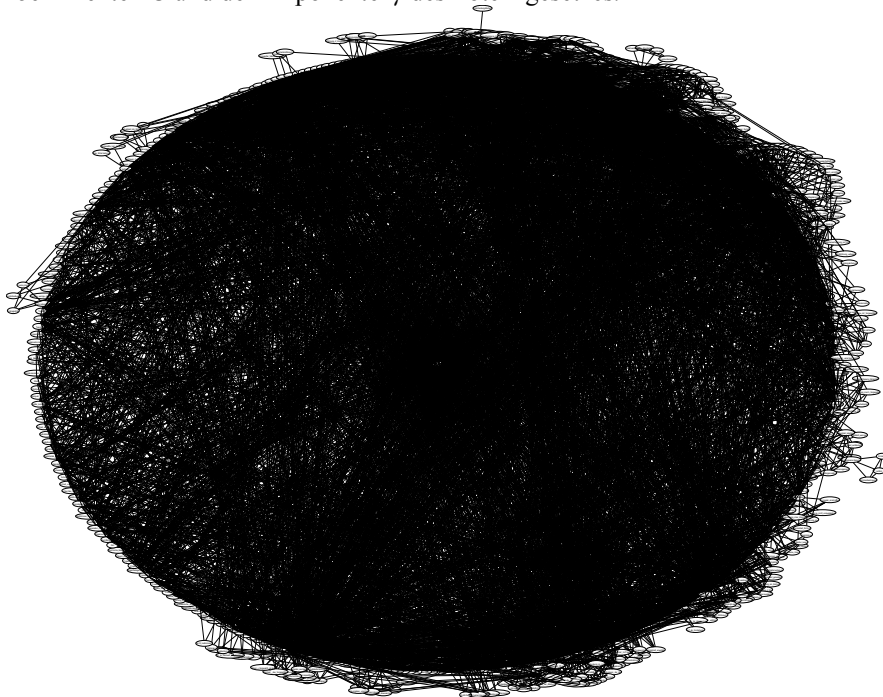


Abbildung 3-6: Das Netz über zwei Arbeitszeitmodelle mit 412 Texteinheiten und 693 Begriffen. Die nicht direkt mit dem Begriff “Arbeitszeitmodell B” verbundenen Begriffe sind wie in einer Sonnencorona um das Netz abgebildet, das als Ausschnitt bereits in Abbildung 20 dargestellt wurde.

Bevor ich aber das Netz der Abbildung 3-6mit dem Assoziationsnetz von Steyvers & Tennenbaum vergleiche, möchte ich auf einen wesentlichen Unterschied der verbalen Daten hinweisen: Steyvers & Tenenbaum verwenden as-

soziativ verbundene Begriffspaare, also bloße Begriffsketten. Mit GABEK[®] werden vollständig verbundene Begriffsgruppen verwendet, die jeweilige Texteinheiten repräsentieren. Damit ist die Basis für Kanten zwischen den Knoten unterschiedlich: Steyvers & Tenenbaum verwenden Assoziationsexperimente, um die begrifflichen Verbindungen zu zeichnen. Bei GABEK[®] ist es entscheidend, dass eine zusammenhängende Menge von Schlüsselbegriffen einer Aussage für eine Person "Sinn macht" (Tabelle 3-1).

	Steyvers & Tenenbaum	GABEK[®]
Einheiten	Begriffspaare	Texteinheiten
Basis für Kanten	Assoziationen in einem Experiment	Subjektive Sinnzusammenhänge
Wachstum des Netzes	Einzelne Begriffe werden hinzugefügt	Texteinheiten werden hinzugefügt
Knoten	Lexikalische Begriffe	Lexikalische Begriffe

Tabelle 3-1: Unterschiede und Gemeinsamkeiten der verglichenen Netze

Wir werden jetzt die Netze nach formalen Attributen vergleichen (Tabelle 3-2). Da unser Projekt über Arbeitszeitmodelle nur eine kleine Datenbasis aufweist, schließe ich in den Vergleich noch ein großes GABEK[®]-Projekt ein, das zur Evaluierung von Planung, Bau und Inbetriebnahme eines riesigen Pumpspeichers für die Vorarlberger Illwerke AG durchgeführt wurde. Die verbalen Daten wurden durch einstündige Interviews mit 58 verantwortlichen Managern erhoben. Die Größe des Netzes (6008 Knoten) ist vergleichbar mit der Größe des semantischen Netzes von Steyvers & Tenenbaum (5018 Knoten).

	Steyvers & Tenenbaum	GABEK[®] Illwerke Project (Österreich)	GABEK[®] Daimler Project (Deutschland)
Anzahl der Knoten n	5018	6008	712
Durchschnittliche Anzahl der Verbindungen k	22.0	16.9	13.5
Anzahl der Kanten E	55,198	50,767	4806
Durchmesser des Netzes D	5	5	5
Durchschnittliche minimale Pfadlänge L	3.04	2.90	2.55
Clusterkoeffizient C	0.186	0.774	0.784
Exponent γ für das	-3.01	-1.227	-1.041

Potenzgesetz der Verteilung der Knotengrade				
--	--	--	--	--

Tabelle 3-2: Formale Attribute zweier GABEK®-Projekte, die händisch kodiert wurden im Vergleich mit dem Assoziationsnetz von Steyvers & Tenenbaum (2005, S. 49-55)⁸

Anhand der Tabelle 3-2 sehen wir, dass beide GABEK®-Projekte die Bedingungen für "kleine Welten" erfüllen. Obwohl sie in ihrer Größe sehr unterschiedlich sind, so ist doch deren Durchmesser D und deren durchschnittliche Pfadlänge L klein und deren Clusterkoeffizient C groß.

Die formalen Attribute der beiden GABEK®-Projekte zeigen außerdem, dass auch die Bedingungen für skalenfreie Strukturen erfüllt werden: So ist die durchschnittliche Anzahl der Verbindungen k niedrig und es gibt einen negativen Exponenten γ für das Potenzgesetz, dem die Verteilung der Knotengrade für die lexikalischen Ausdrücke folgt (Abbildung 3-7).

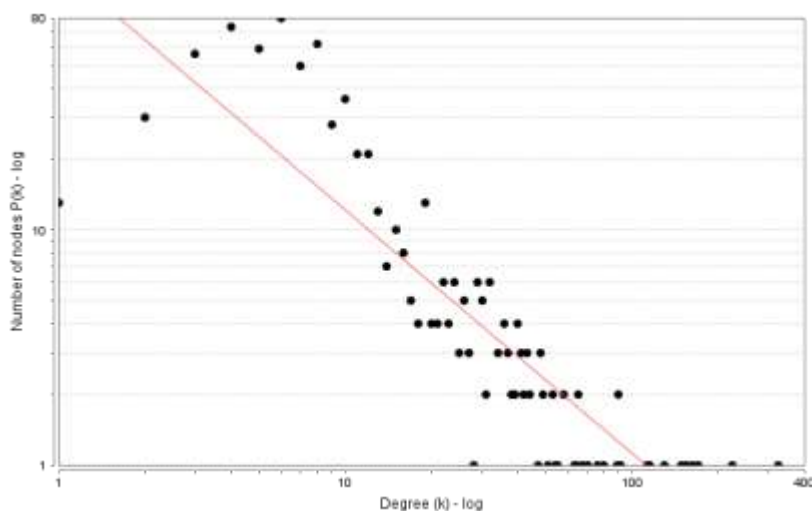


Abbildung 3-7: Der Exponent γ für das Potenzgesetz der Verteilung der Knotengrade beim Daimler Projekt⁹

Wenn wir die GABEK®-Netze mit dem Assoziationsnetz von Steyvers & Tenenbaum vergleichen, so fällt der niedrigere Clusterkoeffizient bei Steyvers & Tenenbaum auf sowie der Exponent γ , der eine steilere Neigung der Geraden ergibt, die die Verteilung der Knotengrade beschreibt. Ich nehme an, dass dies eine Folge der Methode der Datensammlung ist: Steyvers & Tenenbaum protokollieren Begriffssequenzen in einem Assoziationsexperiment. Das hat zur Folge,

⁸Auszug aus Lösch & Zelger (2013, S. 84.)

⁹Ich danke Horst Lösch für die Berechnung der Graphik.

dass immer nur ein neues Wort zum Netz hinzugefügt wird. Bei GABEK® hingegen werden Sätze aneinandergereiht, die durch kleine Begriffsnetze repräsentiert werden. Damit wird zu einem großen Begriffsnetz nicht ein einzelner Begriff sondern eine Begriffsmenge mit drei bis neun Begriffen hinzugefügt, die einen Satz repräsentiert. Dies erklärt den höheren Clusterkoeffizienten und auch die weniger steile Gerade γ . Denn mit jedem neuen Satz werden auch Kanten zwischen Nachbarn eines zentralen Begriffes zum Netz hinzugefügt. Es ist eine Folge der Entscheidung, dass für linguistische GABEK®-Netze Aussagen als elementare Elemente aufgefasst wurden, die für wenigstens eine Person "Sinn machen" und nicht Einzelbegriffe. In der Tradition von Wittgenstein habe ich nämlich Sätze als sprachliche Grundelemente aufgefasst und nicht Begriffe. Diese Grundentscheidung wird uns noch zu wesentlichen Überlegungen führen.

3.3 Das Wachstum semantischer Netze

Da die Größe von Netzen meistens durch die Anzahl der Knoten angegeben wird, denken die Autoren, die sich mit dem Wachstum linguistischer Netze befassen, zunächst an einzelne Begriffe, die zu einem Netz hinzugefügt werden. So auch Barabási & Albert (1999), auf die sich die meisten Autoren beziehen. Barabási & Albert (1999) haben das Prinzip des *preferential attachment* vertreten. Es besagt, dass neue Begriffe mit schon bestehenden in einem linguistischen Netz verbunden werden, mit einer Wahrscheinlichkeit, die proportional ist zum Knotengrad der schon bestehenden Begriffe. Das heißt, dass neue Begriffe vor allem mit jenen Begriffen im Netz verbunden werden, die bereits viele Verbindungen haben.

Nach Ferrer i Cancho & Solé (2001, p. 2264) folgt aus dem Prinzip des *preferential attachment* die skalenfreie Struktur linguistischer Netze. Außerdem kann mit diesem Prinzip erklärt werden, dass es in jeder Sprache einen Kernbereich von Begriffen gibt, die von fast allen Personen, die sie sprechen, verstanden werden.

Auch Steyvers & Tenenbaum bestätigen, dass das Prinzip des *preferential attachment* die Verteilung der Knotengrade nach einem Potenzgesetz zur Folge hat. Sie weisen aber auch darauf hin, dass dieses Prinzip allein nicht zu Strukturen einer "kleinen Welt" führt. "However, it does not yield clustering coefficients that are nearly as high as those we observed." (2005, p. 56) Eine angemessene Theorie über das Wachstum von Netzwerken müsste aber auf jeden Fall beides erklären: die "skalenfreie" und die "kleine Welt"-Struktur. Daher ersetzen Steyvers & Tenenbaum das *preferential attachment* durch drei Prinzipien, die sie wie folgt zusammenfassen: "*Specifically, we assume that when a new node is added to the network, it differentiates an existing node by acquiring a pattern of connections that corresponds to a subset of the existing node's connections. Secondly, we assume that the probability of differentiating a particular node at each time step is*