

Fachrichtung 4.7 Computerlinguistik  
Universität des Saarlandes

# Ähnlichkeitsmaße für FrameNet

Diplomarbeit

Michael Wirth

Angefertigt unter der Leitung von  
Prof. Dr. Manfred Pinkal und  
Dr. Aljoscha Burchardt



# Erklärung

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Saarbrücken, den 7. April 2009

Michael Wirth

## Zusammenfassung

Untersuchungsgegenstand dieser Arbeit ist die Konzeptähnlichkeit von Frames, wie sie in der Berkley FrameNet-Datenbank erfasst sind. Frames repräsentieren außersprachliche Konzepte und liefern eine Abstraktion eines zu analysierenden Textes auf Prädikat-Argument-Ebene. Um solche Abstraktionen in Systemen zur Verarbeitung natürlicher Sprache sinnvoll nutzen zu können, ist es wichtig, dass verschiedene Abstraktionen in Beziehung gesetzt und miteinander verglichen werden können. Bezogen auf Frames untersuche ich in dieser Arbeit deshalb die Modellierung von *Frame-Ähnlichkeit*.

Zuerst werde ich mit Hilfe eines Experimentes zeigen, dass der intuitive Begriff von Frame-Ähnlichkeit empirisch nachweisbar ist. Anschließend wird ein automatisches Frame-Ähnlichkeitsmaß für die FrameNet-Datenbank entwickelt. Hierbei werden sowohl bereits für ähnliche Ressourcen vorhandene Ähnlichkeitsmaße adaptiert, als auch neue ausschließlich auf FrameNet basierende Maße entwickelt.

In dieser Diplomarbeit wird der empirische Nachweis des Phänomens „Frame-Ähnlichkeit“ erbracht. Es werden zwei Gold-Standard Frame-Sortierungen erarbeitet. Ausserdem wird eine erste Definition von „Frame-Ähnlichkeit“ gegeben. Des Weiteren wird ein robustes automatisches Frame-Ähnlichkeitsmaß entwickelt.

## Danksagung

Ich möchte Prof. Dr. Manfred Pinkal danken für die Betreuung meiner Diplomarbeit und dafür, dass er es mir ermöglichte, meine Arbeit in einem vitalen und motivierenden Umfeld zu schreiben.

Auch möchte ich Dr. Aljoscha Burchardt danken, der mich zu meinem Diplomthema geführt hat und mich trotz des Wechsels der Arbeitsstelle weiterhin betreut hat.

Ich danke Dr. Marco Pennacchiotti für seine Begeisterung und für die Unterstützung bei dem praktischen Teil der Arbeit.

Herzlichen Dank an Ines Rehbein und Dr. Josef Ruppenhofer für die Betreuung des schriftlichen Teils der Arbeit, für inspirierende Gespräche und für die konstruktive Kritik.

# Inhaltsverzeichnis

<b>1. Einleitung</b>	<b>1</b>
1.1. Motivation . . . . .	1
1.2. Semantische Ähnlichkeit . . . . .	1
1.3. Struktur der Arbeit . . . . .	3
<b>2. Ressourcen</b>	<b>4</b>
2.1. FrameNet . . . . .	4
2.1.1. Wissenschaftlicher Hintergrund . . . . .	5
2.1.2. Struktur . . . . .	6
2.1.3. Aktuelle Forschung . . . . .	10
2.2. WordNet . . . . .	11
<b>3. Wissenschaftliches Umfeld</b>	<b>14</b>
3.1. Systeme zur automatischen Erkennung von semantischer Ähnlichkeit . . . . .	14
3.1.1. Wörterbuch-basierte Systeme . . . . .	14
3.1.2. Thesaurus-basierte Systeme . . . . .	15
3.1.3. Systeme basierend auf semantischen Netzen . . . . .	15
3.1.4. Hybride Systeme . . . . .	19
3.2. Methodik zur Evaluation von Systemen zur automatischen Berechnung semantischer Ähnlichkeit . . . . .	23
3.3. Einordnung der Fragestellung ins wissenschaftliche Umfeld . . . . .	24
<b>4. Frame-Ähnlichkeit</b>	<b>26</b>
4.1. Ist eine Definition semantischer Ähnlichkeit für FN-Frames möglich? . . . . .	26
4.1.1. Pilotstudie . . . . .	26
4.1.2. Evaluation . . . . .	27
4.2. Möglichkeiten einer Definition semantischer Ähnlichkeit auf der Basis von FrameNet . . . . .	29
4.2.1. Arbeitsdefinition . . . . .	31
<b>5. Entwicklung eines FN basierten Ähnlichkeitsmaßes</b>	<b>32</b>
5.1. Adaption verschiedener Ähnlichkeitsmaße . . . . .	33
5.1.1. Einfache Pfadlänge . . . . .	33
5.1.2. Skalierte Pfadlänge . . . . .	35
5.1.3. Hybrider graph-basierter Ansatz . . . . .	36
5.2. FN-Ähnlichkeitsmaße . . . . .	37
5.2.1. Frame-Name . . . . .	38

5.2.2. Frame-Definition . . . . .	38
5.2.3. Frame Elemente . . . . .	39
5.2.4. Lexical Units . . . . .	41
5.3. Zusammenfassung . . . . .	42
<b>6. Evaluation</b>	<b>44</b>
6.1. Leistung der Ähnlichkeitsmaße . . . . .	44
6.1.1. Experimentdesign und Ergebnisse . . . . .	44
6.1.2. Diskussion der Resultate . . . . .	47
6.2. Qualitative Untersuchung . . . . .	48
6.2.1. Struktur-basierte Ähnlichkeitsmaße . . . . .	48
6.2.2. Nicht-struktur-basierte Maße . . . . .	61
6.3. Zusammenfassung . . . . .	66
<b>7. Zusammenfassung und Ausblick</b>	<b>68</b>
7.1. Erreichte Ziele . . . . .	68
7.2. Zukünftige Arbeit . . . . .	69
<b>A. Gold Standards</b>	<b>72</b>
A.1. $GS_{control}$ . . . . .	72
A.2. $GS_{random}$ . . . . .	76
<b>Literaturverzeichnis</b>	<b>80</b>

# Abbildungsverzeichnis

2.1. „Crime_scenario“ Frame mit Verknüpfungen . . . . .	5
2.2. Beispiel Frame aus FrameNet . . . . .	6
2.3. Causative_of - Inchoative-Kette . . . . .	9
2.4. Pfad Change_of_state_initial_state - Emptying . . . . .	11
2.5. Bedeutungen „car“ . . . . .	12
2.6. Hyperonyme „dolphin“ (Auszug) . . . . .	13
3.1. Gültige Pfadmuster (Hirst und St-Onge, 1998) . . . . .	17
3.2. Konzept - Ähnlichkeit nach Wu und Palmer (1994) . . . . .	19
5.1. Gültige Pfadmuster (Hirst und St-Onge, 1998) . . . . .	34
5.2. Beispiel Pfadmuster h) . . . . .	34
5.3. Nicht akzeptierter Pfad . . . . .	35
5.4. Frame-Definitionen CHANGE_OF_PHASE und ALTERED_PHASE . . . . .	39
5.5. FE-zu-FE-Relationen . . . . .	40
5.6. LUs TEMPERATURE-Frame . . . . .	41
5.7. LUs TEMPERATURE - AMBIENT_TEMPERATURE . . . . .	42
6.1. Architektur Versuchssystem . . . . .	46
6.2. Pfad Knot_creation - Rope_manipulation . . . . .	50
6.3. Pfad Religious_belief - Certainty . . . . .	50
6.4. Beispiel Pfad Commerce_sell - Reciprocity . . . . .	52
6.5. Beispiel Pfadmuster e) . . . . .	53
6.6. Pfadmuster $GS_{control}$ . . . . .	55
6.7. Pfadmuster $GS_{random}$ . . . . .	56
6.8. Gültige Pfadmuster (Hirst und St-Onge, 1998) . . . . .	56
6.9. Beispiel FE-Overlap . . . . .	59
7.1. FATE Text-Hypothese-Paar (FATE T/H-Paar 215) . . . . .	71

# Tabellenverzeichnis

2.1. FN-Hierarchie (Release 1.3) in Zahlen . . . . .	10
2.2. Relationen in WordNet . . . . .	13
4.1. Ergebnisse Wort-Ähnlichkeitsexperimente . . . . .	30
6.1. Korrelation der Ähnlichkeitsmaße mit den Gold-Standards . . . . .	45
6.2. Prozentuale Verteilung der Frame-Relationen in FN . . . . .	49
6.3. Durchschnittliche Anzahl unterschiedlicher Relationen pro Pfad . . . . .	51
6.4. Anzahl der Vorkommen der Pfadmuster in den Gold-Standards . . . . .	53
6.5. Anzahl verbundener FE pro Relation in Prozent für FN gesamt . . . . .	58
6.6. Prozentuale Verteilung der Relationen in den TOP30 Frame-Paaren . . . . .	60
6.7. Anzahl der Frame-Paare mit einer Ähnlichkeit $> 0$ . . . . .	62
6.8. TOP 5 Frame-Paare Sortierung $FP_{control}$ . . . . .	62
6.9. TOP 5 Frame-Paare Sortierung $FP_{random}$ . . . . .	63
6.10. Korrelationswerte LU-basierte Ähnlichkeitsmaße . . . . .	64
6.11. Frame-Name-Distanz-Maß - Jaro-Winkler Werte der 5 besten Frame-Paare . . . . .	65
7.1. Frame-Ähnlichkeiten . . . . .	70

# 1. Einleitung

## 1.1. Motivation

Zugang zu Information aus elektronisch vorliegendem, sprachlichen Datenmaterial gewinnt immer mehr an Bedeutung. Eine Möglichkeit Information zugänglich zu machen, ist die Erschließung der Kernbedeutung der Sprachdaten. Dies geschieht meist durch Abstraktion des zu Grunde liegenden Datenmaterials. *Konzepte* haben sich als gute Abstraktionsebene erwiesen. Solche Konzepte können beispielsweise die Bedeutung von „Kauf“- und „Verkauf“-Situationen umschreiben. Durch die Repräsentation von Sprachdaten durch Konzepte wird der Zugang zu deren Bedeutung jedoch weder erschlossen, noch wird der Zugang zur enthaltenen Information sehr erleichtert. Erst wenn die Konzepte organisiert werden können, kann die Information in einer sinnvoll strukturierten Form angeboten werden. Wichtig sind vor allem die Beziehungen der Konzepte untereinander, vor allem eine Ähnlichkeitsrelation.

Um Sprachdaten automatisch mit Konzepten zu versehen, benötigt man elektronische Lexika. Diese Lexika stellen eine Abbildung von Wörtern auf Konzepte zur Verfügung. Für die Nützlichkeit eines solchen Lexikons ist es von großer Bedeutung, dass die Konzepte strukturiert sind. Mit strukturierten Konzepten können eine Vielzahl von Aufgabenstellungen behandelt werden. So können Texte erschlossen und zusammengefasst werden. Des Weiteren können Konzepte in Information Extraction-, Information Retrieval- und Question Answering- Systemen genutzt werden oder in Systemen zur Modellierung natürlichsprachlicher Folgerungsbeziehungen.

Ziel dieser Arbeit ist die Entwicklung eines Konzept-Ähnlichkeitsmaßes für das elektronische Wörterbuch FrameNet. FrameNet bietet sowohl eine Abbildung von Wörtern auf Konzepte, als auch eine Struktur, in der diese Konzepte organisiert sind. Für einen breiten Einsatz in sprachverarbeitenden Systemen fehlt allerdings noch ein robustes Ähnlichkeitsmaß.

In Abschnitt 1.2 dieses Kapitels wird der Begriff „semantische Ähnlichkeit“ geklärt. Der letzte Abschnitt präsentiert die Struktur der hier vorliegenden Arbeit.

## 1.2. Semantische Ähnlichkeit

Der Gegenstand dieser Arbeit ist die Entwicklung eines Maßes für die semantische Ähnlichkeit von Konzepten. Trotz einer langen Phase wissenschaftlicher Forschung gibt es keine konkrete unstrittige Definition von semantischer Ähnlichkeit. Es ist also notwendig, zu Beginn der Arbeit zu klären, welchem Ansatz wir folgen wollen. Drei Begriffe sind in der wissenschaftlichen Diskussion dominant: *Semantische Gleichheit*, *semanti-*

*sche Ähnlichkeit* und *semantische Distanz*. Zuerst soll geklärt werden, wie diese Begriffe miteinander in Beziehung stehen. In einem zweiten Schritt sollen die verschiedenen Begriffe expliziert werden.

Philip Resnik grenzt in seinem Artikel "Using Information Content to Evaluate Semantic Similarity in a Taxonomy" Resnik (1995) die drei Begriffe voneinander ab. Mit folgendem Beispiel verdeutlicht er den Unterschied zwischen *semantischer Gleichheit* und *semantischer Ähnlichkeit*:

„[...]cars and gasoline would seem to be more closely related than, say, cars and bicycles, but the latter pair are certainly more similar.” Resnik (1995, Seite 1)

(deutsch: „[...] Autos und Benzin scheinen näher verwandt zu sein, als ,sagen wir, Autos und Fahrräder, aber das letztere Paar ist sich sicherlich ähnlicher.“)

Budanitsky und Hirst Budanitsky und Hirst (2001) explizieren weiter, dass semantische Gleichheit ein Sonderfall von semantischer Ähnlichkeit ist.

“It’s important to note that semantic relatedness is a more general concept than similarity; similar entities are usually assumed to be related by virtue of their likeness” Budanitsky und Hirst (2001, Seite 1)

(deutsch: „Es ist wichtig anzumerken, dass semantische Ähnlichkeit ein allgemeineres Konzept ist, als Gleichheit; man nimmt normalerweise an, dass gleiche Entitäten aufgrund ihrer Ähnlichkeit verwandt sind.“)

In ihrem Artikel „Evaluating WordNet-based measures of lexical ambiguity” Budanitsky und Hirst (2006) führen die Autoren weiter aus:

“Among other relationships that the notion of relatedness encompasses are the various kinds of meronymy, antonymy, functional association, and other „non-classical relations”(Morris und Hirst, 2004)” Budanitsky und Hirst (2006, Seite 2)

(deutsch: Unter den Beziehungen, die unter dem Begriff Ähnlichkeit zusammengefasst werden, sind unter anderem die verschiedenen Arten von Meronymie, Antonymie, funktionale Assoziierung und andere „nicht-klassische Relationen“)

Demzufolge sind auch Antonyme semantisch verwandt. Sie sind jedoch nicht semantisch gleich, sondern gerade das Gegenteil. In dieser Arbeit wollen wir den hier vorgestellten Begriff von semantischer Ähnlichkeit benutzen. *Semantische Distanz* verstehen wir, Budanitsky und Hirst (2001) folgend, als das Inverse der semantischen Ähnlichkeit.

Abschließend können die Begriffe *Semantische Gleichheit*, *Semantische Ähnlichkeit* und *Semantische Distanz* wie folgt für diese Arbeit definiert werden:

- Semantische Gleichheit beschreibt identische Bedeutungen.
- Semantische Ähnlichkeit ist ein allgemeineres Konzept, als Gleichheit.
- Semantische Distanz ist das Inverse zu semantischer Ähnlichkeit.

### 1.3. Struktur der Arbeit

In Kapitel 2 werden die in dieser Arbeit benutzten Ressourcen eingeführt. Bei der Vorstellung von FrameNet wird ausführlich auf die Struktur der Ressource eingegangen. WordNet wird nur kurz vorgestellt, um eine grundlegende Vorstellung der Struktur der Ressource zu vermitteln und als Grundlage für die in den folgenden Kapiteln besprochenen WordNet-Ähnlichkeitsmaße.

Kapitel 3 gibt einen Überblick über das wissenschaftliche Umfeld. Es werden verschiedene Ansätze zur Berechnung semantischer Ähnlichkeit vorgestellt. Des Weiteren wird eine verbreitete Methode zur Evaluation von automatischen Ähnlichkeitsmaßen vorgestellt. Im letzten Abschnitt des Kapitels ordne ich die in dieser Diplomarbeit behandelten Fragen nach einem Ähnlichkeitsmaß für FrameNet, in das wissenschaftliche Umfeld ein.

In Kapitel 4 wird der intuitive Begriff von Konzept-Ähnlichkeit für FrameNet experimentell nachgewiesen und eine erste Definition des Ähnlichkeitsbegriffs gegeben. Im Weiteren werden die Experimentresultate zur Validierung eines Gold-Standards genutzt.

Kapitel 5 präsentiert verschiedene Maße zur Berechnung der Konzept-Ähnlichkeit für FrameNet. Es werden insgesamt 10 verschiedene Ähnlichkeitsmaße vorgestellt, darunter sowohl adaptierte Maße, als auch von mir speziell für FrameNet entwickelte Maße.

In Kapitel 6 werden die verschiedenen Ähnlichkeitsmaße evaluiert. Untersucht werden, unter anderem, das Verhalten der Ähnlichkeitsmaßen in unterschiedlich dichten Teilen der FrameNet-Struktur und der Unterschied zwischen struktur-basierten und nicht struktur-basierten Maßen.

In Kapitel 7 werden die Ergebnisse der Diplomarbeit zusammengefasst und eine Anwendungsmöglichkeit für ein FrameNet-basiertes Konzept-Ähnlichkeitsmaß skizziert.

## 2. Ressourcen

Kapitel 2 führt die in meiner Diplomarbeit genutzten lexikalischen Ressourcen ein. In Abschnitt 2.1 wird FrameNet ausführlich vorgestellt. Abschnitt 2.2 stellt WordNet vor. Die Kurzvorstellung von WordNet dient als Grundlage für die in Kapitel 3.1.3 vorgestellten Ähnlichkeitsmaße, die ich für FrameNet adaptiert habe (siehe Kapitel 5.1).

### 2.1. FrameNet

Um gemäß der Definition von semantischer Ähnlichkeit automatisch einen Ähnlichkeitswert berechnen zu können, benötigen wir eine entsprechende Ressource. In dieser Arbeit wird FrameNet als Resource genutzt.

*FrameNet* (kurz: FN) ist ein lexikographisches Projekt der Universität Berkeley (Baker u. a., 1998). Ziel dieses Projektes ist es, eine umfassende semantische Online-Ressource für das Gegenwarts-Englisch zu erstellen. Die theoretische Grundlage hierfür bildet die Frame-Semantik (Kap. 2.1.1 auf der nächsten Seite). FN enthält in der aktuellen Version 1.3 rund 135.000 annotierte Sätze in 795 hierarchisch organisierten Frames.

FrameNet bietet Zugang zu semantischer Information auf Ebene der Prädikat-Argument-Struktur. Mit der Prädikat-Argument-Struktur kann die Kernbedeutung von Sätzen modelliert werden. Ich benutze FrameNet, da es sich besonders gut für ein Konzept-Ähnlichkeitsmaß auf Prädikat-Argument-Struktur-Ebene eignet. Chang u. a. (2002) arbeiten die Vorteile von FrameNet gegenüber anderen verfügbaren Ressourcen heraus. So bietet FrameNet nicht nur eine Sammlung von Prädikat-Argument-Strukturen wie das PropBank Projekt (Kingsbury und Palmer, 2002), sondern die Prädikat-Argument-Strukturen sind in einem semantischen Netz organisiert. Dieses semantische Netz bildet nicht nur taxonomische Beziehungen wie beispielsweise Hyperonymie, Synonymie, Meronymie oder Antonymie ab, wie bei WordNet (Fellbaum, 1998), sondern liefert auch Informationen über kontextuelle Hintergründe.

Auf der Basis dieser hierarchisch organisierten Frames, wird im Verlauf dieser Arbeit ein generisches Maß für semantische Ähnlichkeit entwickelt. Semantisch ähnlich sollen dabei Paare von Frames sein, die über die FrameNet-Hierarchie verbunden sind. Der Grad der Ähnlichkeit wird dann durch die Menge an geteilter Information bestimmt. Dieser Ansatz ist ein hybrider Ansatz. Er hebt sich von distanz- und vektor-basierten Modellen ab, da semantische Ähnlichkeit weder allein über die Länge des Pfades, noch allein durch die Lage von numerischen Vektoren zueinander dargestellt wird. Für eine genaue Beschreibung der Maße siehe Kapitel 5.2 auf Seite 37.

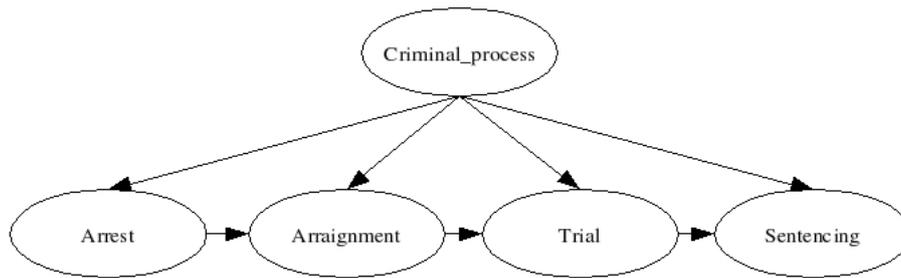


Abbildung 2.1.: „Crime\_scenario“ Frame mit Verknüpfungen

### 2.1.1. Wissenschaftlicher Hintergrund

In seinem 1977er Artikel „Scenes-and-frames semantics“ entwickelt Charles Fillmore die *Frame Semantik* als eine Art Prototypen-Semantik. *Frames* sind der zentrale Baustein dieser Semantik. Sie beschreiben prototypische Situationen oder Aktionen. Die an der Situation oder Aktion beteiligten Entitäten werden durch *Frame-Elemente* beschrieben. So wird mit dem Frame „Sleep“ der Zustand des Schlafens beschrieben und der Schlafende durch das Frame-Element „Sleeper“.

Im Laufe des Spracherwerbs werden Frames verfeinert und miteinander verknüpft. Es entstehen neue Frames und linguistisches Material kann Frames neu zugeordnet werden. Ein gutes Beispiel für Verknüpfungen von Frames ist der Frame „Criminal\_process“ mit seinen Unterframes in Abbildung 2.1. Hier ist zu erkennen, dass sich die komplexe Situation „Criminal\_process“ (dt.: Strafprozeß) untergliedert in die Situationen „Arrest“ (dt.: Verhaftung), „Arraignment“ (dt.: Anklageerhebung), „Trial“ (dt.: Verhandlung), „Sentencing“ (dt.: Verurteilung). Die Konzepte sind auch zeitlich geordnet, die Verhaftung geht der Anklageerhebung, die Anklageerhebung der Verhandlung und die Verhandlung der Verurteilung voraus. Die Pfeile stellen Relationen dar. Diese Relationen zwischen den Frames erlauben es, semantische Ähnlichkeit zwischen Frames zu definieren. Fillmore schreibt dazu:

„[...] frames are associated in memory with other frames by virtue of shared linguistic material, and that scenes are associated with other scenes by virtue of sameness or similarity of the entities or relations or substances in them or their contexts of occurrence.“ Fillmore (1977, Seite 63)

(deutsch: „[...] Frames im Gedächtnis untereinander verbunden sind durch geteiltes linguistisches Material, und dass Szenen mit anderen Szenen verbunden sind durch die Gleichheit oder Ähnlichkeit der Entitäten oder Relationen oder Substanzen, in dem oder in deren Kontext sie vorkommen.)

Die Beschreibung der komplexen Situation „Strafprozeß“ enthält das Frame-Element

Frame: Abusing	
In this frame an ABUSER repeatedly treats the VICTIM in a cruel and violent way, including physically harming or forcing the VICTIM to engage in sexual activity against their will. The VICTIM usually lacks the power to resist or fight back.	
⊞ ABUSER	<b>Fred</b> <u>abused</u> Martha by locking her into a closet .
⊞ VICTIM	Fred <u>abused</u> <b>Martha</b> by locking her into a closet .
LU	abuse.n, abuse.v, abusive.a, batter.v, domestic violence.n, maltreat.v, maltreatment.n

Abbildung 2.2.: Beispiel Frame aus FrameNet

“Court” (dt.: Gericht). Das gleiche Frame-Element findet sich sowohl im Frame „Verhandlung“, als auch im Frame „Verurteilung“. Die Situationen sind also miteinander verbunden, sind sich ähnlich. Sie teilen sich auch die Rolle “Place” (dt.: Ort), was darauf schließen läßt, dass die Situationen an gleicher Stelle stattfinden.

Im folgenden Abschnitt wird die Struktur von FN detailliert vorgestellt.

## 2.1.2. Struktur

### 2.1.2.1. Frames

In Frame-Semantik beschreiben Frames außersprachliche Situationen oder Szenen (siehe Kapitel 2.1.1). Die schematische Darstellung eines Frames in der FN-Datenbank enthält den Frame-Namen, die Definition des Frames, die beteiligten Rollen - Frame-Elemente genannt -, Informationen über die Verbindung zu anderen Frames - kodiert in Frame-zu-Frame-Relationen - und Wörter, die einem Frame zugeordnet sind. Diese Wörter werden Lexical Unit (kurz: LU) genannt.

Neben der semantischen Ebene werden auch syntaktische Informationen durch FrameNet erfasst (Ruppenhofer u. a., 2006, Kapitel 3). Es werden der phrasale Typ und die grammatikalische Funktion zur frame-evozierenden LU annotiert. Abbildung 2.2 zeigt ein Beispiel für einen FN-Eintrag.

Frame-Namen sind eindeutige, informelle Bezeichner für Frames (Fillmore u. a., 2003). Der Frame *Abusing* (dt.: mißbrauchen; Abbildung 2.2) beispielsweise beschreibt eine Missbrauchssituation. Die Definition umschreibt die zugrunde liegende Situation so abstrakt wie möglich und so konkret wie nötig. Die an der Situation beteiligten Rollen werden durch Frame-Elemente (kurz: FE) ausgedrückt. Für *Abusing* sind dies die FEs *ABUSER* und *VICTIM*. Die Benennung der einzelnen Frame-Elemente erfolgt nicht durch Rückgriff auf eine kleine Menge universeller Rollen, sondern beschreibt bedeutungsvoll die einzelnen Aktanten des Frames.

Gemäß den theoretischen Vorgaben (siehe Kapitel 2.1.1) sollten Frames im Gedächtnis untereinander verbunden sein. In FrameNet sind sie durch die so genannten Frame-

zu-Frame-Relationen und die einzelnen FE untereinander durch FE-zu-FE-Relationen verbunden. Es gibt sieben verschiedene Frame-zu-Frame Relationen: Inheritance, Using, Subframe, Precedes, Perspective\_on, Causative\_of, Inchoative\_of, See\_also.

### 2.1.2.2. Frame-zu-Frame-Relationen

Frame-zu-Frame-Relationen drücken Generalisierungen über Frames aus. Die verbundenen Frames bilden ein semantisches Netz, beziehungsweise einen semantischen Raum. Dieser semantische Raum erlaubt es, ähnliche Frames bzw. FEs zu suchen oder zu inferieren.

„We intend, that the Frame Relations, Frame Element mappings, and Lexical Units provide a significant amount of information for doing inferencing on natural language. We don't provide every kind of information that an inferencing engine might require [...], but the frame-to-frame relations should provide a good basis for further specifications [...]” (Ruppenhofer u. a., 2006, Seite 125)

(deutsch: Wir beabsichtigen, dass Frame-Relationen, Frame-Element-Abbildungen und Lexical Units eine signifikante Menge an Informationen zum Inferieren über natürliche Sprache zur Verfügung stellen. Wir stellen nicht jede Art von Information zur Verfügung, die eine Inferenzmaschine eventuell benötigt [...], aber die Frame-zu-Frame-Relationen sollten eine gute Basis für weitere Ausarbeitungen bilden [...])

In dieser Arbeit wird mit dem Begriff *FN-Hierarchie*, auf den semantischen Raum referiert. Frame-zu-Frame-Relationen sind gerichtet. Sie verbinden einen Sub-Frame mit einem Super-Frame. BEISPIEL Die einzelnen Frame-zu-Frame-Relationen - Inheritance, Using, Subframe, Precedes, Perspective\_on, Inchoative\_of, Causative\_of und See\_also - sind, wie folgt, definiert:

- **INHERITANCE:** Die Inheritance-Relation ist die stärkste Frame-zu-Frame-Relation. Alles was für den Super-Frame gilt, muss auch für den Sub-Frame gelten. Inheritance bezeichnet eine vollständige und monotone Relation mit der Möglichkeit der Mehrfachvererbung. Dies bedeutet, dass jedes FE im Sub-Frame ein korrespondierendes FE im Super-Frame haben muss und dass der semantische Typ des FE des Sub-Frames mit dem semantischen Typ des Super-Frame identisch sein muss. Mehrfachvererbung bedeutet, dass ein Sub-Frame mehr als einen Super-Frame haben kann. Die FEs des Sub-Frames können also aus mehreren Super-Frames stammen.
- **USING:** Die Using-Relation ist eine Abschwächung der Inheritance-Relation. Sie ist monoton mit der Möglichkeit der Mehrfachvererbung, aber sie ist nicht vollständig. Es hat also nicht jedes FE des Sub-Frames eine Entsprechung in einem Super-Frame. Inhaltlich gleicht die Using-Relation der Perspective\_on-Relation.

- SUBFRAME: Frames können komplexe Situationen beschreiben, die sich wiederum aus einfacheren Situationen zusammensetzen. Die Beziehung zwischen solch einem komplexen Frame und einer einfacheren Teilsituation wird als Subframe-Relation bezeichnet. Meist sind die FE des Sub-Frames mit FEs im Super-Frame verbunden. So kann beispielsweise über eine Kette von FE-zu-FE Relationen die *Suspect* Rolle des Arrest-Frames mit der *Defendant*-Rolle des *Verdict*-Frames verbunden werden. Dies bedeutet, da Arrest und *Verdict* Subframes des Frames *Criminal\_process* sind, dass beide Rollen auf den gleichen Aktanten verweisen. Sollte ein komplexer Frame B, der Subframes besitzt, von einem Frame A erben, so müssen die Subframes von B gemäß der Vollständigkeit der Inheritance-Relation auch Subframes des Super-Frames sein.
- PRECEDES: Subframes können auch zeitlich geordnet werden. Diese Ordnung stellt die Precedes-Relation her. Sie tritt nur zwischen Subframes eines komplexen Frames auf.
- PERSPECTIVE\_ON: Die Perspective-on-Relation ist eine Erweiterung der Using-Relation. Perspective\_on-Relationen verbinden mindestens zwei Frames, die die selbe Situation aus unterschiedlichen Sichtweisen beschreiben, mit einem neutralen Frame. Der neutrale Super-Frame sollte keine LUs enthalten, sondern ausschließlich die Beschreibung der abstrakten Situation bereitstellen, die aus verschiedenen Blickwinkeln betrachtet wird. Der neutrale Frame darf nicht selbst Sub-Frame einer Perspective\_on Relation sein.
- INCHOATIVE\_OF / CAUSATIVE\_OF: In FrameNet wird zwischen stativen, inchoativen und kausativen Frames unterschieden. Um den systematischen Zusammenhang zwischen solchen Frames ausdrücken zu können, gibt es die Relationen Inchoative\_of und Causative\_of. Oft bilden solche Frames eine Kette vom kausativen über den inchoativen zum stativen Frame. Ein gutes Beispiel für solch eine Kette von Frames ist: Cause\_temperature\_change -> Inchoative\_change\_of\_temperature -> Temperature (siehe Abbildung 2.3 auf der nächsten Seite)

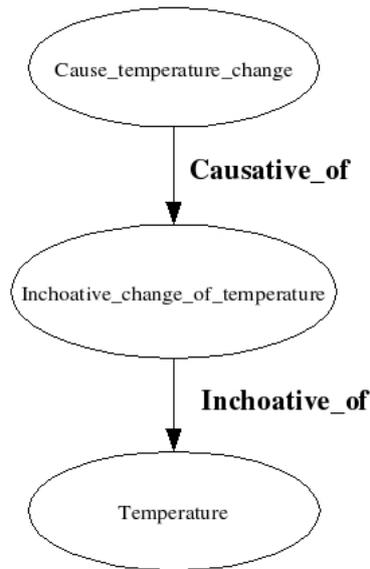


Abbildung 2.3.: Causative\_of - Inchoative-Kette

- **SEE\_ALSO**: Eng verwandte Frames werden durch die See\_also-Relation miteinander verbunden. Formal definierte Bedingungen, die für die anderen Relationen, müssen hier nicht erfüllt sein. Frames, die sich nur durch Bedeutungsnuancen voneinander unterscheiden, werden durch die See\_also-Relation verbunden. Die Unterschiede zwischen den Bedeutungen werden im Super-Frame erklärt. Die See\_also Relation ist keine semantische Relation und ist nicht konzipiert für die Benutzung in maschinellen Systemen (Ruppenhofer u. a., 2006, Fußnote 22, Seite 125). Sie dient lediglich der Arbeitserleichterung bei der Annotation.

### 2.1.2.3. FrameNet-Hierarchie

Die FN-Hierarchie ist ein gerichteter, zyklischer, unzusammenhängender Multigraph<sup>1</sup>, bestehend aus 795 Knoten, den Frames, in 26 unabhängigen Teilgraphen (siehe Tabelle 2.1). Sieben dieser Knoten sind isoliert. Insgesamt gibt es 86 verschiedene Wurzeln<sup>2</sup>. Die Gesamtzahl der eingehenden Kanten beträgt 1136, das ergibt durchschnittlich 2,86 eingehende Kanten pro Frame. 559 Frames sind von mehr als einem Startknoten aus erreichbar. Der Durchmesser<sup>3</sup> beträgt 15, gemessen in der Anzahl der Kanten ohne Zyklen.

<sup>1</sup>In einem Multigraphen sind zwei Knoten durch mehr als eine Kante verbunden.

<sup>2</sup>Wurzeln sind hier Knoten ohne eingehende Kanten. Es können jedoch nicht alle Knoten des FN-Graphen von solch einer Wurzel erreicht werden.

<sup>3</sup>Der Durchmesser beschreibt den längsten Pfad in einem Graphen.

Anzahl Frames	795
Anzahl Wurzeln	86
Anzahl isolierter Knoten	7
Anzahl unabhängiger Teilgraphen	26
Anzahl Frames erreichbar von mehr als einem Startknoten	559
Anzahl eingehender Kanten	1136
Durchschnittl. Anzahl eingehender Kanten pro Frame	2,86
Durchmesser	15

Tabelle 2.1.: FN-Hierarchie (Release 1.3) in Zahlen

Aus diesen Zahlen läßt sich bereits auf Probleme bei der Arbeit mit der FN-Hierarchie schließen. Da die FN-Hierarchie keine Baumstruktur aufweist, ist eine 1-zu-1 Adaption von baumbasierten Ähnlichkeitsmaßen für FN nicht ohne Weiters möglich.

Die Abdeckung stellt ein weiteres Problem dar. Betrachtet man nur die *Inheritance* Relation, so zeigt sich, dass 40% der Frames im FN-Graph Startknoten sind (Burchardt, 2008). Dadurch wird die Benutzung von semantischen Ähnlichkeitsmaßen, die ausschließlich über *Inheritance*-Beziehungen definiert sind, schwierig bis unmöglich. Eine Modellierung von semantischer Ähnlichkeit auf Basis der *Inheritance*-Relation wäre allerdings wünschenswert, da die *Inheritance*-Relation die am besten und weitesten ausgearbeitete FN-Relation ist und die Hyponym/Hyperonym - Relation ausdrückt.

Ein drittes Problem in FN besteht darin, dass die semantische Distanz, die durch die Kanten im FN-Graphen modelliert wird, von Frame zu Frame sehr unterschiedlich sein kann (vgl. Kapitel 3.1.3 für WordNet). In Abbildung 2.4 ist der Pfad von *EVENT* nach *EMPTYING* dargestellt. Der kürzeste Pfad hat die Länge 3 und der längste 8. Man erreicht vom Frame *INTENTIONALLY\_ACT* den Frame *TRANSITIVE\_ACTION* entweder über eine *Inheritance*-Kante oder über die Frames *SUBJECTIVE\_INFLUENCE* und *OBJECTIVE\_INFLUENCE* und die Kanten *Using, See\_also* und *Inheritance*. Hier hat man die Wahl zwischen einer *Inheritance*-Kante oder einem Pfad mit der Länge 3. Zwischen den Frames *TRANSITIVE\_ACTION* und *EMPTYING* kann man zwischen einer *Inheritance*-Kante oder einem Pfad der Länge 4 wählen.

Darüber hinaus veranschaulicht das Beispiel in Abbildung 2.4 ein weiteres Problem: Das Vorhandensein von Zyklen in der FN-Hierarchie. Der *PLACING*-Frame ist durch die *See\_also*-Relation mit dem *REMOVING*-Frame verbunden. Von diesem wiederum führt eine *Precedes*-Kante zurück zum *PLACING*-Frame

In der vorliegenden Arbeit werde ich zeigen, dass es trotz dieser Probleme möglich ist, ein semantisches Ähnlichkeitsmaß auf Basis der FN-Hierarchie zu definieren.

### 2.1.3. Aktuelle Forschung

Der Fokus der wissenschaftlichen Arbeit bei FN liegt auf der manuellen und automatischen Erweiterung der Ressource für das Englische und für weitere Sprachen, wie zum Beispiel für das Spanische (Subirats und Petruck, 2003), das Japanische (Ohara u. a., 2004), das Deutsche (Erk u. a., 2003) und das Italienische (Tonelli und Pianta, 2008).

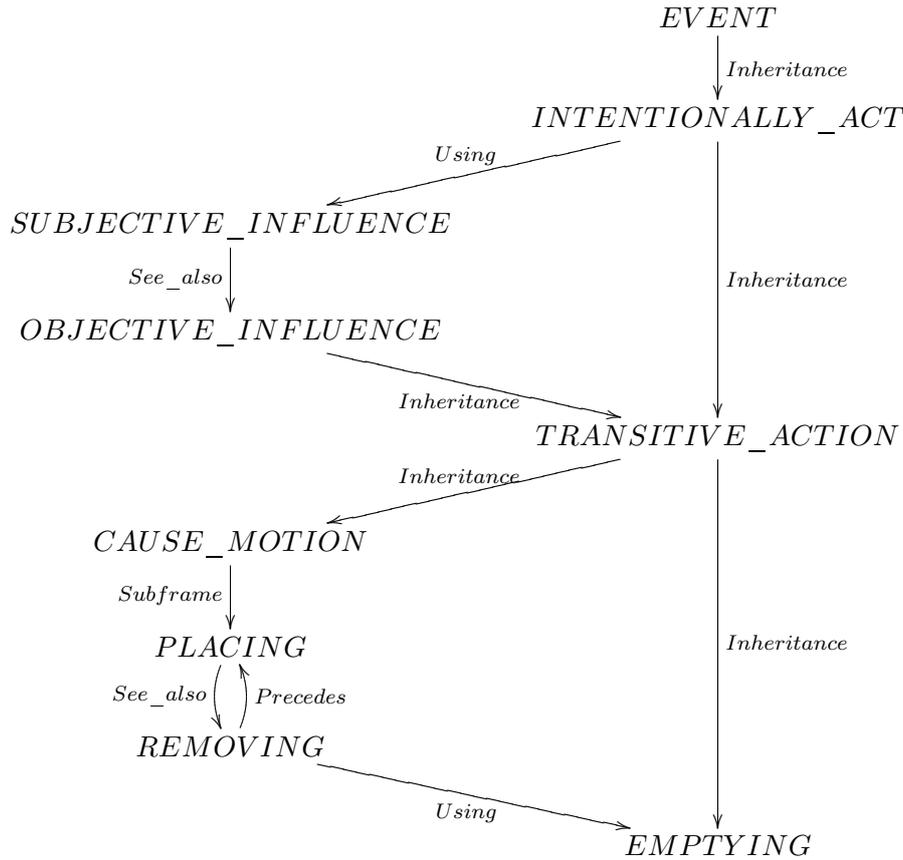


Abbildung 2.4.: Pfad Change\_of\_state\_initial\_state - Emptying

Des weiteren wird FN in einigen wenigen NLP Systeme als Ressource genutzt. Zum einen in *Question Answering*- Systemen (Narayanan und Harabagiu (2004), Frank u. a. (2007), Shen und Lapata (2007)) und in Systemen zur Erkennung semantischer Folgebeziehungen (Burchardt und Frank (2006), Burchardt u. a. (2007), Burchardt (2008), Burchardt und Pennachiotti (2008)).

## 2.2. WordNet

Dieser Abschnitt enthält eine kurze Einführung in WordNet (Fellbaum, 1998). WordNet ist ein elektronisches Wörterbuch. Es wird seit 1985 an der Princeton University entwickelt. Es ist eine der größten lexikalisch semantischen Ressourcen für das Gegenwartsg Englisch. Im Folgenden wird die Struktur vorgestellt.

- |   |
|---|
| <ol style="list-style-type: none"> <li>1. {<b>car</b>, auto, automobile, machine, motorcar (a motor vehicle with four wheels; usually propelled by an internal combustion engine)}</li> <li>2. {<b>car</b>, railcar, railway car, railroad car (a wheeled vehicle adapted to the rails of railroad)}</li> <li>3. {<b>car</b>, gondola (the compartment that is suspended from an airship and that carries personnel and the cargo and the power plant)}</li> <li>4. {<b>car</b>, elevator car (where passengers ride up and down)}</li> <li>5. {cable car, <b>car</b> (a conveyance for passengers or freight on a cable railway)}</li> </ol> |
|---|

Abbildung 2.5.: Bedeutungen „car“

**Synset** Die in dieser Arbeit genutzte Version 3 von WordNet enthält 155000 Wörter. Davon sind rund 117000 Nomen. Desweiteren enthält WordNet Verben, Adjektive und Adverbien in jeweils separaten Hierarchien. Die einzelnen Wörter werden zu Synonym-Gruppen, den *Synsets*, zusammengefasst. Desweiteren enthält ein *Synset* eine Umschreibung der Bedeutung des *Synsets*, die sogenannte Glosse. Insgesamt enthält WordNet über 115000 *Synsets*.

Eine Wortform in einem *Synset* steht für eine Bedeutung des Wortes. Ambige Wörter kommen dementsprechend in mehr als einem *Synset* vor. Abbildung 2.5 präsentiert die Synsets, in denen „car“ (dt.: Auto) enthalten ist. Das Wort „car“ besitzt also fünf verschiedene Bedeutungen in WordNet. Die Glosse von Bedeutung 1. umschreibt „car“ als:

„a motor vehicle with four wheels; usually propelled by an internal combustion engine“

(dt.: „Ein motorisiertes Gefährt mit vier Rädern; gewöhnlich angetrieben durch einen Verbrennungsmotor“)

**Relationen** *Synsets* sind in Hierarchien organisiert. Die Hierarchien sind nach Wortarten getrennt. Es gibt Hierarchien für Nomen, Verben, Adjektive und Adverbien. Die Nomen-Hierarchien sind die am weitesten ausgearbeiteten Hierarchien. Insgesamt gibt es 11 unabhängige Nomen-Hierarchien. Die Synsets in den Nomen-Hierarchien sind durch die Hyperonym-Relation organisiert. Abbildung 2.6 zeigt ein Beispiel für die Hyperonyme von „dolphin“ (dt.: Delphin). Wir sehen, dass ein „dolphin“ ein „mammal“ (dt.: Säuge-

```

dolphin
  -> whale
    -> mammal
      -> vertebrate, carniate
        -> chordate
          -> animal, animate being, beast, brute, creature, fauna
            -> orgnism, being
              -> living thing
                -> physical entity
                  -> entity

```

Abbildung 2.6.: Hyperonyme „dolphin“ (Auszug)

<b>Nomen</b>	<b>Verb</b>	<b>Adjektiv</b>	<b>Adverb</b>
Antonym	Antonym	Antonym	Antonym
Hyponym	Troponym	Gleichheit	Abgeleitet von
Hyperonym	Hyperonym	Relationales Adj.	
Meronym	Folge	Siehe auch	
Holonym	Grund	Attribut	
Attribut	Siehe auch	Partizip	

Tabelle 2.2.: Relationen in WordNet

tier) ist. Und wir können ablesen, dass ein „mammal“ ein „animal“ (dt.: Tier) ist. Neben der Hyperonym-/Hyponym-Relation gibt es noch weitere semantische Relationen. Tabelle 2.2 zeigt die Relation aufgeschlüsselt nach der Wortart. Ein Troponym ist dabei eine Sonderform eines Hyponyms speziell für Verben in WordNet (siehe Fellbaum und Miller, 1990). Die Synonym-Relation ist nicht explizit aufgeführt, da alle Wörtern eines *Synsets* Synonyme sind. Die Relationen werden beispielsweise genutzt, um Wortähnlichkeit zu berechnen. Einige WordNet-basierte Ähnlichkeitsmaße werden in Kapitel 3.1.3 vorgestellt.

## 3. Wissenschaftliches Umfeld

Dieses Kapitel setzt sich aus drei Teilen zusammen. Im ersten Teil, Abschnitt 3.1, wird ein Überblick über den aktuellen Stand an Systemen zur Erkennung von semantischer Ähnlichkeit gegeben. Im zweiten Teil, Abschnitt 3.2, wird ein Experimentdesign zur Evaluation von Ähnlichkeitsmaßen vorgestellt. Der letzte Teil, Abschnitt 3.3, ordnet die aktuelle Fragestellung in das wissenschaftliche Umfeld ein.

### 3.1. Systeme zur automatischen Erkennung von semantischer Ähnlichkeit

Ganz grundsätzlich können folgende Typen von Systemen unterschieden werden:

- Wörterbuch-basierte -Systeme
- Systeme basierend auf semantischen Netzen
- Thesaurus-basierte Systeme
- Integrierte Systeme

In den nächsten Abschnitten werden jeweils die verschiedenen Typen von Systemen exemplarisch an jeweils ein bis zwei Systemen vorgestellt.

#### 3.1.1. Wörterbuch-basierte Systeme

Maschinen-lesbare Wörterbücher, wie das Longman Dictionary Of Contemporary English (LDOCE) (Summers, 2003) oder das Oxford Advanced Learner's Dictionary (OALD) (Hornby, 1995) werden genutzt, um semantische Ähnlichkeiten zu berechnen.

Die Ansätze von Stevenson und Wilks (2001) oder von Kozima und Furugori (1993) beispielsweise benutzen zur Errechnung der semantischer Ähnlichkeit das LDOCE. Während Stevenson und Wilks eine „normalisierte Wortüberlappung“<sup>1</sup> Methode benutzen, um die Ähnlichkeit von Wörtern zu modellieren, überführen Kozima und Furugori das LDOCE in ein semantisches Netz und benutzen einen „spreading activation“ - Ansatz zur Errechnung von semantischer Ähnlichkeit.

---

<sup>1</sup>Bei Wortüberlappungsansätzen errechnet sich die Ähnlichkeit zweier Wörter über die Überlappung der Kontexte, in denen die Wörter vorkommen. Im einfachsten Fall werden die Wörter gezählt, die in beiden Kontexten vorkommen. Kontexte können beispielsweise die Definitionen der Wörter in einem Wörterbuch sein oder Textabschnitte aus Korpora.

Kozima und Furugori (1993) erzeugen hierfür in einem ersten Schritt aus dem „Longman Defining Vocabulary“ (Abk.: LDV) ein semantisches Netz. Das LDV beschreibt das Vokabular, das in den Definitionen der Wörterbucheinträge verwendet wird. Jedes dieser Wörter wird auf einen Knoten abgebildet und diese Knoten sind untereinander durch Kanten verbunden. Knoten besitzen ein Attribut „activation“. Beim Vergleich zweier Wörter, werden deren jeweilige Knoten im semantischen Netz aktiviert, indem dem „activation“ Attribut ein Wert zugewiesen wird. Diese Aktivierung pflanzt sich in mehreren Schritten fort, indem benachbarte Knoten ebenfalls aktiviert werden. Ergebnis dieses Prozesses sind „activation pattern“. Zur Bestimmung der Ähnlichkeit zwischen zwei Wörtern werden diese „activation pattern“ verglichen.

### 3.1.2. Thesaurus-basierte Systeme

In diesen Systemen werden Thesauri, wie Roget's Thesaurus (Kipfer, 2001) oder der Macquarie Thesaurus (Bernard, 1986), benutzt. Thesauri klassifizieren Wörter und Phrasen nach Ideen oder Konzepten. Je nachdem, sind diese Klassen selbst wieder strukturiert in Kategorien, Unterkategorien oder Gruppen. Ein wichtiger Punkt ist, dass die Klassen lediglich indiziert werden. Welche konkrete Idee oder welches konkrete Konzept zu dieser Klassifizierung führt, muss nicht näher ausgeführt sein. (Morris, 1988)

Morris und Hirst (1991) nutzen, in ihrem speziell für Roget's Thesaurus angepassten Ansatz, die Ressource, um Muster lexikalischer Kohäsion zu identifizieren. Ein Muster besteht aus ähnlichen Wörtern, die eine Kette bilden. Zwei Wörter sind sich semantisch ähnlich, wenn sie entweder die gleiche Grundform besitzen, oder auf eine der folgenden Arten miteinander in Beziehung stehen:

1. Beide Wörter haben eine Kategorie in ihrem Index gemeinsam
2. Ein Wort hat eine Kategorie in seinem Index, die auf eine Kategorie im Index des anderen Wortes verweist
3. Ein Wort ist entweder ein label im Index des anderen Wortes, oder ist in der Kategorie des anderen Wortes
4. Zwei Wörter sind in derselben Gruppe enthalten, also semantisch ähnlich
5. Beide Wörter haben eine Kategorie in ihrem Index, die auf eine gemeinsame dritte Kategorie verweist.

### 3.1.3. Systeme basierend auf semantischen Netzen

Ein semantisches Netz ist ein Netzwerk von Knoten verbunden durch Kanten, wobei die Knoten Konzepte sind und die Kanten bedeutungsvolle Verbindungen zwischen den Knoten darstellen (Lee u. a., 1993). Beispiele hierfür sind WordNet (Miller u. a. 1990; Fellbaum 1998) und FrameNet (vgl. Kapitel 2.1). Die im Weiteren in diesem Abschnitt vorgestellten Systeme nutzen WordNet als Ressource.

Ansätze, die mit semantischen Netzen arbeiten, berechnen die Pfadlängen, als Maß für semantische Ähnlichkeit (Rada u. a. 1989, Hirst und St-Onge 1998). Die grundlegende Annahme in diesem Ansatz ist, dass semantisch eng verwandte Konzepte auch im semantischen Netz eng beieinander liegen:

„[...] to evaluate the distance between the nodes corresponding to the items being compared - the shorter the path from one node to another, the more similar they are. Given multiple paths, one takes the length of the shortest one.” (Resnik, 1995)

(deutsch: „[...] um die Distanz zwischen den Knoten, die mit den zu vergleichenden Objekten korrespondieren, auszuwerten - je kürzer der Pfad zwischen den Knoten ist, desto ähnlicher sind sie sich. Wenn mehrere Pfade gegeben sind, wählt man den kürzesten.“)

Dieser Ansatz hat jedoch den Nachteil, dass Kanten in semantischen Netzen nicht immer die gleiche konzeptuelle Distanz modellieren (Resnik, 1995).

Es wurden Ansätze entwickelt, die diesem Effekt entgegen wirken. Sussna (1993; 1997) berücksichtigt bei seinem Ähnlichkeitsmaß die Tiefe in der Hierarchie. Die grundlegende Idee ist hierbei, dass Paare von Konzepten, die tief in der Hierarchie eingebettet sind, sich ähnlicher sind, als Paare von Konzepten, die weit oben in der Hierarchie stehen.

Einen ähnlichen Ansatz verfolgen Wu und Palmer (1994), die bei der Berechnung der Ähnlichkeit die Tiefe des höchsten gemeinsamen Knoten auf dem Pfad zwischen den beiden Konzepten berücksichtigen .

Die Systeme von Hirst und St-Onge (1998), Wu und Palmer (1994) und Leacock und Chodorow (1998) werden im Folgenden ausführlicher vorgestellt, da sie in dieser Arbeit für FrameNet adaptiert werden (siehe Kapitel 5.1).

### **3.1.3.1. Hirst & St. Onge**

Hirst und St-Onge (1998) entwickelten ihr Maß, um Wortverwechslungen mit Hilfe lexikalischer Ketten zu erkennen. Lexikalische Ketten sind dabei definiert als Kette von Wörtern, die durch kohäsive Relationen verbunden sind. Hirst und St-Onge adaptieren den von Morris und Hirst (1991) entwickelten Algorithmus zur Bestimmung der semantischen Distanz zweier Wörter in WordNet. Es werden drei Arten von Relationen unterschieden: extra-starke, starke und mittel-starke Relationen. Die durch Relationen verbundenen Synsets in WordNet bilden dabei kohäsive Ketten als Pfade durch die WordNet-Hierarchie.

Eine extra-starke Relation besteht nur zwischen zwei identischen Wörtern. Eine starke Relation besteht zwischen zwei Wörtern, die entweder dem selben Synset zugeordnet sind, die durch höchstens eine horizontale Relation miteinander verbunden sind, oder zwischen zwei Wörtern, von denen eines Bestandteil des anderen ist. Eine mittel-starke Relation besteht zwischen Wörtern, die nach den Pfadmustern in Abbildung 3.1 miteinander verbunden sind. Diese Pfadmuster folgen den folgenden von Hirst und St-Onge aufgestellten Regeln:

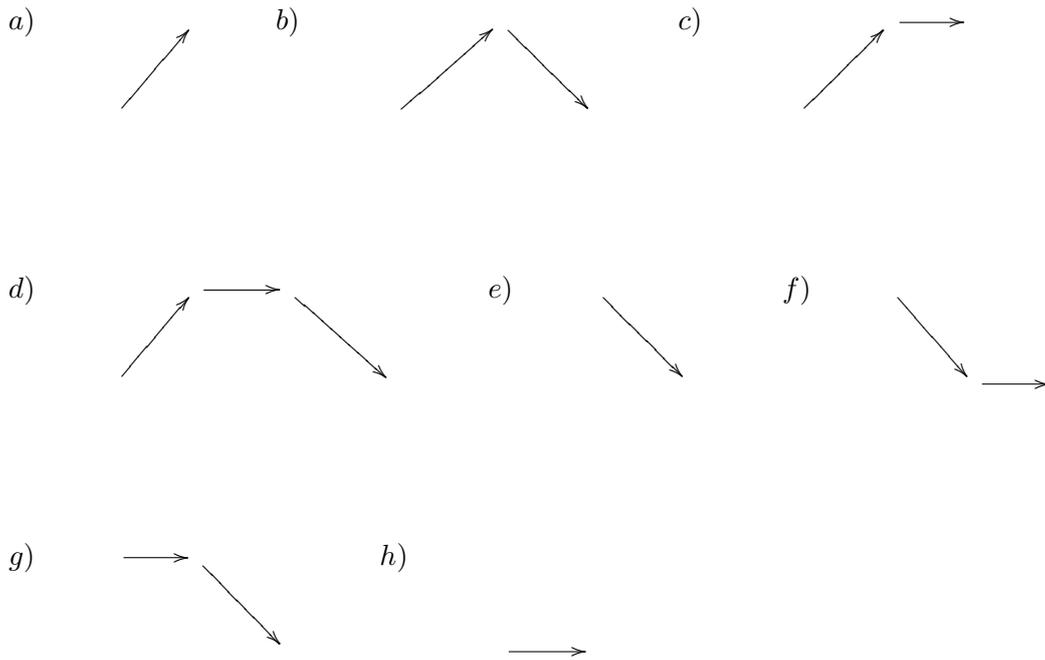


Abbildung 3.1.: Gültige Pfadmuster (Hirst und St-Onge, 1998)

„(R1) No other direction may precede an upward link.

(R2) At most one change of direction is allowed.

(R2') It is permitted to use a horizontal link to make a transition from an upward to a downward direction.” (Hirst und St-Onge, 1998, ;

Seite 7)

(dt.: (R1) Keine andere Richtung geht einer aufwärts gerichteten Kante voraus.

(R2) Höchstens eine Richtnugsänderung ist erlaubt.

(R2') Es ist erlaubt eine horizontale Verbindung zu nutzen, um von aufwärts gerichteten Kanten zu abwärts gerichteten Kanten zu wechseln.)

Das Gewicht eines solchen Pfades wird durch folgende Formel errechnet:

$$weight = C - path\ length - k * number\ of\ changes\ of\ direction \quad (3.1)$$

C und k sind Konstanten. Insgesamt werden durch dieses Maß kurze Pfade mit wenigen Richtungsänderungen bevorzugt. Dies ist das einzige taxonomie-basierte Ähnlichkeitsmaß, das neben den Hyponym-/Hyperonym-Relationen der WordNet Hierarchie weitere Relationen nutzt. Als horizontale Relationen werden beispielsweise die Implikation, die

Antonymie, die Meronymie/Holonymie und die Kausation bezeichnet. Das Maß ist jedoch speziell für die gut ausgearbeitete Nomen-Hierarchie von WordNet entwickelt worden.

### 3.1.3.2. Leacock & Chodorow

Leacock und Chodorow (1998) entwickeln in ihrem Papier „Combining local context and WordNet similarity for word sense identification“ ein Ähnlichkeitsmaß zur Verbesserung eines Systems zur Identifikation von Wortbedeutungen. Das System arbeitet mit einem korpusbasierten, statistischen Klassifizierer.

Der Klassifizierer benötigt eine große Menge an Trainingsdaten, um in der Testphase gute Ergebnisse liefern zu können. Sind nicht genug Trainingsdaten vorhanden, nennt man dieses Problem das *sparse data* Problem (Katz, 1987). Um die Effekte dieses Problems zu mildern, benutzen Leacock und Chodorow WordNet als lexikalische Ressource.

Die Idee ist, dass semantisch ähnliche Wörter auf einen semantisch ähnlichen Kontext schließen lassen. Semantisch ähnliche Wörter sollen die Lücken in den Trainingsbeispielen schließen, indem in Bezug auf die Trainingsbeispiele unbekannte Wörter durch semantisch gleiche Wörter ersetzt werden, die dem System bekannt sind.

Leacock und Chodorow benutzen zur Errechnung der semantischen Ähnlichkeit die Nomen-Taxonomien von WordNet. Die semantische Ähnlichkeit wird modelliert durch eine skalierte Pfadlänge. Der Pfad wird skaliert durch die Tiefe der Taxonomie. Formal errechnet sich die Ähnlichkeit, wie folgt:

$$sim = \max[-\log(Np/2D)] \quad (3.2)$$

Hierbei bezeichnet  $Np$  die Länge des Pfades und  $D$  ist die maximale Tiefe der Taxonomie. Ein langer Pfad in einer Taxonomie mit geringer Tiefe wird schlechter bewertet, als ein langer Pfad in einer Taxonomie mit großer Tiefe. Um sicherzustellen, dass immer ein Pfad zwischen beliebigen Wortbedeutungen gefunden werden kann, führen Leacock und Chodorow einen Wurzelknoten ein, der die Wurzelknoten der 11 Nomen-Taxonomien von WordNet verbindet.

Wie schon das System von Hirst und St-Onge (1998), arbeitet auch dieses Maß ausschließlich auf der WordNet-Nomen-Taxonomie.

### 3.1.3.3. Wu & Palmer

In ihrem Artikel behandeln Wu und Palmer (1994) das Problem der lexikalischen Selektion bei der maschinellen Übersetzung chinesischer Verben ins Englische.

Ihre Lösung besteht darin, Verben auf konzeptuelle Domänen abzubilden und diese in einer hierarchischen Struktur zu organisieren. Auf dieser hierarchischen Struktur wird die konzeptuelle Ähnlichkeit von Verben modelliert. Die konzeptuelle Ähnlichkeit *ConSim* ist hierbei umso größer, je näher die beiden Konzepte in der Hierarchie zueinander liegen. Gemessen wird der Abstand der beiden Konzepte  $C1$  und  $C2$  zum ersten gemeinsamen Superkonzept  $C3$ , dass formal, wie folgt, definiert ist:

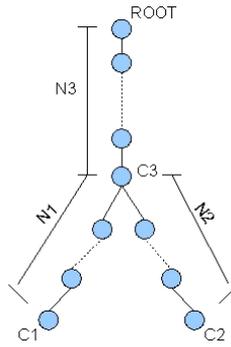


Abbildung 3.2.: Konzept - Ähnlichkeit nach Wu und Palmer (1994)

$$ConSim(C1, C2) = \frac{2 * N3}{N1 + N2 + 2 * N3} \quad (3.3)$$

$N1$  und  $N2$  bezeichnen den Abstand der Konzepte zum ersten gemeinsamen Superknoten;  $N3$  ist der Abstand des ersten gemeinsamen Superknotens zur Wurzel. Siehe Abbildung 3.2.

### 3.1.4. Hybride Systeme

Die Systeme, die in diesem Abschnitt vorgestellt werden, reichern die bereits im semantischen Netz vorhandene Information durch Korpusinformationen an, um das *sparse data*-Problem zu lösen.

#### 3.1.4.1. Graph-basierte Ähnlichkeitsmaße

Alle in diesem Abschnitt beschriebenen Systeme nutzen WordNet als Ressource.

Resnik (1995) beschreibt die semantische Ähnlichkeit von zwei Konzepten als den Informationsgehalt des ersten gemeinsamen Konzeptes auf dem Pfad zwischen den zu untersuchenden Konzepten. Der Begriff Informationsgehalt ist der Informationstheorie (Ross, 1976) entliehen. Formal wird das Ähnlichkeitsmaß folgendermaßen beschrieben:

$$sim_R(c_1, c_2) = -\log p(lso(c_1, c_2)) \quad (3.4)$$

Hierbei bezeichnet  $lso$  das erste übergeordnete Konzept. Die Wahrscheinlichkeit  $p(c)$  errechnet sich unter Zuhilfenahme des Brown Corpus of American English (Francis u. a., 1982).

$$freq(c) = \sum_{n \in words(c)} count(n) \quad (3.5)$$

$$p(c) = \frac{freq(c)}{N} \quad (3.6)$$

Die Wahrscheinlichkeit eines Konzeptes  $p(c)$  ist die Anzahl der vorkommenden Nomen, die das Konzept erfasst, normalisiert mit der Gesamtzahl der im Brown Corpus vorkommenden Nomen  $N$ , die in WordNet annotiert sind.

Jiang und Conrath (1997) kritisieren Resniks Ansatz, da dort die Taxonomie nur benutzt wird, um das erste übergeordnete Konzept zu bestimmen. Ihr Ansatz berücksichtigt die lokale Dichte der Hierarchie, die Tiefe des Knotens in der Hierarchie und die Art der Relation zwischen den Konzepten. Die semantische Distanz errechnet sich als die Summe der Gewichtungen der Kanten des kürzesten Pfades.

### 3.1.4.2. Informationstheoretische Ähnlichkeitsmaße

Die in diesem Abschnitt vorgestellten Ansätze lösen sich von einer speziellen Resource und formulieren semantische Ähnlichkeit auf Grundlage von informationstheoretischen Überlegungen.

Lin (1998) definiert auf der Basis der Informationstheorie (Cover und Thomas, 1991) einen ressourcenunabhängigen Begriff von semantischer Ähnlichkeit:

„**Similarity Theorem:** The similarity between A and B is measured by the ratio between the amount of information needed to state the commonality of A and B and the information needed to fully describe what A and B are.”  
(Lin, 1998, Seite 2)

(dt.: Die Ähnlichkeit zwischen A und B berechnet sich durch die Menge an Informationen, die benötigt werden, um die Gemeinsamkeiten zwischen A und B auszudrücken, und die Menge an Informationen, die benötigt werden, um A und B vollständig zu beschreiben.)

$$sim_L(A, B) = \frac{\log P(common(A, B))}{\log P(description(A, B))} \quad (3.7)$$

Maguitman u. a. (2005) entwickeln den Ansatz von Lin weiter und stellen ein graph-basiertes Ähnlichkeitsmaß für den Open Directory Project-Graph vor, das auch „horizontale Verbindungen“ berücksichtigt. Das Open Directory Project<sup>2</sup> (kurz: ODP) erstellt eine Ontologie zur Klassifizierung von Internet-URLs. Zur Zeit sind über 4,5 Millionen URLs in über 590000 Kategorien erfasst.

Die Erweiterung um „horizontale Verbindungen“ ist notwendig, da

---

<sup>2</sup>[www.dmoz.org](http://www.dmoz.org)

„An important distinction between taxonomies and ontologies such as the ODP graph is that edges in a taxonomy are all of the same type („is-a” links), while in the ODP graph edges can have diverse types (e.g., „is-a”, „symbolic”, „related”). Different types of edges have different meanings and should be used accordingly.” (Maguitman u. a., 2006, Seite 436)

(deutsch: Ein bedeutender Unterschied zwischen Taxonomien und Ontologien, wie der ODP Graph, besteht darin, dass die Kanten in einer Taxonomie alle vom selben Typ sind („is-a” Verbindungen), während im ODP-Graphen Kanten unterschiedliche Typen besitzen können (e.g. „is-a”, „symbolic”, „related”). Verschiedene Typen von Kanten haben unterschiedliche Bedeutungen und sollten dementsprechend benutzt werden.)

Diese „horizontalen Verbindungen” sind vergleichbar mit den bei Hirst und St-Onge (1998) genutzten „nicht-klassischen” Relationen (siehe Abschnitt 3.1.3.1 auf Seite 16).

Der ODP-Graph wird zur Berechnung der semantischen Ähnlichkeit in zwei Teile, einen hierarchischen und einen nicht-hierarchischen Teil, aufgespalten. Den hierarchischen Teil bilden die Teile des Graphen, die durch „is-a”-Kanten verbunden sind. Den nicht-hierarchischen Teil bilden die Teile des Graphen, die durch „symbolic” oder „related”-Kanten verbunden sind. Die einzelnen Komponenten werden durch Adjazenzmatrizen kodiert. Anschließend wird mit Formel 3.12 die semantische Ähnlichkeit von zwei Klassen des ODP-Graphen errechnet. Matrix  $\mathbf{T}$  kodiert die hierarchische Komponente (siehe Gleichung 3.8)

$$\mathbf{T}_{ij} = \begin{cases} 1 & \text{wenn } i = j, \\ \alpha & \text{wenn } i \neq j \text{ und } (i, j) \in T, \\ 0 & \text{sonst.} \end{cases} \quad (3.8)$$

$T$  steht für die Menge aller „is-a”-Kanten und  $\alpha$  beschreibt das Gewicht einer speziellen „is-a”-Kante. Die nicht-hierarchischen Komponenten werden analog, als Matrix  $\mathbf{S}$ , bzw. Matrix  $\mathbf{R}$ , codiert. Alle drei Komponenten werden im nächsten Schritt zu einer Adjazenzmatrix zusammengefasst. Dies geschieht mit Hilfe der Fuzzy-Vereinigungsoperation  $[A \vee B]_{ij} = \max(A_{ij}, B_{ij})$ .

$$G = T \vee S \vee R \quad (3.9)$$

$\mathbf{G}$  bildet die Adjazenzmatrix des ODP-Graphen. Ein Eintrag  $G_{ij}$  beschreibt die semantische Ähnlichkeit direkt verbundener Klassen des ODP-Graphen  $t_i$  und  $t_j$ .

In einem nächsten Schritt wird die Hülle  $\mathbf{T}^+$  der Matrix  $\mathbf{T}$  errechnet. Hierbei ist  $\mathbf{T}_{ij}^+ = 1$ , wenn  $t_j \in subtree(t_i)$ , sonst  $\mathbf{T}_{ij}^+ = 0$ . Abschließend wird eine „fuzzy membership matrix” (Maguitman u. a., 2005)  $\mathbf{W}$  von  $\mathbf{G}$  berechnet.

$$W = T^+ \odot G \odot T^+ \quad (3.10)$$

$\odot$  steht für die „MaxProduct fuzzy composition“-Funktion (Kandel, 1986), die für Matrizen, wie folgt, definiert ist:

$$[A \odot B]_{ij} = \max_k (A_{ik} \cdot B_{kj}) \quad (3.11)$$

Ein Eintrag  $W_{ij}$  beschreibt das Gewicht des Pfades der die beiden Klassen  $t_i$  und  $t_j$  verbindet. Wichtig ist, mit der Hülle  $T^+$  wird ein Pfad beliebiger Länge im hierarchischen Teil des ODP-Graphen beschrieben. Die Matrix  $G$  beschreibt jedoch nur die Verbindung direkt benachbarter Klassen. Allerdings erfasst  $G$  alle Arten von Relationen im ODP-Graphen.

Ein durch  $W$  beschriebener Pfad setzt sich aus drei Teilen zusammen. Der erste Teil ist ein beliebig langer Pfad im hierarchischen Teil des ODP-Graphen. Der zweite Teil, codiert in  $G$ , ist ein Pfad der Länge 1, der möglicherweise eine „symbolic“- oder „related“-Kante enthält. Der dritte Teil ist wieder ein beliebig langer Pfad aus dem hierarchischen Teil des ODP-Graphen.

Mit der Matrix  $\mathbf{W}$  wird nun die Ähnlichkeit zweier Klassen der Ontologie folgendermaßen berechnet:

$$\sigma_S^G(t_1, t_2) = \max_k \frac{2 \cdot \min(W_{k1}, W_{k2}) \cdot \log Pr[t_k]}{\log(Pr[t_1|t_k] \cdot Pr[t_k]) + \log(Pr[t_2|t_k] \cdot Pr[t_k])} \quad (3.12)$$

Die Wahrscheinlichkeit  $Pr[t_k]$  errechnet sich dabei als:

$$Pr[t_k] = \frac{\sum_{t_j \in V} (W_{kj} \cdot |t_j|)}{|U|} \quad (3.13)$$

wobei  $|U|$  die Anzahl aller Dokumente in der Ontologie beschreibt und  $V$  die Menge aller Klassen, die Dokumente enthalten. Die bedingte Wahrscheinlichkeit  $Pr[t_i|t_k]$  errechnet sich folgendermaßen:

$$Pr[t_i|t_k] = \frac{\sum_{t_j \in V} (\min(W_{ij}, W_{kj}) \cdot |t_j|)}{\sum_{t_j \in V} (W_{kj} \cdot |t_j|)} \quad (3.14)$$

Das hier vorgestellte Ähnlichkeitsmaß von Maguitman u. a. (2006) wird ebenfalls in dieser Arbeit für FrameNet adaptiert (siehe: Kapitel 5.1). Es erweitert das durch Lin (1998) vorgestellte universelle Ähnlichkeitsmaß so, dass es für beliebige Graphen adaptiert werden kann.

**Zusammenfassung** In diesem Abschnitt wurden verschiedene Ansätze zur Errechnung semantischer Ähnlichkeit vorgestellt. Ansätze, die mit semantischen Netzen arbeiten, sind von besonderem Interesse, da sie auch auf FrameNet anwendbar sein sollten. Dies wird an Hand der Ansätze von Hirst und St-Onge (1998), Leacock und Chodorow (1998), Wu und Palmer (1994) und Maguitman u. a. (2006) in den Kapiteln 5 und 6 dieser Arbeit untersucht. Diese Ansätze stehen stellvertretend für pfadbasierte Ansätze (Hirst und

St-Onge, 1998), Ansätze mit skaliertem Pfadlänge (Leacock und Chodorow (1998); Wu und Palmer (1994)) und hybride Ansätze (Maguitman u. a., 2006). Budanitsky (1999) und Budanitsky und Hirst (2001, 2006) evaluieren und diskutieren ausführlich die hier vorgestellten Ansätze und Systeme.

### 3.2. Methodik zur Evaluation von Systemen zur automatischen Berechnung semantischer Ähnlichkeit

Um die Performanz der in Kapitel 3.1 vorgestellten Ähnlichkeitsmaße beurteilen zu können, benötigt man ein einheitliches Evaluationsschema.

Ein wegweisendes (vgl. Miller und Charles (1991), Resnik (1995), Budanitsky und Hirst (2001), Finkelstein u. a. (2002), Gurevych (2006)) Experimentdesign beschreiben Rubenstein und Goodenough (1965) in ihrem Artikel „Contextual Correlates of Synonymy“. Rubenstein und Goodenough untersuchen den Zusammenhang zwischen Wortüberlappungen im Kontext zweier Wörter und der semantischen Ähnlichkeit. Als Basis dafür wird eine Referenzordnung bedeutungsgleicher Wörter benötigt. Diese Referenzordnung erhalten die Autoren durch folgenden Versuch:

65 Wörter sollen gemäß ihrer semantischen Ähnlichkeit geordnet werden. Hierfür werden die 65 Wortpaare auf 65 Karten geschrieben und den Versuchspersonen mit folgenden Instruktionen überreicht:

„1. After looking through the whole deck, order the pairs according to amount of "similarity of meaning" so that the slip containing the pair exhibiting the greatest amount of "similarity of meaning" is at the top of the deck and the pair exhibiting at least amount is on bottom.

2. Assign a value from 4.0-0.0 to each pair –the greater the "similarity of meaning," the higher the number. You may assign the same value to more than one pair.“ Rubenstein und Goodenough (1965, Seite 628)

(deutsch: 1. Nachdem Sie sich den ganzen Stapel angeschaut haben, ordnen Sie die Paare gemäß ihres Gehalts an Bedeutungsgleichheit, so dass die Karte mit dem Paar mit der größten „Bedeutungsgleichheit“ ganz oben auf dem Stapel liegt und das Paar mit dem geringsten Gehalt an „Bedeutungsgleichheit“ ganz unten im Stapel liegt.

2. Weise jedem Paar einen Wert zwischen 4.0 und 0.0 zu – je größer die „Bedeutungsgleichheit“, desto größer die Zahl. Sie dürfen den selben Wert bei mehr als einem Paar vergeben.“)

Die erste Gruppe von Versuchspersonen, bestehend aus 15 Studierenden, beurteilt in einem ersten Durchlauf 48 Wortpaare und zwei Wochen später in einem zweiten Versuchsdurchlauf 65 Wortpaare. 36 Wortpaare sind bei beiden Versuchen identisch und dienen zur Bestimmung der Interrater-Reliabilität. Zur Bestimmung der Korrelation benutzen Rubenstein und Goodenough Spearman's  $\rho$  (Siegel und Castellan, 1988). Die beiden Stichproben korrelieren höchst signifikant mit einem Wert von  $\rho = 0.89$ .

Anschließend bewertet ein zweite Gruppe von 36 Versuchspersonen alle 65 Wortpaare. Die vergebenen numerischen Ähnlichkeitswerte beider Versuchspersonengruppen werden gemittelt und dann die Korrelation zwischen beiden Gruppen bestimmt. Die Korrelation beträgt  $\rho = 0.99$  und ist höchst signifikant.

Als letztes werden die beiden Ordnungen mit den gemittelten Werten zu einer Referenzordnung zusammengeführt. Diese Referenzordnung dient Rubenstein und Goodenough im Weiteren als Diskussionsgrundlage.

Dieses Experimentdesign wird in Kapitel 4.1.1 wieder aufgegriffen und zur Evaluation des Frameähnlichkeitsexperiments genutzt.

### 3.3. Einordnung der Fragestellung ins wissenschaftliche Umfeld

In dieser Arbeit wird ein Ähnlichkeitsmaß für Frames basierend auf der FrameNet-Hierarchie entwickelt. Die FrameNet-Hierarchie ist ein semantisches Netz. Aus diesem Grund werden in Kapitel 5.1 die in Abschnitt 3.1.3 vorgestellten struktur-basierten Ähnlichkeitsmaße und in Abschnitt 3.1.4.2 vorgestellten informations-basierten Ähnlichkeitsmaße für FrameNet adaptiert und auf ihren Nutzen hinsichtlich eines FrameNet-Ähnlichkeitsmaßes untersucht.

In der Linguistik gibt es eine lange Tradition in der Untersuchung von Wortbedeutung und der Ähnlichkeit von Wortbedeutungen (Quillian, 1966; Collins und Loftus, 1975; Fillmore, 1977; Rubenstein und Goodenough, 1965; Morris, 1988; Resnik, 1995; Leacock und Chodorow, 1998; Stevenson und Wilks, 2001; Morris und Hirst, 2004). Die Aufgabe, die semantische Ähnlichkeit von abstrakten Konzepten, in diesem Fall den Frames, zu bestimmen, ist nicht direkt vergleichbar mit der Aufgabe, die semantische Ähnlichkeit von Wörtern zu bestimmen. Trotzdem werden zur Evaluation von Konzeptähnlichkeitsmaßen als Referenzdaten Daten aus Wortähnlichkeitsexperimenten genutzt (vgl. Kim und Candan, 2006). Unter der Verwendung der gleichen Datenbasis und des gleichen Experimentdesign werden die Ergebnisse des Konzeptähnlichkeitssystems direkt mit den vorhandenen Daten verglichen.

Diese Vorgehensweise für die Bewertung eines Frame-Ähnlichkeitssystems zu benutzen ist problematisch. Frames beschreiben eine Szene eindeutig. Bei Wörtern existiert das Problem der lexikalischen Ambiguität und man kann in einem Experimentdesign, wie Rubenstein und Goodenough (1965) es benutzen, nicht feststellen, welche Bedeutung von der Versuchsperson beurteilt wurde.

Vergleicht man die Leistung der in diesem Kapitel vorgestellten Systeme mit der Leistung eines Systems, das ausschließlich FrameNet als Wissensbasis nutzt, so stößt man auf ein weiteres Problem. Die meisten hier vorgestellten Systeme arbeiten lediglich auf der Nomen-Hierarchie von WordNet. Sie berechnen die semantische Ähnlichkeit von Nomen. Nomen bezeichnen allerdings oft konkrete Dinge. Frames hingegen beschreiben überwiegend Situationen und dies auch auf unterschiedliche Art.

Burchardt (2008) schreibt dazu:

„Modeling *semantic similarity* for events is bigger challenge, as they form a heterogeneous class. If we look at FrameNet, we find frames that have different ontological status in that they describe, e.g., complex events (COMMERCIAL\_TRANSACTION), sub-events (COMMERCE\_GOODS-TRANSFER), linguistic perspectives on events (COMMERCE\_BUY), or partial aspects of events (RECIPROCALITY).“ (Burchardt, 2008, Seite 70)

(deutsch: Das Modellieren von semantischer Gleichheit für Events ist eine größere Herausforderung, da sie eine heterogene Klasse bilden. Wenn wir uns FrameNet anschauen, finden wir Frames, die unterschiedlichen ontologischen Status besitzen. So beschreiben sie zum Beispiel komplexe Events (COMMERCIAL\_TRANSACTION), Teil-Events (COMMERCE\_GOODS-TRANSFER), linguistische Blickwinkel auf Events (COMMERCE\_BUY) oder Teilaspekte von Events (RECIPROCALITY))

Diese Problematik ist zu berücksichtigen bei der Auswertung des in Kapitel 4.1 vorgestellten Experiments zur Ähnlichkeit von Frames. Ein direkter Vergleich mit Referenzdaten für Frame-Ähnlichkeit ist nicht möglich, da keine geeigneten Daten zur Verfügung stehen.

## 4. Frame-Ähnlichkeit

In diesem Kapitel wird Frame-Ähnlichkeit empirisch nachgewiesen und FrameNet als Ressource zur Modellierung von semantischer Ähnlichkeit von Frames vorgestellt. In einem abschließenden Abschnitt werden die Vor- und Nachteile von FrameNet bei der Modellierung von Frame-Ähnlichkeit zusammengefasst und eine erste Definition von Frame-Ähnlichkeit gegeben.

### 4.1. Ist eine Definition semantischer Ähnlichkeit für FN-Frames möglich?

In diesem Abschnitt wird empirisch nachgewiesen, dass eine manuelle Annotation von Frame-Ähnlichkeit mit hinreichender Verlässlichkeit möglich ist. Im folgenden beschreibe ich ein im Rahmen dieser Diplomarbeit durchgeführtes Experiment, das sich am Design einer Studie von Rubenstein und Goodenough (1965) orientiert (siehe Kapitel 3.2). Abgeschlossen wird das Kapitel mit der Diskussion der Ergebnisse der Untersuchung und einer ersten Definition von semantischer Ähnlichkeit von Frames.

#### 4.1.1. Pilotstudie

Ziel der Pilotstudie ist es zu zeigen, ob semantische Ähnlichkeit auf der Ebene abstrakter Situationsbeschreibungen beurteilt werden kann. Fillmore (1977) legt dar, dass kognitive Frames im Gedächtnis durch semantische Ähnlichkeit organisiert sind (siehe: Kapitel 2.1.1). In FN korrespondiert dieses Organisationskonzept mit der Ordnung der Frames im semantischen Raum der FN-Hierarchie. Ruppenhofer u. a. (2006) beschreiben die Vorteile, die die FN-Hierarchie bietet, folgendermaßen:

„One obvious benefit of defining relations between frames is that it enables the look-up of words that are related.” Ruppenhofer u. a. (2006, Seite 120)

(deutsch: Ein offensichtlicher Vorteil der Definition von Frame-Relationen besteht darin, verwandte Wörter auffinden zu können.)

Ausgehend von diesen theoretischen Grundlagen, führt dies zu folgender Hypothese:

(H1) Der intuitive Begriff von semantischer Ähnlichkeit ist zwischen FN-Frames nachweisbar und reproduzierbar.

**Experiment** 15 Versuchspersonen bekamen die Aufgabe gestellt, die relative Ähnlichkeit von 30 FN-Frame-Paaren zu bestimmen. Die Frame-Paare wurden in zwei Sets zu je 15 Frame-Paaren aufgeteilt. 5 der 15 Frame-Paaren dienten zur Berechnung der Interrater-Reliabilität. Insgesamt wurden 310 Frame-Paare beurteilt.

Die beiden Datensätze unterscheiden sich durch ihre Zusammenstellung. Für den ersten Datensatz  $FP_{random}$  wurden die Frame-Paare zufällig ausgewählt. Um den Effekt der Unvollständigkeit der FN-Hierarchie (siehe Kapitel 2.1.2.3) zu reduzieren, wurden für den zweiten Datensatz  $FP_{control}$  Frame-Paare ausgewählt, die in der FN-Hierarchie einem gut ausgearbeiteten „\_scenario“-Frames zugeordnet werden können, wie zum Beispiel dem „Commerce\_scenario“, oder durch maximal eine Kante mit einem solchen Frame verbunden sind.

Zur Bestimmung der semantischen Ähnlichkeit der Frames stehen den Versuchspersonen der Frame-Name, die Frame-Definition, die FE des jeweiligen Frames und ein Beispielsatz zur Verfügung. Diese Informationen ermöglichen es den Versuchspersonen, die beschriebene Situation zu erfassen.

**Gold-Standard** Zu Zwecken der Evaluation wurde ein Golden-Standard erstellt, indem die beiden Datensätze  $FP_{random}$  und  $FP_{control}$  von 3 Experten sortiert wurden. Nach Abschluss des Experiments wurde die Korrelation der Rangfolge der Frame-Ähnlichkeit in den *Golden-Standard*-Daten (siehe A.1, A.2) mit der durch die Versuchspersonen hergestellten Rangfolge der Frame-Ähnlichkeiten berechnet (siehe Abschnitt 4.1.2).

#### 4.1.2. Evaluation

Evaluiert werden die Interrater-Reliabilität und die Korrelation der Ordnungen der einzelnen Versuchspersonen mit dem Gold-Standard. Da es sich um ordinal skalierte Daten handelt, sind Korrelationskoeffizienten, wie Spearman’s  $\rho$ , Kendall’s  $\tau$  oder Kendall  $W$ , zur Evaluation geeignet (Siegel und Castellan, 1988).

##### 4.1.2.1. Evaluationsmetriken

Um die Interrater-Reliabilität zu errechnen, wird der Konkordanzkoeffizient  $W$  nach Kendall benutzt (siehe Formel 4.1).

$$W_{Kendall} = \frac{12 \sum \bar{R}_i^2 - 3N(N+1)^2}{N(N^2-1) - k \sum T_j} \quad (4.1)$$

mit

$$T_j = \sum_{i=1}^{g_j} (t_i^3 - t_j) \quad (4.2)$$

wobei  $k$  für die Anzahl der Beurteiler,  $N$  für die Anzahl der zu beurteilenden Frame-Paare und  $\bar{R}_i$  für den durchschnittlichen Rang des  $i$ -ten Frame-Paares steht.  $T_j$  ist die Korrektur für gebundene Ränge.  $t_i$  steht dabei für die Anzahl der gebundenen Ränge an der Position  $i$  und  $g_j$  bezeichnet die Anzahl der gebundenen Rängen bei Beurteiler  $j$ .

Die Kendall W-Metrik ist speziell dazu entwickelt worden, die Übereinstimmung zwischen mehreren Beurteilern zu quantifizieren. Ein großer  $W$  Wert kann so interpretiert werden, dass die Beurteiler bei ihrer Beurteilung denselben Standard angewendet haben (Siegel und Castellan, 1988). Die Errechnung des Konkordanzkoeffizienten über den Durchschnitt der Kendall  $\tau$  - oder Spearman  $\rho$  -Werte ist möglich, aber mit erheblich größerem Aufwand verbunden (vgl. Siegel und Castellan, 1988).

Um die Korrelation zwischen den einzelnen Versuchspersonen und dem Gold-Standard zu bestimmen, wird Kendalls Rangkorrelationskoeffizient  $\tau$  (Kendall, 1938) benutzt:

$$\tau_{Kendall} = \frac{2S}{\sqrt{N(N-1) - T_x} \sqrt{N(N-1) - T_y}} \quad (4.3)$$

mit

$$T_x = \sum t(t-1) \quad (4.4)$$

wobei  $N$  für die Anzahl der zu beurteilenden Frame-Paare steht.  $S$  bezeichnet die Differenz zwischen gleichen und unterschiedlichen Beurteilungen.  $T_x$  steht für den Korrekturfaktor für gebundene Ränge, mit  $t$  als Anzahl der gebundenen Ränge in jeder Gruppe von gebundenen Rängen.

Kendalls  $\tau$  hat gegenüber Spearmans  $\rho$  den Vorteil, dass es sich besser für kleine Stichproben eignet und eine klarere Interpretation zulässt. Kendalls  $\tau$  kann gedeutet werden, als die Differenz der Wahrscheinlichkeit, dass zwei Beobachtungen gleich geordnet sind und der Wahrscheinlichkeit, dass zwei Beobachtungen unterschiedlich geordnet sind (Lapata, 2006).

Die Korrelation zwischen Versuchspersonen und Gold-Standard ergibt sich aus dem Mittelwert der  $\tau_{Kendall}$  Werte der einzelnen Versuchspersonen mit dem Gold-Standard und der Anzahl der beurteilten Frame-Paare  $k$ .

$$Corr_{VP_{GS}} = \frac{\sum_i \tau_{Kendall}(VP_i, GS)}{k} \quad (4.5)$$

#### 4.1.2.2. Diskussion

In diesem Abschnitt wird einerseits die Gültigkeit von (H1) (siehe Kapitel 4.1.1, im Folgenden wiederholt) überprüft.

(H1) Der intuitive Begriff von semantischer Ähnlichkeit ist zwischen FN-Frames nachweisbar und reproduzierbar.

Andererseits wird untersucht, ob die durch die drei Experten hergestellten Referenzordnungen mit der Sortierung der Versuchspersonen korrelieren und als Gold-Standard für die Evaluation der automatischen Ähnlichkeitsmaße benutzt werden können.

Um H1 zu überprüfen, wird die Interrater-Reliabilität für die beiden Datensätze  $FP_{random}$  und  $FP_{controll}$  errechnet. Zur Errechnung der Interrater-Korrelation wird das

im vorigen Abschnitt eingeführte Maß  $W_{Kendall}$  benutzt. Der  $W_{Kendall}$  -Wert der Versuchspersonen für die 5 Kontroll-Frame-Paare des Datensatzes  $FP_{random}$  beträgt 0,789, und für den Datensatz  $FP_{controll}$  0,697. Die Daten sind  $\chi^2$ -verteilt. Bei 4 Freiheitsgraden liegt die Wahrscheinlichkeit, dass die Daten zufällig miteinander korrelieren bei unter 0,1 %. Dies zeigt, dass es eine gemeinsame Intuition bei der Beurteilung gegeben hat. Somit ist der intuitive Begriff von Frame-Ähnlichkeit empirisch nachgewiesen. Es ist also sinnvoll zu versuchen, diese Frame-Ähnlichkeit zu erfassen und zu explizieren.

Die Werte zeigen weiter, dass die Korrelation für  $FP_{random}$  größer ist, als für  $FP_{controll}$ . Es scheint einfacher zu sein, die semantische Ähnlichkeit von Frames zu beurteilen, wenn die semantische Distanz größer wird.

Die Evaluation des Gold-Standard mit den Annotationen der Versuchspersonen ergibt einen durchschnittlichen  $\tau_{Kendall}$  -Wert von 0,530 mit einer Standardabweichung von 0,416 für  $FP_{random}$  und 0,566 mit einer Standardabweichung von 0,173 für  $FP_{controll}$ . Dies bedeutet, dass beide Gold-Standards korrelieren mit  $\alpha < 0.001$  mit den Ordnungen der Versuchspersonen. Somit ist gezeigt, dass die beiden Referenzordnungen verlässlich genug sind, um im Laufe der weiteren Untersuchung als Gold-Standard zu dienen.

**Vergleich mit Wort-Ähnlichkeitsexperimenten** Ein weitere Frage beschäftigt sich damit, ob die Beurteilung von Frame-Ähnlichkeit eine schwierigere Aufgabe darstellt, als die Beurteilung von Wort-Ähnlichkeit. Hierzu können die Ergebnisse aus Wortähnlichkeitsexperimenten von Rubenstein und Goodenough (1965), Miller und Charles (1991), Resnik (1995) und Charles (2001) herangezogen werden und mit den Ergebnissen des hier beschriebenen Experiments verglichen werden. Alle vier Autoren benutzen das Experimentdesign von Rubenstein und Goodenough (1965), und es gibt eine Menge von 28 Wortpaaren, die in allen vier zitierten Arbeiten untersucht werden. Für diese 28 Wortpaare (vgl.: Tabelle 4.1) ergibt sich ein  $W_{Kendall}$ -Wert von 0,892. Die  $W_{Kendall}$  Werte für die Korrelation mit den GS sind für  $FP_{controll}$  0,697 und für  $FP_{random}$  0,789. Diese drei Werte sind höchst signifikant. Ein signifikanter Unterschied zwischen der Beurteilung von Wort-Ähnlichkeit und Frame-Ähnlichkeit kann nicht festgestellt werden.

Es muss allerdings berücksichtigt werden, dass bei den angesprochenen Experimenten nicht Wortbedeutungen miteinander verglichen wurden, sondern Wörter. Wörter können jedoch lexikalisch ambig sein und es gibt keine Möglichkeit zu überprüfen, welche Wortbedeutungen die Versuchspersonen beurteilt haben. Beim Vergleich von Frames sind die Situationen bzw. die Bedeutung des Frames vollständig disambiguiert. Dieser Vergleich liefert folglich nur einen Anhaltspunkt für die Abschätzung des Schwierigkeitsgrades der Problemstellung.

## 4.2. Möglichkeiten einer Definition semantischer Ähnlichkeit auf der Basis von FrameNet

Zusammenfassend sollen hier die Möglichkeiten von FN als Ressource und die sich daraus ergebenden Forderungen an ein automatisches FN-Ähnlichkeitsmaß betrachtet werden.

Synonympaare		Charles (2001)	Rubenstein und Goodenough (1965)	Miller und Charles (1991)	Resnik (1995)
midday	noon	4.00	3.94	3.42	12.3925
automobile	car	3.89	3.92	3.92	8.0411
gem	jewel	4.00	3.94	3.84	14.9286
boy	lad	3.62	3.82	3.76	8.4240
coast	shore	3.24	3.6	3.70	10.8076
journey	voyage	3.64	3.58	3.84	6.7537
magician	wizard	3.38	3.21	3.50	13.6656
furnace	stove	2.97	3.11	3.11	1.7135
bird	crane	2.91	2.63	2.97	9.3139
food	fruit	2.74	2.69	3.08	5.0076
bird	cock	2.72	2.63	3.05	9.3139
asylum	madhouse	3.10	3.04	3.61	15.6656
brother	lad	2.17	2.41	1.66	2.9355
tool	implement	2.48	3.66	2.95	6.0787
brother	monk	2.02	2.74	2.82	2.9683
coast	forest	2.00	0.85	0.42	0.0000
car	journey	1.70	1.55	1.16	0.0000
noon	string	0.95	0.04	0.08	0.0000
monk	oracle	1.58	0.91	1.10	2.9683
food	rooster	1.48	1.09	0.89	1.0105
coast	hill	1.14	1.26	0.87	6.2344
crane	implement	1.31	2.37	1.68	2.9683
rooster	voyage	0.33	0.04	0.08	0.0000
monk	slave	1.26	0.57	0.55	2.9683
forest	graveyard	1.24	1.00	0.84	0.0000
lad	wizard	1.07	0.99	0.42	2.9683
chord	smile	1.03	0.02	0.13	2.3544
glass	magician	1.05	0.44	0.11	1.0105

Tabelle 4.1.: Ergebnisse Wort-Ähnlichkeitsexperimente

Die Pilotstudie hat gezeigt, dass semantische Ähnlichkeit zwischen Frames mit hinreichender Verlässlichkeit beurteilt werden kann. Hiervon ausgehend ist zu untersuchen, welche Informationen aus FN am besten zur Modellierung eines Ähnlichkeitsmaßes geeignet sind. Als Informationen stellt FN die Datenbank der Frames und die FN-Hierarchie zur Verfügung.

Die FN-Hierarchie ist dabei eine Schlüsselkomponente. Die Nutzung der FN-Hierarchie allein, als Basis für rein pfadbasierte Ansätze, ist jedoch nicht zu empfehlen. Die FN-Hierarchie ist kein zusammenhängender Graph. Es ist also nicht sichergestellt, ob ein Pfad zwischen zwei beliebigen Frames existiert. So sind in  $FP_{random}$  32,9% und in  $FP_{controll}$  16,13% der Frame-Paare in der FN-Hierarchie nicht verbunden. Dies bedeutet jedoch nicht, dass ein Frame-Paar von unähnlich ist. So gibt es beispielsweise keine Verbindung zwischen den Frames *Locale* und *Being\_located* oder zwischen *Increment* und *Undergo\_change*. Eine klare Anforderung an ein automatisches Frame-Ähnlichkeitsmaß ist es, solche Fälle von Unvollständigkeit zu berücksichtigen.

Ist es möglich einen Pfad zwischen zwei Frames zu finden, stellt die Gewichtung ein Problem dar. Die FN-Hierarchie ist kein kantengewichteter Graph, und die Definitionen der einzelnen Relationen deuten, wenn überhaupt, nur eine vage Ordnung an. Dies ist ein schwerwiegendes Problem, da für 94,19% der Pfade in  $FP_{random}$  und 87,74% der Pfade in  $FP_{controll}$  mehr als einen Relationstyp benutzt wird.

Die FN-Hierarchie kann also nicht die Einzige von FN genutzte Komponente sein, wenn man ein robustes Ähnlichkeitsmaß definieren will. Als weitere Informationsquellen können der Frame-Name, die Definition der Frames, die FEs und die LUs genutzt werden. So können die Frame-Definition oder die LUs einen Kontext liefern, der, in Anlehnung an wörterbuch-basierte Ansätze, genutzt wird, um Ähnlichkeit über Wortüberlappungen oder Vektorräume zu modellieren.

#### 4.2.1. Arbeitsdefinition

Eine Arbeitsdefinition von semantischer Ähnlichkeit für Frames kann dementsprechend lauten:

Zwei Frames sind sich semantisch ähnlich,  
wenn sie entweder in der FN-Hierarchie miteinander verbunden sind  
oder  
wenn es Überlappungen der FEs, LUs, in den Frame-Definitionen oder der Frame-Namen gibt.

## 5. Entwicklung eines FN basierten Ähnlichkeitsmaßes

Ziel dieses Kapitels ist es, ein robustes Frame-Ähnlichkeitsmaß zu entwickeln, das ausschließlich FN als Ressource nutzt. Hierfür werden verschiedene vorhandene Maße adaptiert und eigene FN-basierte Maße entwickelt. In diesem Kapitel werden 10 verschiedene Ähnlichkeitsmaße zur Bestimmung von Frame-Ähnlichkeit vorgestellt. Vier dieser Maße (siehe Kapitel 5.1) sind Adaptionen bereits vorhandener Ähnlichkeitsmaße. Da bislang keine Arbeiten zu diesem Thema veröffentlicht wurden, werden zuerst die bereits für WordNet entwickelten Ähnlichkeitsmaße von Hirst und St-Onge (1998), Leacock und Chodorow (1998) und Wu und Palmer (1994) adaptiert. Dies sind strukturbasierte Ähnlichkeitsmaße, die ausschließlich auf der WordNet-Taxonomie arbeiten. Außerdem wird mit dem Ansatz von Maguitman u. a. (2006) ein hybrider Ansatz adaptiert, der sowohl strukturbasierte, als auch informationsbasierte Komponenten aufweist (siehe Abschnitt 3.1.4.2). In der Evaluation soll anhand dieser Adaptionen untersucht werden, wie gut sich bereits entwickelte Lösungsansätze auf FN übertragen lassen. Offene Fragen sind dabei:

- Wie aussagekräftig sind Pfade in der FN-Hierarchie?
- Wie groß ist der Einfluss verschiedener Parameter, wie Pfadlänge oder die Tiefe des Pfades in der Hierarchie?
- Gibt es erkennbare Muster in Pfaden?
- Welche Relationen werden bevorzugt benutzt?

In Abschnitt 5.2 werden Maße präsentiert, die nicht allein auf der FN-Hierarchie basieren, sondern auch Informationen aus den Frames, wie Frame-Name, -Definition, etc. nutzen. Diese Maße werden unabhängig voneinander definiert, um das Potential jedes Informationsträgers besser beurteilen zu können. Wieviel die einzelnen Teile zur Bestimmung der Ähnlichkeit beitragen und ob die einzelnen Maße dieselbe Art von Ähnlichkeit modellieren, wird die Evaluation in Kapitel 6 zeigen.

Es wird darauf geachtet, dass sowohl die Adaption, als auch die neu entwickelten Maße konzeptuell einfach umgesetzt werden. Dies soll die Interpretation erleichtern und die Einflüsse verschiedener mathematischer Verfahren auf die Daten minimieren. Solche mathematischen Verfahren sind Dimensionsreduktionsverfahren, wie *Singular Value Decomposition* (Golub und Kahan, 2007).

## 5.1. Adaption verschiedener Ähnlichkeitsmaße

In diesem Abschnitt werden bereits vorhandene strukturbasierte Ansätze genutzt, um Frame-Ähnlichkeit zu modellieren. Mit dieser Vorgehensweise soll überprüft werden, ob bereits entwickelte Lösungsansätze für FrameNet übernommen werden können.

Die adaptierten Ähnlichkeitsmaße sind die Ähnlichkeitsmaße von Hirst und St-Onge (1998), Leacock und Chodorow (1998), Wu und Palmer (1994) und Maguitman u. a. (2005). Diese Maße repräsentieren verschiedene Ansätze. Das Ähnlichkeitsmaß von Hirst und St-Onge (1998) nutzt die einfache Pfadlänge, kombiniert mit einer Heuristik für gültige Pfade, zur Berechnung der Ähnlichkeit. Das Maß von Leacock und Chodorow (1998) berechnet die Ähnlichkeit mit Hilfe einer mit der Tiefe der Taxonomie skalierten Pfadlänge. Ebenfalls die Pfadlänge benutzt das Maß von Wu und Palmer (1994), wobei hier der kürzeste Pfad zum ersten gemeinsamen Superkonzept und die Tiefe dieses Superkonzepts in der Taxonomie den Grad der semantischen Ähnlichkeit bestimmen.

Durch die Skalierung der Pfadlänge begegnen Leacock und Chodorow (1998) und Wu und Palmer (1994) dem Problem, dass nicht jede Kante in der WordNet-Taxonomie die gleiche semantische Distanz modelliert. Allerdings nutzen beide Ähnlichkeitsmaße nur die Hyponym/Hypernym-Relation der Nomen-Taxonomie von WordNet. Im Gegensatz dazu nutzen die adaptierten strukturbasierten Ähnlichkeitsmaße alle in FrameNet zur Verfügung stehenden Relationen mit Ausnahme der See\_also-Relation. Die See\_also Relation ist deshalb ausgeschlossen, da sie keine semantische Relation ist, sondern einen Querverweis in FrameNet darstellt (siehe Kapitel 2.1.2.2 auf Seite 7). Ein weiterer Unterschied besteht darin, dass die LUs eines Frames verschiedene Wortarten besitzen können, wohingegen die WordNet-Ähnlichkeitsmaße nur für die Nomen-Taxonomie von WordNet definiert sind. Somit ist die lexikalische Abdeckung der adaptierten Ähnlichkeitsmaße größer.

Die Evaluation der adaptierten Maße wird jedoch zeigen, dass rein strukturbasierte Ansätze nicht ausreichen, um Frame-Ähnlichkeit adäquat zu beschreiben. Das graphbasierte Ähnlichkeitsmaß von Maguitman u. a. (2006) ist eine Weiterentwicklung des von Lin (1998) vorgestellten "Universal Similarity Measure". Maguitman u. a. skalieren hier den gefundenen Pfad mit dem Informationsgehalt der untersuchten Konzepte. Dieser Ansatz adressiert ebenfalls das Problem, dass Ontologien hinsichtlich ihrer Topologie und der Aussagekraft ihrer Knoten unbalanciert sein können (Maguitman u. a., 2005, S. 108).

Im Folgenden wird die Adaption der einzelnen Ähnlichkeitsmaße beschrieben.

### 5.1.1. Einfache Pfadlänge

Das von Hirst und St-Onge (1998) entwickelte Ähnlichkeitsmaß (siehe Kapitel 3.1.3.1) nutzt zur Berechnung der Ähnlichkeit die einfache Pfadlänge, kombiniert mit einer Heuristik für gültige Pfade. Die in Abbildung 5.1 dargestellten Pfadmuster beschreiben die gültigen Pfade. Die maximal zulässige Pfadlänge beträgt 5. Hirst und St-Onge (1998) unterscheiden zwischen vertikalen (auf- bzw. abwärtsgerichtete Pfeile) und horizontalen (waagerechte Pfeile) Kanten in der WordNet-Taxonomie. Bei der Adaption wer-

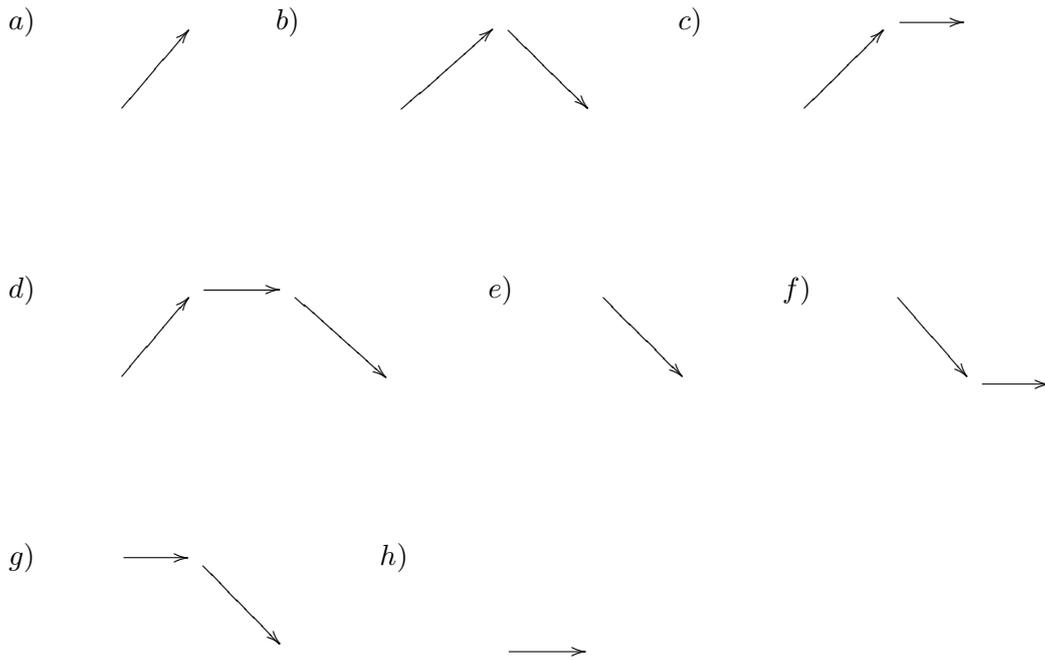


Abbildung 5.1.: Gültige Pfadmuster (Hirst und St-Onge, 1998)

den die Kanten der FN-Hierarchie entsprechend aufgeteilt. Die Inheritance-, Using- und Subframe-Kanten werden als vertikale Kanten gewertet, da sie Hyperonymie bzw. Meronymie modellieren. Die Perspective\_on-, Precedes-, Causative\_of- und Inchoative\_of-Kanten werden als horizontale Kanten gewertet. Die Pfadmuster werden übernommen. Da die Frame-zu-Frame-Relationen gerichtet sind, ist es möglich zwischen aufwärts und abwärts gerichteten Kanten zu unterscheiden. Die Beschränkung der Pfadlänge auf 5 wird nicht übernommen, da keine Evidenz in FrameNet für diesen Wert vorliegt. Stattdessen wird immer der kürzeste Pfad zwischen zwei Frames genutzt, der einem Pfadmuster aus Abbildung 5.1 entspricht. Die Pfeile beschreiben jeweils die Kantenart, horizontal bzw. vertikal, und die Ausrichtung der Kante, aufwärts bzw. abwärts. Abbildung 5.2 zeigt

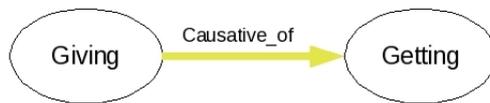


Abbildung 5.2.: Beispiel Pfadmuster h)

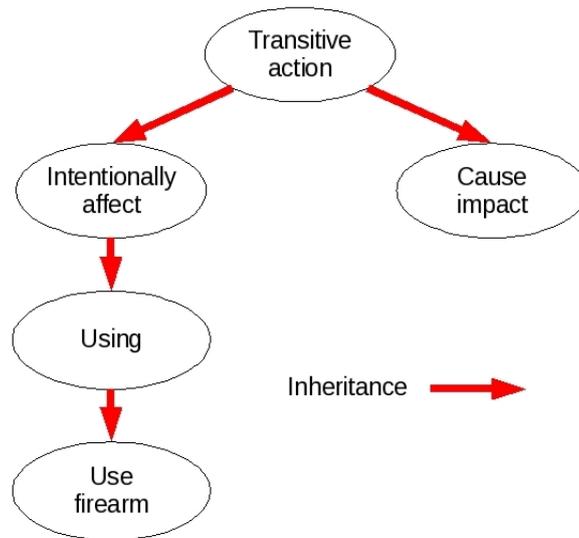


Abbildung 5.3.: Nicht akzeptierter Pfad

einen von Pfadmuster h) akzeptierten Pfad. Die Frames GIVING und GETTING sind durch eine „horizontale“ *Causative\_of*- Kante miteinander verbunden. Abbildung 5.3 veranschaulicht ein Beispiel für einen nicht durch die Pfadmuster in Abbildung 5.1 lizenzierten Pfad. Der Pfad verbindet die Frames USE\_FIREARM und CAUSE\_IMPACT. Der Teilpfad von USE\_FIREARM nach TRANSITIVE\_ACTION besteht nur aus abwärts gerichteten *Inheritance*-Kanten. Da es kein Pfadmuster gibt, dass abwärts gerichtete Kanten vor aufwärts gerichteten Kanten erlaubt, ist der Pfad von USE\_FIREARM nach CAUSE\_IMPACT nicht gültig.

### 5.1.2. Skalierte Pfadlänge

Leacock und Chodorow (1998) skalieren die Pfadlänge mit der Tiefe der Taxonomie (siehe Kapitel 3.1.3.2). Sie führen einen künstlichen Super-Wurzelknoten ein, der die einzelnen Teilgraphen der Nomen-Taxonomie von WordNet verbindet. Damit wird sichergestellt, dass alle Knoten in der Taxonomie verbunden sind. Bei der Adaption des Ähnlichkeitsmaßes wird kein künstlicher singulärer Wurzelknoten in die FN-Hierarchie eingefügt. Dies führt zwar zu Problemen bei der Skalierung der Pfadlänge mit der Tiefe der Taxonomie, aber ein Pfad, der eine Verbindung zwischen zwei Frames über einen künstlichen Wurzelknoten herstellt, würde keine semantische Ähnlichkeit mehr implizieren. Die einzelnen Komponenten der FN-Hierarchie besitzen unterschiedliche Tiefen. Die größte Tiefe der FN-Hierarchie ist 15 und wird zur Skalierung der Pfadlänge herangezogen.

Bei der Adaption des Ähnlichkeitsmaßes von Wu und Palmer (1994) (siehe Kapitel 3.1.3.3) auf FrameNet wurden die Konzepte durch Frames und die konzeptuelle Domäne

durch die FrameNet-Hierarchie ersetzt. Hier wird der jeweils kürzeste Pfad zum ersten gemeinsamen Super-Frame und der kürzeste Pfad von diesem Super-Frame zu einer möglichen Wurzel zur Berechnung der semantischen Ähnlichkeit benutzt.

### 5.1.3. Hybrider graph-basierter Ansatz

Im Folgenden beschreibe ich die Adaption des Ähnlichkeitsmaßes von Maguitman u. a. (2005) (siehe Kapitel 3.1.4.2). Für dieses Ähnlichkeitsmaß spricht, dass alle Arten von Relationen in einem Graphen berücksichtigt werden (siehe Gleichung 3.8 auf Seite 21). Dieses modulare Design erlaubt es, die Relationen des Graphen beliebig aufzuteilen und mit Hilfe der Adjazenzmatrix  $G$  wohldefiniert zusammenzufassen.

Die Frame-zu-Frame-Relationen werden in drei Kategorien aufgeteilt. In die erste Kategorie mit der höchsten Gewichtung werden die Inheritance-, Using- und Subframe-Relation eingeordnet. Die zweite Kategorie enthält die Perspective\_on-, Inchoative\_of- und Causative\_of-Relation. Die Precedes-Relation wird der dritten Kategorie zugeordnet. Die Gewichtung der einzelnen Kategorien nimmt von Kategorie zu Kategorie ab. Im folgenden werden die einzelnen Relationen der FN-Hierarchie formalisiert und die den Relationskategorien entsprechenden Matrizen definiert. Wir formalisieren die Frame-zu-Frame-Relationen als  $\rightarrow_R \subseteq E$ , für die  $x \rightarrow_R y$  gilt, genau dann wenn  $y$  von  $x$  aus mit der Relation  $R$  direkt erreicht werden kann und  $x$  und  $y$  adjazent sind.  $E$  steht für die Menge aller Kanten in der FN-Hierarchie.

Die Inheritance-Relation beschreibt die Menge aller Kanten  $\mathcal{I}$ , für die gilt:

$$\mathcal{I} = \{(v, w) \mid v \rightarrow_{\text{Inheritance}} w; \text{ mit } v, w \in V\} \quad (5.1)$$

Analog beschreiben die Mengen  $U, S, C, I, P, \mathcal{P}$  die Using-, Subframe-, Causative\_of-, Inchoative\_of-, Perspective\_on- und Precedes-Relation. Die Vereinigungsmenge  $RK_1 = \mathcal{I} \cup U \cup S$ , bestehend aus der Inheritance-, Using- und Subframe-Relation bildet die erste Relationskategorie. Die Menge  $RK_2 = C \cup I \cup P$  beschreibt die zweite Relationskategorie mit der Causative\_of-, Inchoative\_of- und Perspective\_on-Relation. Die dritte Relationskategorie  $RK_3 = \mathcal{P}$  entspricht der Precedes-Relation. Mit Hilfe der Relationskategorien können die den Kategorien entsprechenden Adjazenzmatrizen definiert werden.

Die Adjazenzmatrix  $A_1$  ist definiert als:

$$A_{RK_1}_{ij} = \begin{cases} 1 & , \text{ wenn } i = j \\ \alpha & , \text{ wenn } i \neq j \text{ und } (i, j) \in RK_1 \\ 0 & , \text{ sonst} \end{cases} \quad (5.2)$$

Die Adjazenzmatrizen für die Relationskategorien  $RK_2$  und  $RK_3$  sind analog definiert. Die Gewichtungen  $\alpha, \beta$  und  $\gamma$  wurden von mir auf 1, 0.75 und 0.5 festgelegt, da eine Gewichtung zur Errechnung von  $G_{FN}$  notwendig ist. Die Zusammenfassung der einzelnen, den Relationskategorien entsprechenden Matrizen zur Adjazenzmatrix  $G_{FN}$  (siehe Gleichung 5.3) führt dazu, dass die FN-Hierarchie in dieser Darstellung kein Multi-Graph

mehr ist, sondern zwei Knoten jeweils mit der Kante mit der höchsten Gewichtung verbunden sind.  $G_{FN}$  ist analog zu Maguitman u. a. (2005) definiert als:

$$G_{FN} = A_{RK\_1} \vee A_{RK\_2} \vee A_{RK\_3} \quad (5.3)$$

Die „fuzzy membership matrix“ (Maguitman u. a., 2005)  $W_{FN}$  wird durch

$$W_{FN} = A_{RK\_1}^* \odot G_{FN} \odot A_{RK\_1}^* \quad (5.4)$$

berechnet, wobei  $A_{RK\_1}^*$  die Hülle von  $A_{RK\_1}$  ist.

Die Berechnung der semantischen Ähnlichkeit erfolgt, wie in Gleichung 5.5 beschrieben, analog zu Maguitman u. a. (2005).

$$\sigma_S^{FN}(f_1, f_2) = \max_k \frac{2 \cdot \min(W_{FN_{k1}}, W_{FN_{k2}}) \cdot \log Pr_{FN}(f_k)}{\log(Pr_{FN}(f_1|f_k) \cdot Pr_{FN}(f_k)) + \log(Pr_{FN}(f_2|f_k) \cdot Pr_{FN}(f_k))} \quad (5.5)$$

Die Wahrscheinlichkeit  $Pr_{FN}(f_k)$  beschreibt die Wahrscheinlichkeit, dass ein beliebiger Frame auf einem Pfad mit Startknoten  $f_k$  liegt. Sie ist, wie folgt, definiert:

$$Pr_{FN}(f_k) = \frac{\sum_{f_j \in FN} (W_{FN_{kj}} \cdot |lu(f_j)|)}{\sum_{x \in FN} |lu(x)|} \quad (5.6)$$

$|lu(f)|$  beschreibt die Anzahl aller LUs des Frame  $f$ , wobei  $Pr_{FN}(f_i|f_k)$  ausdrückt, dass ein beliebiger Frame sowohl von Frame  $f_i$ , als auch von Frame  $f_k$  erreichbar ist. Die Definition lautet:

$$Pr_{FN}(f_i|f_k) = \frac{\sum_{f_j \in FN} (\min(W_{ij}, W_{kj}) \cdot |lu(j)|)}{\sum_{f_j \in FN} (W_{FN_{kj}} \cdot |lu(f_j)|)} \quad (5.7)$$

## 5.2. FN-Ähnlichkeitsmaße

In diesem Abschnitt wird untersucht, inwieweit mit durch Frames zur Verfügung gestellten Informationen Frame-Ähnlichkeit modelliert werden kann. Im Gegensatz zum vorigen Kapitel liegt nun der Fokus auf den im Frame selbst enthaltenen Informationen und nicht auf der FN-Hierarchie. Ein Frame stellt als Informationen den Frame-Namen, die Frame-Definition, eine Menge von FEs und eine Menge von LUs bereit. Da nicht mehr direkt auf der FN-Hierarchie gearbeitet wird, fallen einige Probleme weg, wie die lückenhafte Abdeckung oder die Unverbundenheit der FN-Hierarchie. Das „sparse-data“-Problem bleibt jedoch bestehen, verlagert sich aber auf eine andere Ebene. Es gibt insgesamt nur 795 Frames in FN (Release 1.3) und die Frame-internen Informationen stellen jeweils nur kleine Datenmengen zur Verfügung. Die Ursachen hierfür liegen einerseits auf konzeptueller Seite, wie zum Beispiel die geringe Länge der FN-Definitionen, andererseits ist die Arbeit an FN noch nicht abgeschlossen und die Zahl der Frames und annotierten LUs wächst beständig.

Traditionelle Ansätze, die diese Frame-internen Informationen verwerten können, sind Distanzmaße, distributionelle Ansätze sowie informationstheoretische Ansätze. Für die verschiedenen Frame-Komponenten wurden verschiedene Ansätze zur Modellierung von Frame-Ähnlichkeit verwendet. Im folgenden werden die einzelnen Frame-Komponenten und die verwendeten Mittel zur Berechnung von Frame-Ähnlichkeit beschrieben.

### 5.2.1. Frame-Name

Frame-Namen sollen die durch den Frame erfasste aussersprachliche Szene knapp umschreiben. Sie geben der Szene eine informelle Struktur und eine Perspektive (Fillmore u. a., 2003). Der Unterschied zwischen den beiden Frames `COMMERCE_BUY` und `COMMERCE_SELL` ist an den Namen erkennbar. Beide Frames beschreiben eine Situation, in der gehandelt wird. Sie tun dies jedoch aus unterschiedlichen Perspektiven. `COMMERCE_BUY` bezeichnet eine Kauf-Situation und `COMMERCE_SELL` eine Verkauf-Situation. Diese Wortähnlichkeit kann ein erstes Indiz für eine Frame-Ähnlichkeit sein. Ein Distanzmaß auf Wortebene kann die Ähnlichkeit zwischen `COMMERCE_BUY` und `COMMERCE_SELL` erfassen.

Als Maß zur Bestimmung der Wortähnlichkeit wird das Jaro-Winkler-Distanzmaß benutzt. Die Distanz zweier Wörter wird bestimmt durch die Zahl und die Reihenfolge der gemeinsam vorkommenden Buchstaben der zwei Wörter, korrigiert durch die Länge des gemeinsamen Präfixes<sup>1</sup> der beiden Wörter (Jaro, 1995; Winkler, 1999). Dieses Maß ist effizient in der Berechnung der Distanz, besonders für kurze Wörter (Cohen u. a., 2003).

Problematisch für diesen Ansatz sind Fälle, in denen die Namen der Frames unterschiedlich sind, aber semantisch ähnliche Szenen beschreiben. Ein sehr schönes Beispiel hierfür sind der `TRIAL`- und der `SENTENCING`-Frame. Beides sind Subframes des `CRIMINAL_PROCESS`-Frames und es besteht eine zeitliche Relation zwischen beiden Frames. Die Jaro-Winkler-Distanz für diese beiden Frames beträgt 0, was auf eine geringe Ähnlichkeit hinweist. Im Gegensatz dazu beträgt die Jaro-Winkler-Distanz von `COMMERCE_BUY` und `COMMERCE_SELL` 0.926, was auf eine große Ähnlichkeit hinweist. Für eine ausführliche Diskussion siehe Kapitel 6.

### 5.2.2. Frame-Definition

Frame-Definitionen umschreiben die Bedeutung des Frames. Sie stellen einen Text mit einer durchschnittlichen Wortlänge von 74 Wörtern bereit. Diese Wörter werden genutzt, um mit einem „word overlap“-Ansatz die Ähnlichkeit zu modellieren. Die Hypothese ist, dass je größer die Übereinstimmung im Vokabular der Definitionen sind, desto ähnlicher sind sich die Frames. Dieser Ansatz wird in Wörterbuch-basierten Systemen bereits erfolgreich genutzt (siehe Stevenson und Wilks, 2001). Zur Berechnung des „word overlap“ der Definitionen werden die Stämme der in den Definitionen enthaltenen Nomen, Verben und Adjektive herangezogen. Problematisch für diesen Ansatz ist, dass die Lexikographen des FrameNet-Projektes kein einheitliches Definitionsvokabular benutzen, wie zum

---

<sup>1</sup> Als Präfix wird eine Kette von gleichen Zeichen am Wortanfang bezeichnet.

<p><b>CHANGE_OF_PHASE</b>  In this <b>frame</b> an <b>Undergoer</b> <b>undergoes</b> a <b>change</b> of <b>phase</b>.</p> <p><b>ALTERED_PHASE</b>  In this <b>frame</b> an <b>Undergoer</b> is in a <b>phase brought</b> about after a <b>change</b> of <b>phase</b>.  This <b>can be</b> its <b>inherent phase</b> or <b>indicate</b> that it is <b>returned</b> to its <b>inherent state</b>.</p>
---

Abbildung 5.4.: Frame-Definitionen CHANGE\_OF\_PHASE und ALTERED\_PHASE

Beispiel bei der Umschreibung der Wortbedeutungen im Longman Dictionary of Contemporary English (Summers, 2003). So können ähnliche Frames durch unterschiedliche Formulierungen in ihren Definitionen als unähnlich bewertet werden. Ein Beispiel hierfür ist das Frame-Paar CHANGE\_OF\_PHASE - ALTERED\_PHASE. Die Definitionen (siehe Abbildung 5.4) weisen nur eine geringe Übereinstimmung auf, obwohl sie sehr ähnliche Konzepte beschreiben.

Deshalb nutzt ein zweites Maß zusätzlich eine Abbildung dieser Wörter auf WordNet Synsets und berechnet die Übereinstimmung der in den beiden Frame-Definitionen enthaltenen Synsets. Es wird sich jedoch in der Evaluation zeigen, dass auch die Erweiterung um Synsets die Fähigkeit des Maßes nicht verbessert, Frame-Ähnlichkeit zu beschreiben.

### 5.2.3. Frame Elemente

Die Frame-Elemente beschreiben die im Frame beteiligten Rollen. Parallel zu den Frame-zu-Frame-Relationen gibt es eine FE-zu-FE-Relation. Diese Relation bildet FEs eines Super-Frames auf FEs eines Sub-Frames ab.

Abbildung 5.5 zeigt ein Beispiel für durch Relationen verbundene FEs. Der Frame CHANGE\_OF\_PHASE\_SCENARIO ist sowohl mit dem Frame CHANGE\_OF\_PHASE, als auch mit dem Frame ALTERED\_PHASE über eine Subframe-Relation verbunden (in Abbildung 5.5 der unterbrochene blaue Pfeil). Parallel zu diesen Subframe-Relationen sind die Rollen des Super-Frame CHANGE\_OF\_PHASE\_SCENARIO mit den jeweiligen Rollen der Sub-Frames CHANGE\_OF\_PHASE und ALTERED\_PHASE verbunden. So ist in Abbildung 5.5 zu erkennen, dass die *Undergoer*-Rolle von CHANGE\_OF\_PHASE\_SCENARIO sowohl mit der *Undergoer*-Rolle von ALTERED\_PHASE, als auch mit der *Undergoer*-Rolle von CHANGE\_OF\_PHASE verbunden ist. Das heißt, dass die *Undergoer*-Rolle des CHANGE\_OF\_PHASE-Frames dieselbe Rolle beschreibt, wie die *Undergoer*-Rolle des ALTERED\_PHASE Frames. Allgemein bedeutet dies, dass zwei Rollen zwei verschiedener Frames, die über einen Pfad von FE-zu-FE-Relationen verbunden sind, denselben Mitspieler kennzeichnen.

Dieser Umstand kann nun genutzt werden, um Frame-Ähnlichkeit zu beschreiben. Die Hypothese lautet, dass zwei Frames sich ähnlich sind, wenn sich möglichst viele ihrer Rollen aufeinander abbilden lassen. Je mehr solcher Abbildungen möglich sind, desto

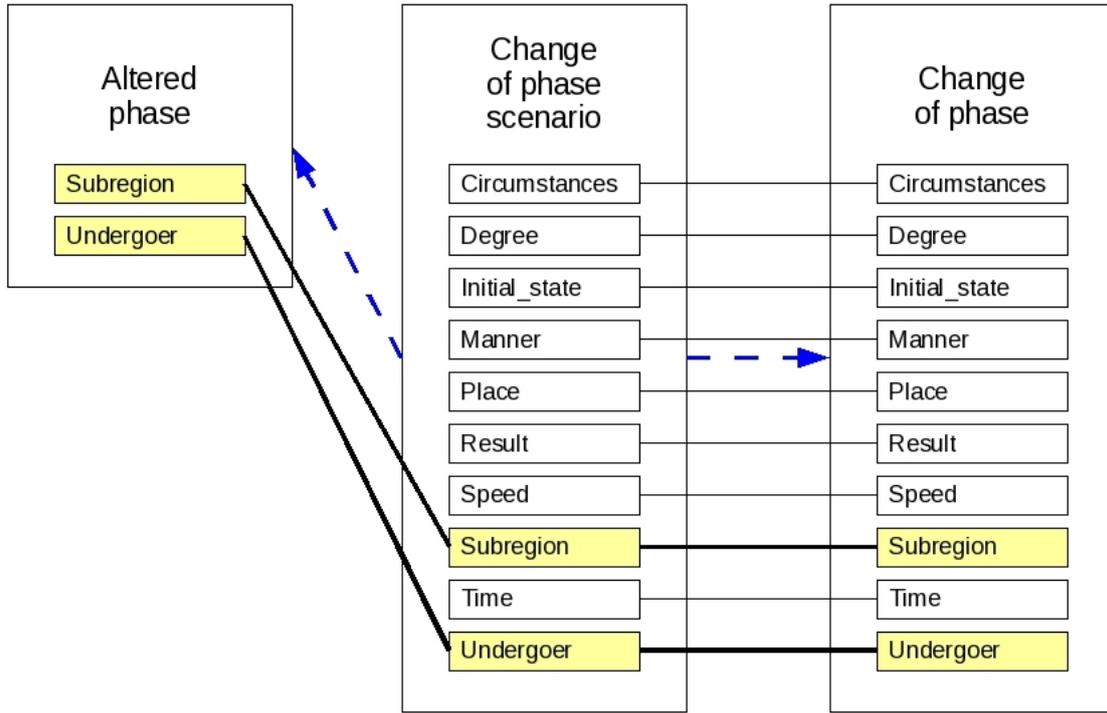


Abbildung 5.5.: FE-zu-FE-Relationen

größer ist die Ähnlichkeit zwischen den Frames. Formal ist dieses Maß, wie folgt, definiert:

$$sim_{FE}(F_1, F_2) = \frac{|FE_1 \cap FE_2|}{\max(|FE_1|, |FE_2|)} \quad (5.8)$$

$F_1$  und  $F_2$  stehen hierbei für beliebige Frames.  $FE_1$  und  $FE_2$  beschreiben die Menge aller Rollen von  $F_1$  bzw.  $F_2$ .

Eine weitere Hypothese besagt, dass die Ähnlichkeit zunimmt, wenn die FEs nicht nur FE-zu-FE-Relationen verbunden sind, sondern auch einen identischen Namen tragen. Formal definiert sich dieses Maß wie folgt:

$$sim_{FEName}(F_1, F_2) = \frac{|\{(fe_1, fe_2) | (fe_1, fe_2) \in FE_1 \cap FE_2 \wedge name(fe_1) = name(fe_2)\}|}{\max(|FE_1|, |FE_2|)} \quad (5.9)$$

wobei  $name(fe)$  den Namen des FE zurückgibt.

Die Ähnlichkeit eines Pfades zwischen Frame  $F_i$  und Frame  $F_j$  errechnet sich, wie folgt:

$$sim_{FEPfad}(F_i, F_j) = \prod_{(f_v, f_w) \in pfad(F_i, F_j)} sim_{FE}(f_v, f_w) \quad (5.10)$$

cold (adj)	frigid (adj)	scalding (adj)
cool (adj)	hot (adj)	temperature (n)
freezing (adj)	lukewarm (adj)	tepid (adj)

Abbildung 5.6.: LUs TEMPERATURE-Frame

$sim_{FE}$  steht hierbei für eines der beiden Ähnlichkeitsmaße  $sim_{Fe}$  oder  $sim_{FeName.pfad}(F_x, F_y)$  beschreibt die Menge der Kanten auf dem Pfad in der FN-Hierarchie zwischen den Frames  $F_x$  und  $F_y$ .

Im Gegensatz zu den über den Frame-Namen und die Frame-Definition definierten Ähnlichkeitsmaßen ist es für die Abbildung der FE notwendig, die FN-Hierarchie zu benutzen. FE sind nur verbunden, wenn auch ihre Frames über Frame-zu-Frame-Relationen verbunden sind. So ist dieses Ähnlichkeitsmaß ein strukturbasiertes Ähnlichkeitsmaß. Die Evaluationsergebnisse zeigen, dass dieses Maß die besten Ergebnisse der in diesem Abschnitt beschriebenen Maße liefert. Dies ist ein weiterer Hinweis auf das Potential der FN-Hierarchie, die Ähnlichkeit komplexer Situationen zu modellieren.

#### 5.2.4. Lexical Units

Die LUs eines Frames sind die mit dem Frame assoziierten Wörter. Frames beschreiben meist mehrere LUs. Abbildung 5.6 zeigt beispielhaft die LUs des TEMPERATURE-Frames.

Die Hypothese besagt, je mehr LUs in zwei Frames übereinstimmen, desto größer ist die Ähnlichkeit. In Abbildung 5.7 sind die LUs der Frames TEMPERATURE und AMBIENT\_TEMPERATURE gegenübergestellt. Die Frames besitzen 9 bzw. 7 LUs. 4 von diesen LUs kommen in beiden Frames vor.

FN enthält in Version 1.3 10195 LUs in 795 verschiedenen Frames. Berücksichtigt man Mehrfachvorkommen von LUs (406 Fälle) in verschiedenen Frames und Mehrwort-LUs (3383 Fälle), so erhält man 6407 verschiedene einfache LUs (siehe Pennacchiotti u. a., 2008).

Im folgenden werden zwei Ähnlichkeitsmaße vorgestellt, die LUs benutzen. Der erste Ansatz ist ein reiner LU-Overlap-Ansatz, der die LUs auf einfache und direkte Weise nutzt, um den Prozentsatz übereinstimmender LUs zu berechnen. Im zweiten Ansatz dienen die LUs als Grundlage für die Berechnung von Kookkurenz-Vektoren.

Der reine LU-Overlap-Ansatz ist formal, wie folgt, definiert:

$$sim_{LU_{simple}}(F_1, F_2) = \frac{|LU_{F_1} \cap LU_{F_2}|}{\max(|LU_{F_1}|, |LU_{F_2}|)} \quad (5.11)$$

Die mit 406 von 10196 LUs nur geringe Zahl von mehrfach vorkommenden LUs, zeigt schon eine deutliche Schwäche des einfachen  $sim_{LU_{simple}}$ -Ähnlichkeitsmaßes auf, nämlich die geringe Abdeckung. Da bei diesem Ähnlichkeitsmaß die FN-Hierarchie nicht genutzt

TEMPERATURE	AMBIENT_TEMPERATURE
	chilly (adj)
<b>cold (adj)</b>	<b>cold (adj)</b>
<b>cool (adj)</b>	<b>cool (adj)</b>
freezing (adj)	
frigid (adj)	
<b>hot (adj)</b>	<b>hot (adj)</b>
lukewarm (adj)	
scalding (adj)	
	scorcher (n)
<b>temperature (n)</b>	<b>temperature (n)</b>
tepid (adj)	
	warm (adj)

Abbildung 5.7.: LUs TEMPERATURE - AMBIENT\_TEMPERATURE

wird, kann keine Ähnlichkeit zwischen beispielsweise den verbundenen Frames FLEEING und ESCAPING modelliert werden, da dort keine gemeinsamen LUs existieren. Um dieses Problem zu lösen, wurde mit einem weiteren Ähnlichkeitsmaß experimentiert.

Das vektorraum-basierte Maß nutzt das von Pennacchiotti u. a. (2008) entwickelte LsaWord-Modell, um die Ähnlichkeit zweier Frames zu bestimmen. Pennacchiotti u. a. (2008) haben gezeigt, dass mit Hilfe der LUs Vektormodelle von Frames konstruiert werden können. Die Frame-Ähnlichkeit errechnet sich dann gemäß Formel 5.12:

$$sim_{LU_{semSpace}}(F_1, F_2) = \cos(\angle(\vec{F}_1, \vec{F}_2)) \quad (5.12)$$

$\vec{F}_1, \vec{F}_2$  bezeichnen die mit Hilfe der LUs erstellten Kookkurenzvektoren der Frames  $F_1$  und  $F_2$ . Der Ähnlichkeitswert der beiden Frames ist die Kosinus-Ähnlichkeit der beiden Vektoren  $\vec{F}_1$  und  $\vec{F}_2$ . Das zuvor für das einfache Ähnlichkeitsmaß  $sim_{LU_{simple}}$  festgestellte Abdeckungsproblem ist in  $sim_{LU_{semSpace}}$  gelöst, da jeder LUs enthaltende Frame durch einen Vektor repräsentiert wird. Somit kann auch zwischen Vektoren, die keine LUs gemein haben, Frame-Ähnlichkeit berechnet werden.

Die Evaluation wird zeigen, dass beide Maße weniger gut geeignet sind, um Frame-Ähnlichkeit zu beschreiben.

### 5.3. Zusammenfassung

10 verschiedene Ähnlichkeitsmaße zur Bestimmung von Frame-Ähnlichkeit wurden in diesem Kapitel vorgestellt. Vier dieser Maße (siehe Kapitel 5.1) sind Adaptionen bereits vorhandener Wort-Ähnlichkeitsmaße. Sechs weitere Maße sind speziell für FN entwickelt

worden. In Kapitel 6 werden die in diesem Kapitel besprochenen Maße evaluiert. Die Evaluation wird zeigen, dass erstens die FN-Hierarchie trotz aller Probleme eine gute Basis zur Definition eines Frame-Ähnlichkeitsmaßes bietet, und zweitens frame-basierte Maße, die die FN-Hierarchie nutzen, Frame-Ähnlichkeit am besten approximieren.

## 6. Evaluation

In diesem Kapitel werden die in Kapitel 5 vorgestellten Ähnlichkeitsmaße evaluiert. Ein Hauptziel der Evaluation wird es sein, zu untersuchen, wie gut Frame-Ähnlichkeit mit Hilfe von FN modelliert werden kann. Besonderes Augenmerk liegt dabei auf den Informationen, die durch die Frames selbst und die FN-Hierarchie zur Verfügung gestellt werden. Der Vergleich mit den adaptierten WordNet-Ähnlichkeitsmaßen soll zeigen, wie gut bereits entwickelte Ähnlichkeitsmaße Frame-Ähnlichkeit modellieren können. Dies wird untersucht mit Hilfe der in Kapitel 4 vorgestellten Gold-Standards. Ein abschließender Vergleich der Ähnlichkeitsmaße untereinander soll zeigen, ob die Maße alle die gleiche Art von Ähnlichkeit modellieren.

Dieses Kapitel ist wie folgt strukturiert: Abschnitt 6.1 präsentiert die Ergebnisse der Evaluation der einzelnen Ähnlichkeitsmaße mit dem Gold-Standard. In Abschnitt 6.2 folgt eine qualitative Analyse der Experimentergebnisse. Das Augenmerk liegt hierbei sowohl auf dem Verhalten der Ähnlichkeitsmaße in dichten und weniger dichten Regionen der FN-Hierarchie, als auch auf dem Einfluß der Pfadstruktur in der FN-Hierarchie. In Abschnitt 6.3 werde ich die Ergebnisse der Evaluation zusammenfassen.

### 6.1. Leistung der Ähnlichkeitsmaße

In diesem Abschnitt wird die Fähigkeit der Ähnlichkeitsmaße untersucht, Frame-Paare nach ihrer Ähnlichkeit zu ordnen. Als Gold-Standard stehen die in Kapitel 4 beschriebenen Ordnungen zur Verfügung. Diese beiden Referenzordnungen erlauben es uns, die Leistung der Frame-Ähnlichkeitsmaße sowohl in gut ausgearbeiteten Teilen von FrameNet als auch in weniger gut ausgearbeiteten Teilen zu untersuchen. In Abschnitt 6.1.1 wird das Experimentdesign beschrieben und die Ergebnisse präsentiert. In Abschnitt 6.1.2 werden abschließend die Ergebnisse ausgewertet und diskutiert.

#### 6.1.1. Experimentdesign und Ergebnisse

In dem hier durchgeführten Experiment wurden alle in Kapitel 5 vorgestellten Maße benutzt, um Frame-Paare nach ihrer Ähnlichkeit zu ordnen. Als Eingabe für die Versuche dienen die beiden in Kapitel 4.1 vorgestellten Gold-Standards (kurz: GS). Dabei stehen durch das Korpusdesign der GS Referenzdaten aus unterschiedlich dichten Regionen von FN zur Verfügung. Datensatz  $FP_{control}$  stellt Frame-Paare aus dichteren Regionen zur Verfügung. Datensatz  $FP_{random}$  stellt Frame-Paare aus weniger verbundenen Regionen von FN bereit. Das Versuchssystem besitzt zwei Konfigurationen: eine Konfiguration für die struktur-basierten Ähnlichkeitsmaße und eine Konfiguration für

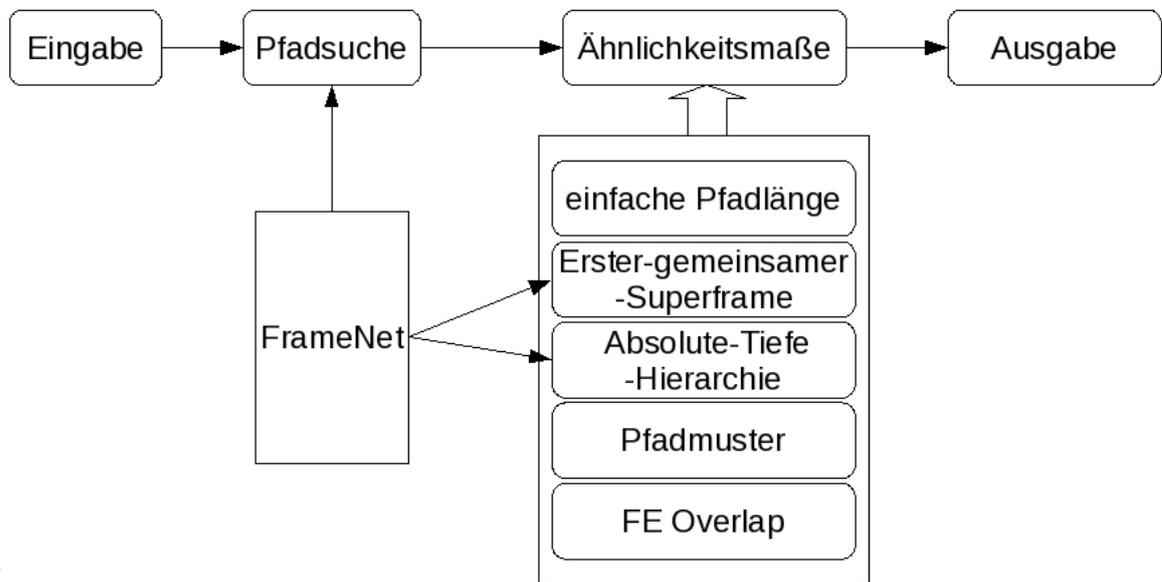
die nicht struktur-basierten Ähnlichkeitsmaße. Abbildung 6.1 zeigt schematisch den Aufbau des Versuchssystems. Die verschiedenen Konfiguration sind mit I und II bezeichnet. Die Eingabe ist eine Menge von Frame-Paaren. In Konfiguration I wird zuerst ein Pfad zwischen den Frame-Paaren berechnet, bevor in der Ähnlichkeitsmaß-Komponente die Ähnlichkeit errechnet wird. Die Ausgabe ist eine Menge von Frame-Paaren, gewichtet mit Ähnlichkeit. Konfiguration II unterscheidet sich von Konfiguration I nur dadurch, dass keine Pfadsuch-Komponente vorhanden ist. Die Pfadsuche wird in dieser Konfiguration nicht benötigt, da die Ähnlichkeitsmaße nicht struktur-basiert arbeiten, bzw. im Falle von Maguitman u. a. nicht direkt auf die FN-Hierarchie zugreifen. Das Design der Ähnlichkeitsmaß-Komponente erlaubt es, die einzelnen Maße innerhalb ihrer Konfiguration auszutauschen und so verschiedene Maße zu testen.

Die resultierenden Ordnungen wurden anschließend mit den GS verglichen und die Korrelation errechnet. Zur Errechnung der Korrelation wurde das Kendall  $\tau$ -Maß benutzt. Bei den Versuchen wurde FrameNet Release 1.3 verwendet. Tabelle 6.1 zeigt die Ergebnisse der Durchläufe mit den beiden Gold-Standards<sup>1</sup>. Angegeben sind die Korrelationswerte der einzelnen Maße mit dem jeweiligen Gold-Standard.

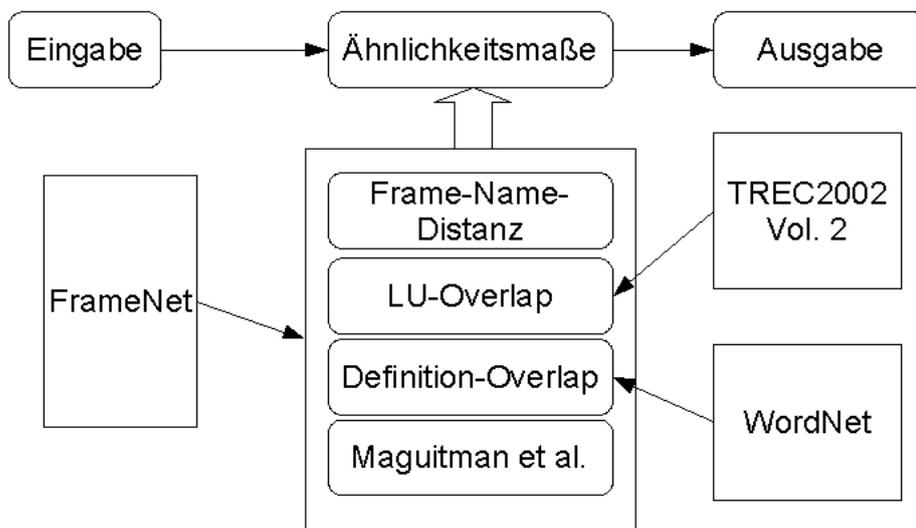
	$GS_{random}$	$GS_{control}$
struktur-basiert		
einfache Pfadlänge	0,199	0,473
WN-Adaptionen		
Erster-Gemeinsamer-Superframe	0,153	0,320
Absolute-Tiefe-Hierarchie	0,199	0,473
Pfadmuster	0,237	0,266
FN		
FE-Overlap Connected	0,221	0,389
FE-Overlap Name	0,263	0,332
Hybrid		
Maguitman et al.	<b>0,036</b>	0,360
nicht struktur-basiert		
Definition Word-Overlap Stem	<b>0,066</b>	0,310
Definition Word-Overlap SynSet	<b>0,039</b>	0,305
LU-Overlap Stem	0,093	0,285
LU-Overlap Semantic Space	<b>0,021</b>	<b>0,043</b>
Frame Name Distanz	<b>0,080</b>	0,390

Tabelle 6.1.: Korrelation der Ähnlichkeitsmaße mit den Gold-Standards

<sup>1</sup>Fett rot eingefärbte Korrelationswerte sind nicht signifikant.



I



II

Abbildung 6.1.: Architektur Versuchssystem

### 6.1.2. Diskussion der Resultate

Man sieht, dass die Ähnlichkeitsmaße für Frame-Paare aus  $FP_{control}$  eine höhere Korrelation mit dem GS aufweisen als Paare aus  $FP_{random}$ . Die Werte reichen von  $\tau_{sim_{LU-SemSpace-random}} = 0,021$  bis zu  $\tau_{sim_{einfache-Pfadlänge-control}} = 0,473$ . Das Signifikanzniveau für die  $z$ -Werte von 0,388, beziehungsweise 8,741, liegt zwischen  $\alpha < 0,352$  und  $\alpha < 0,00003$ . Die Standardabweichung  $\sigma^2$  beträgt 0,0029. Die getesteten Maße sind somit geeignet, Frame-Ähnlichkeit zu berechnen. Dies gilt für alle Maße, für die die Korrelation der Ordnungen der Frame-Paare mit dem GS  $\tau > 0,09$  ist. Die Werte sind signifikant mit  $\alpha < 0,0485$ . Die Korrelation ist für folgende Ähnlichkeitsmaße nicht signifikant:

- Maguitman et al. mit  $GS_{random}$
- Definition Word-Overlap Stem mit  $GS_{random}$
- Definition Word-Overlap SysnSet mit  $GS_{random}$
- LU-Overlap Semantic Space mit beiden  $GS$
- Frame-Name-Distanz mit dem  $GS_{random}$

Alle übrigen Maße korrelieren signifikant mit den Gold-Standards. Sie können also zur Beschreibung von Frame-Ähnlichkeit genutzt werden. Das beste Frame-Ähnlichkeitsmaß für weniger gut ausgearbeitete Teile von FN ist das Maß, das die Übereinstimmung des Rolleninventars -unter gleichzeitiger Berücksichtigung des Rollennamens- berechnet. Für gut ausgearbeitete Teile approximiert die einfache Pfadlänge die Frame-Ähnlichkeit am Besten.

Für alle Ähnlichkeitsmaße ist die Korrelation mit dem  $GS_{control}$  höher als mit  $GS_{random}$ . Der größte Unterschied ist bei dem „einfache Pfadlänge“-Maß mit 0,274 festzustellen. Der durchschnittliche Unterschied zwischen den  $\tau$ -Werten beträgt für die struktur-basierten Maße, ohne die adaptierten Maße von Maguitman u. a.<sup>2</sup> und Leacock und Chodorow<sup>3</sup>, jedoch nur 0,072. Die Methoden zur Skalierung der Pfadlänge, die in den übrigen Ähnlichkeitsmaßen zur Anwendung kommen, reduzieren die Auswirkungen des *sparse-data* -Problems also wirksam. Die kleinsten Unterschiede in den Korrelationswerten mit  $GS_{random}$  und  $GS_{control}$  liefern das adaptierte Ähnlichkeitsmaß von Hirst und St-Onge (1998) und das FE-Overlap-Maß mit der Überprüfung auf Namensgleichheit. Gleichzeitig ist festzustellen, dass die Korrelationswerte aller Maße deutlich von dem Spitzenwert des „einfache Pfadlänge“-Maßes abweichen.

Die fünf Ähnlichkeitsmaße, die nicht die FN-Hierarchie nutzen, korrelieren weniger stark mit den Gold-Standards, als die struktur-basierten Maße. Bei drei der fünf Maße ist mindestens ein  $\tau$ -Wert so niedrig, dass die Korrelation nicht signifikant ist (siehe fette,rote Werte in Tabelle 6.1).

---

<sup>2</sup>Das Ähnlichkeitsmaß von Maguitman et al. bleibt unberücksichtigt, da die FN-Hierarchie nicht direkt genutzt wird (siehe Kapitel 5.1.3 auf Seite 36).

<sup>3</sup>Das Maß liefert identische Korrelationswerte mit dem „einfache Pfadlänge“-Maß. (Siehe Abschnitt 6.2.1.2 auf Seite 57)

Ein weiteres überraschendes Ergebnis ist der hohe Korrelationswert des Frame-Name-Distanz-Maßes mit dem  $GS_{control}$ , das mit 0,39 den zweit-höchsten Korrelationswert aufweist. Da allerdings der Korrelationswert mit  $GS_{random}$  nicht signifikant ist, sollte dieses Maß nicht benutzt werden, um Frame-Ähnlichkeit zu modellieren.

## 6.2. Qualitative Untersuchung

In diesem Abschnitt werden die Ergebnisse der einzelnen Ähnlichkeitsmaße genauer betrachtet. Unter anderem werden wir auf die Fragen aus Kapitel 5 eingehen.

- Wie groß ist der Einfluss verschiedener Parameter, wie Pfadlänge oder die Tiefe des Pfades in der Hierarchie?
- Welche Relationen werden bevorzugt benutzt?
- Gibt es erkennbare Muster in Pfaden?

Zuerst werden die struktur-basierten, dann die nicht struktur-basierten Ähnlichkeitsmaße untersucht. Abschließend wird zusammenfassend auf noch offene Fragen eingegangen.

### 6.2.1. Struktur-basierte Ähnlichkeitsmaße

Der Schwerpunkt dieses Abschnittes ist es, zu klären, welche Eigenschaften der Pfade in der FN-Hierarchie besonders zur Modellierung von Frame-Ähnlichkeit geeignet sind.

#### 6.2.1.1. Pfadlänge

**Einfache Pfadlänge** Die einfache Pfadlänge ist eine der einfachsten Möglichkeiten, Ähnlichkeit in Netzen zu bestimmen. Dieser Ansatz liefert für unsere Problemstellung, die Ähnlichkeit zwischen zwei über die FN-Hierarchie verbundenen Frames zu bestimmen, die beste Lösung. Mit einem Wert von  $\tau = 0,437$  korreliert die Sortierung dieses einfachen Maßes um 0,063 besser mit dem  $GS_{control}$  als das zweit-beste Maß, das Frame-Name-Distanz-Maß (siehe Abschnitt 6.2.2.3). Für die Sortierung von  $FP_{random}$  nimmt die Korrelation jedoch deutlich ab. Die Differenz der  $\tau$ -Werte beträgt 0,274 und liegt damit über dem Durchschnitt. Im arithmetischen Mittel beträgt die Differenz 0,1675.

Dieses Ergebnis ist ein klares Indiz für den uneinheitlichen Entwicklungsstand der FN-Hierarchie. Der  $FP_{control}$  beinhaltet nur Frame-Paare aus gut ausgearbeiteten Teilen der FN-Hierarchie (siehe Kapitel 4.1.1), während  $FP_{random}$  auch Paare aus weniger gut ausgearbeiteten Teilen der FN-Hierarchie enthält. Die beiden Datensätze unterscheiden sich sowohl in der Anzahl der verbundenen Frame-Paare, als auch in der durchschnittlichen Länge der Pfade, wie auch in der Dichte der Teile der FN-Hierarchie.

Im Datensatz  $FP_{control}$  sind 132 der 155 enthaltenen Frame-Paare miteinander verbunden, im  $FP_{random}$  hingegen nur 102 der Frame-Paare. Ein Pfad zwischen Paaren aus  $FP_{control}$  besitzt durchschnittlich eine Länge von 2,98. Ein Pfad zwischen Frame-Paaren

	$GS_{random}$	$GS_{control}$	$TOP30_{random}$	$TOP30_{control}$
Inheritance	62,45	40,8	75,79	35,55
Using	29,87	28,11	14,74	28,88
Subframe	0,21	10,2	0	4,44
Perspective_on	1,87	13,93	6,31	24,44
Causative_of	3,53	2,24	3,15	6,66
Inchoative_of	1,24	0,25	0	0
Precedes	0,82	4,48	0	0

Tabelle 6.2.: Prozentuale Verteilung der Frame-Relationen in FN

aus  $FP_{random}$  hat durchschnittlich eine Länge von 4,68. Die Pfade zwischen Frame-Paaren aus  $FP_{random}$  sind also im Durchschnitt ca. 2 Kanten länger als Pfade zwischen Paaren aus  $FP_{control}$ . Ein weiterer Faktor, der das Ergebnis beeinflusst, ist die Dichte der FN-Hierarchie. Da es gut und weniger gut ausgearbeitete Teile der FN-Hierarchie gibt, ist die Hierarchie an unterschiedlichen Stellen auch unterschiedlich dicht. Die durchschnittliche Dichte der FN-Hierarchie, errechnet nach Formel 6.1, ist  $D_{gesamt} = 0,0035$ .

$$D = \frac{|E|}{|V|^2} \quad (6.1)$$

wobei  $E$  die Menge aller Kanten bezeichnet und  $V$  die Menge aller Knoten.

Die durchschnittliche Dichte der Teilgraphen unter allen „scenario“-Frames ist dagegen  $D_{scenarios} = 0,0118$ . Die Paare des Datensatzes  $FP_{control}$  stammen aus diesen dichteren Teilen der FN-Hierarchie und weisen deshalb eine kürzere Pfadlänge auf.

Eine weitere Frage behandelt, welche Frame-Relationen in den Pfaden vorkommen. Tabelle 6.2 stellt den prozentualen Anteil der einzelnen Relationen an der Gesamtzahl der Relationen dar. Mehr als 92% aller Relationen auf Pfaden zwischen Frame-Paaren aus dem  $FP_{random}$  sind *Inheritance*- oder *Using*-Relationen. Alle übrigen Relationen teilen sich einen Anteil von ungefähr 8%. Bei Pfaden zwischen Frame-Paaren aus  $FP_{control}$  haben die *Inheritance*- und *Using*-Relation einen Anteil von rund 69%. Einen ebenfalls hohen Anteil mit jeweils mehr als 10% haben die *Perspective\_on*- und die *Subframe*-Relation. Der restliche Teil von rund 7% verteilt sich auf die *Precedes*-, *Causative\_of*- und *Inchoative\_of*-Relationen. Dies ist ein weiteres Beispiel dafür, dass die FN-Hierarchie unterschiedlich gut ausgearbeitet ist. Durch die *Perspective\_on*- und *Subframe*-Relation werden feinere Unterschiede modelliert, als durch *Inheritance*- oder *Using*-Relationen. Vergleicht man die Werte der TOP30-Evaluation in Tabelle 6.2, so stellt man fest, dass gerade die *Perspective\_on*-Relation mit 24,44% einen hohen Anteil der Relationen zwischen den Frames der Paare aus  $FP_{control}$  ausmacht. Aus den Daten der TOP30-Evaluation kann ebenfalls abgelesen werden, dass die Relationen *Inchoative\_of* und *Precedes* keinen Einfluß auf diese Pfade haben. Ihr Anteil liegt für beide Datensätze bei jeweils 0%.

Ein weiterer Punkt ist der überdurchschnittlich hohe Anteil der *Perspective\_on*- und

*Causative\_of*-Relationen in den TOP30 Frame-Paaren. Sowohl in den TOP30<sub>random</sub>, als auch in den TOP30<sub>control</sub> verschieben sich die Anteile von den *Inheritance*-, *Using*- und *Subframe*-Relationen, hin zu den *Perspective\_on*- und *Causative\_of*-Relationen. Diese Relationen verbinden sehr eng verwandte Konzepte, so dass die Ähnlichkeit dieser Konzepte als sehr groß empfunden wird.

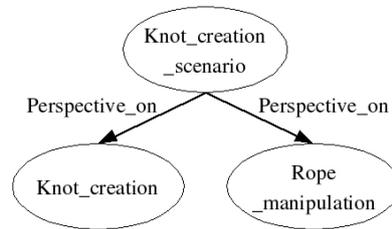


Abbildung 6.2.: Pfad Knot\_creation - Rope\_manipulation

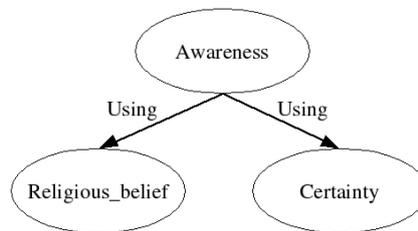


Abbildung 6.3.: Pfad Religious\_belief - Certainty

Die Pfade in den Abbildungen 6.2 und 6.3 sind dafür ein gutes Beispiel. Beide Pfade haben die Länge 2, und in beiden Pfaden kommt nur jeweils eine Art von Relation vor. Der Pfad zwischen KNOT\_CREATION und ROPE\_MANIPULATION besteht ausschließlich aus *Perspective\_on*-Kanten. Der Pfad zwischen RELIGIOUS\_BELIEF und CERTAINTY besteht ausschließlich aus *Using*-Kanten. Das Frame-Paar KNOT\_CREATION - ROPE\_MANIPULATION (3,8 von 4 Punkten) wird, als ähnlicher empfunden als das Frame-Paar RELIGIOUS\_BELIEF - CERTAINTY (1,5 von 4 Punkten)<sup>4</sup>. Mit dem „einfache Pfadlänge“-Maß ist dieser Unterschied jedoch nicht zu erfassen. Beide Pfade besitzen die Pfadlänge 2 und werden dementsprechend sortiert. Eine Möglichkeit, diese unterschiedlichen Eigenschaften der Relationen zu berücksichtigen ist, die Relationen zu gewichten.

Anschließend an diese Problematik stellt sich die Frage, wie viele unterschiedliche Relationen einen Pfad bilden. Die durchschnittliche Anzahl verschiedener Relationen pro Pfad ist in Tabelle 6.3 eingetragen. Auch hier ist ein Unterschied zwischen Pfaden zwischen den Frame-Paaren aus  $FP_{random}$  und Paaren aus  $FP_{control}$  zu erkennen. Pfade von

<sup>4</sup>Im Rahmen des in Kapitel 4.1 auf Seite 26 vorgestellten Experiments sollten die Versuchspersonen die Ähnlichkeit mit Werten zwischen 0 und 4 bewerten.

Paaren aus  $FP_{random}$  enthalten durchschnittlich eine größere Zahl verschiedener Relationen als Pfade von Frame-Paaren aus  $FP_{control}$ . Vergleicht man den Durchschnitt auf den gesamten Gold-Standards mit den Werten für die 30 besten Paare, so stellt man fest, dass die Zahl unterschiedlicher Relationen abnimmt. Daraus kann geschlossen werden, dass Pfade aus wenigen unterschiedlichen Relationen semantische Ähnlichkeit besser repräsentieren.

Zusammenfassend kann also festgestellt werden, dass kurze Pfade mit wenigen, unterschiedlichen Relationen eine sehr enge semantische Verwandtschaft anzeigen. Als zweites kann festgestellt werden, dass der Anteil an *Perspective\_on*- und *Causative\_of*-Relationen für ähnliche Paare höher ist, als für weniger ähnliche Paare. Es ist also für ein robustes Ähnlichkeitsmaß wichtig, die unterschiedlichen Arten der Relationen zu berücksichtigen. Die in den folgenden Abschnitten evaluierten Maße gewichten den Pfad auf unterschiedliche Art. Eine genaue Untersuchung wird zeigen, wo die Stärken und Schwächen der einzelnen Ansätze liegen.

**Pfadmuster** In diesem Abschnitt wird die Adaption des Ähnlichkeitsmaßes von Hirst und St-Onge (1998) evaluiert. Der Algorithmus nutzt Pfadmuster als Filter, um weniger aussagekräftige Pfade von der Berechnung auszuschließen.

Die Benutzung von Pfadmustern führt zu einer Annäherung der Korrelationswerte mit  $GS_{random}$  und  $GS_{control}$ .  $\tau_{random}$  beträgt 0,237 und  $\tau_{control}$  beträgt 0,266 (siehe Tabelle 6.1). Wie schon bei dem „einfache Pfadlänge“-Ansatz stellen wir fest, dass der Korrelationswert mit  $GS_{control}$  leicht besser ist, als die Korrelation  $GS_{random}$ . Allerdings liegt der Korrelationswert für die Frame-Paare aus  $FP_{control}$  unter dem Wert des „einfache Pfadlänge“-Maßes. Für Paare aus  $FP_{random}$  ist der Wert jedoch leicht besser.

Da der Vergleich mit Pfadmustern die einzige Änderung des Algorithmus gegenüber dem „einfache Pfadlänge“-Algorithmus darstellt, werden wir in diesem Abschnitt die Pfadmuster genauer betrachten. Wir werden vergleichen, welche Muster tatsächlich vorkommen und welche Pfadmuster Hirst und St-Onge (1998) benutzt haben. Ausserdem werden wir uns die Häufigkeiten der Pfadmuster anschauen. Dies erlaubt uns abschließend, unter Rückgriff auf Ergebnisse aus der Evaluation des „einfache Pfadlänge“-Ansatzes, eine Erklärung für die gesunkenen Korrelationswerte anzubieten.

Abbildung 6.6 zeigt die vorkommenden Pfadmuster zwischen den Frame-Paaren des  $FP_{control}$ -Datensatzes. Nach oben oder nach unten gerichtete Pfeile beschreiben aufwärts, beziehungsweise abwärts weisende vertikale Relationen; waagrecht verlaufende Pfeile beschreiben horizontale Relationen. Die Einteilung der FrameNet-Relationen in

	$FP_{random}$	$FP_{control}$	$TOP30_{random}$	$TOP30_{control}$
$\emptyset$	2	1,78	1,53	1,06
max.	3	3	3	2

Tabelle 6.3.: Durchschnittliche Anzahl unterschiedlicher Relationen pro Pfad

vertikale Relationen (*Inheritance*, *Using* und *Subframe*) und horizontale Relationen (*Perspective\_on*, *Precedes*, *Causative\_of* und *Inchoative\_of*) ist die gleiche, die auch bei der Adaption des Ähnlichkeitsmaßes von Hirst und St-Onge (1998) verwendet wurde (siehe Kapitel 5.1.1 auf Seite 33). Die 13 in Abbildung 6.6 dargestellten Muster beschreiben die 132 Pfade zwischen den Frame-Paaren aus  $FP_{control}$  vollständig. Ohne eine Gruppierung der Relationen erhält man 84 verschiedene Pfadmuster. Die hohe Anzahl der Pfadmuster hängt damit zusammen, dass es nur selten Pfade mit nur einer Relation, wie zum Beispiel der *Inheritance*-Relation, gibt. Meist bilden verschiedene Relationen einen Pfad.

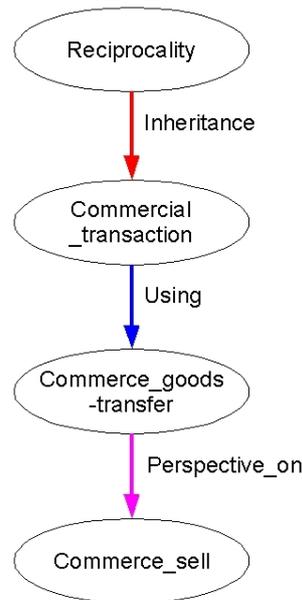


Abbildung 6.4.: Beispiel Pfad Commerce\_sell - Reciprocity

Ein typisches Beispiel aus  $FP_{control}$  ist der Pfad zwischen den Frames COMMERCE\_SELL und RECIPROCALITY (siehe Abbildung 6.4). Der Pfad hat die Länge drei und es werden drei unterschiedliche Relationen, *Perspective\_on*, *Using* und *Inheritance* benutzt. Der Pfad passt auf Pfadmuster i) (siehe Abbildung 6.6)<sup>5</sup>.

Abbildung 6.7 zeigt die Pfadmuster der Verbindungen des Datensatzes  $FP_{random}$ . 7 Muster beschreiben 102 Pfade zwischen den  $FP_{random}$ -Frame-Paaren. 4 Pfadmuster in Abbildung 6.6 und Abbildung 6.7 stimmen überein. Es sind die Pfadmuster a), c), e) und g). Tabelle 6.4 zeigt die Zahl der Vorkommen der einzelnen Pfadmuster. Pfadmuster, die sowohl für  $FP_{random}$  als auch für  $FP_{control}$  vorkommen, sind hervorgehoben. Die Spaltensummen überschreiten die maximale Anzahl von Pfaden pro Gold-Standard, da es für einige Frame-Paare alternative Pfade gleicher Länge gibt.

<sup>5</sup>Die Pfeile beschreiben jeweils ein oder mehrere Kanten (siehe Kapitel 5.1.1 auf Seite 33)

	$FP_{control}$		$FP_{random}$	
a)	<b>3</b>	(2,2%)	<b>8</b>	(7,6%)
b)	9	(6,5%)	2	(2%)
c)	<b>19</b>	(13,8%)	<b>2</b>	(2%)
d)	14	(10,1%)	9	(8,6%)
e)	<b>51</b>	(36,9%)	<b>74</b>	(70,5%)
f)	2	(1,4%)	1	(1%)
g)	<b>5</b>	(3,6%)	<b>9</b>	(8,6%)
h)	9	(6,5%)		
i)	1	(0,7%)		
j)	12	(8,7%)		
k)	7	(5,1%)		
l)	2	(1,4%)		
m)	4	(2,9%)		

Tabelle 6.4.: Anzahl der Vorkommen der Pfadmuster in den Gold-Standards

Das häufigste Pfadmuster ist Muster e), vertikal aufwärts gerichtete Relationen gefolgt von vertikal abwärts gerichteten Relationen.

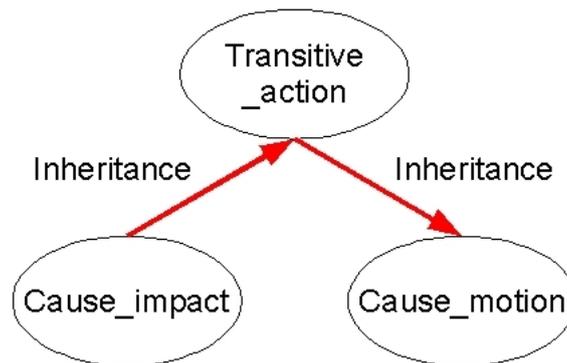


Abbildung 6.5.: Beispiel Pfadmuster e)

Abbildung 6.5 zeigt ein Beispiel für Muster e). CAUSE\_IMPACT ist über eine aufwärts gerichtete *Inheritance*-Kante mit TRANSITIVE\_ACTION, und TRANSITIVE\_ACTION mit einer abwärts gerichteten *Inheritance*-Kante mit CAUSE\_MOTION verbunden. Dies bestätigt die Beobachtungen aus dem ersten Abschnitt, dass kurze Pfade häufig vorkommen und gut geeignet sind, um semantische Ähnlichkeit beschreiben.

Weitherhin ist festzustellen, dass die Pfadmuster für  $FP_{control}$  einfacher sind als die Muster für  $FP_{random}$ . Dies ist wiederum mit der Dichte der FN-Hierarchie zu erklären. Wie bereits dargelegt, sind Pfade zwischen Frame-Paaren aus  $FP_{control}$  im Durchschnitt kürzer als Pfade zwischen Paaren aus  $FP_{random}$ , und die Pfade bestehen aus weniger unterschiedlichen Relationen. Dies führt zu einfacheren Pfadmustern.

Die Betrachtungen des vorigen Abschnittes zeigen, dass die unterschiedliche Dichte der FN-Hierarchie sich auch auf die Pfadmuster der Gold-Standards auswirkt. Die unterschiedliche Dichte der FN-Hierarchie erklärt auch, warum die Adaption des Hirst und St-Onge (1998) Ähnlichkeitsmaßes ebenfalls einen Unterschied in den Korrelationswerten zwischen Paaren aus  $FP_{random}$  und  $FP_{control}$  aufweist. In Tabelle 6.4 erkennt man, dass die Häufigkeit der Pfadmuster c) und e), die auch durch Hirst und St-Onge (1998) lizenziert sind, variiert. Pfadmuster c), das horizontale Relationen erlaubt, wird häufiger durch Pfade aus  $FP_{control}$  genutzt als durch Pfade aus  $FP_{random}$ .

Warum jedoch sind die Korrelationswerte für  $GS_{random}$  besser und für  $GS_{control}$  schlechter als bei dem „einfache Pfadlänge“-Ähnlichkeitsmaß? Um diese Frage zu beantworten, vergleichen wir die Pfadmuster von Hirst und St-Onge (1998) mit den tatsächlich vorkommenden Pfadmustern genauer. Abbildung 6.8 zeigt erneut die Pfadmuster von Hirst und St-Onge (1998), die für WordNet entwickelt wurden. Vergleichen wir diese Pfadmuster, die auch in dem adaptierten System genutzt wurden, mit den Pfadmustern der Pfade aus den beiden Datensätzen, so stellen wir fest, dass nur zwei der Muster von Hirst und St-Onge (1998) bei beiden Datensätzen vorkommen, die Muster b) und d). Die meisten Übereinstimmungen gibt es bei den Pfadmustern aus  $FP_{control}$ . Die Muster b), c), d), e), h), j) und k) kommen auch in den übernommenen Pfadmustern vor, jedoch nicht das Muster f) von Hirst und St-Onge (1998). Zwischen den Pfadmustern von  $FP_{random}$  und denen von Hirst und St-Onge (1998) gibt es jedoch nur zwei Übereinstimmungen. Die Muster c) und e) des  $FP_{random}$  stimmen mit den Pfadmustern b) und d) von Hirst und St-Onge (1998) überein. Alle übrigen Pfadmuster von Hirst und St-Onge (1998) kommen nicht als Pfadmuster in den  $FP_{random}$ -Daten vor.

Ein weiterer wichtiger Unterschied zwischen den Pfadmustern aus  $FP_{control}$  und den Pfadmustern des adaptierten Ähnlichkeitsmaßes von Hirst und St-Onge (1998) besteht darin, dass das Pfadmuster j) aus den Pfadmustern des  $FP_{control}$  in der Menge der Pfadmuster des adaptierten Ähnlichkeitsmaßes von Hirst und St-Onge (1998) fehlt. Dieses Muster ist deshalb so bedeutend, da es Pfade aus nur horizontalen Relationen beschreibt (siehe Abbildung 6.2 auf Seite 50). Diese horizontalen Relationen bündeln Relationen wie die *Perspective\_on-* Relation, die für die Bestimmung der semantischen Ähnlichkeit besonders wichtig sind (siehe Abschnitt 6.2.1.1). Eine bessere Gruppierung der einzelnen Relationen erscheint hier sinnvoll.

Zusammenfassend kann festgehalten werden, dass nur rund 54% der Pfadmuster von  $FP_{control}$  und nur rund 29% der Pfadmuster von  $FP_{random}$  durch die Pfadmuster von Hirst und St-Onge (1998) abgedeckt werden. Dies führt dazu, dass 25,5% der Frame-Paare des Datensatzes  $FP_{random}$  und 22,7% der Paare von  $FP_{control}$  ausgefiltert werden. Unter anderem werden auch gerade diejenigen Pfade von Paaren aus dem  $GS_{control}$  ausgesondert, die nur durch horizontale Relationen gebildet werden. Dies ist eine mögliche Erklärung für den schlechten Korrelationswert des adaptierten Ähnlichkeitsmaßes von Hirst und St-Onge (1998) mit dem  $GS_{control}$ . Bei den Pfaden aus Frame-Paaren des  $GS_{random}$  führt das Filtern durch die Pfadmuster von Hirst und St-Onge (1998) dazu, dass Pfade ausgesondert werden, die den Korrelationswert negativ beeinflussen. Daraus folgt, dass die Pfadmuster von Hirst und St-Onge (1998) nur teilweise übernommen werden sollten und die Einteilung der Relationen in nur zwei Kategorien nicht optimal ist.

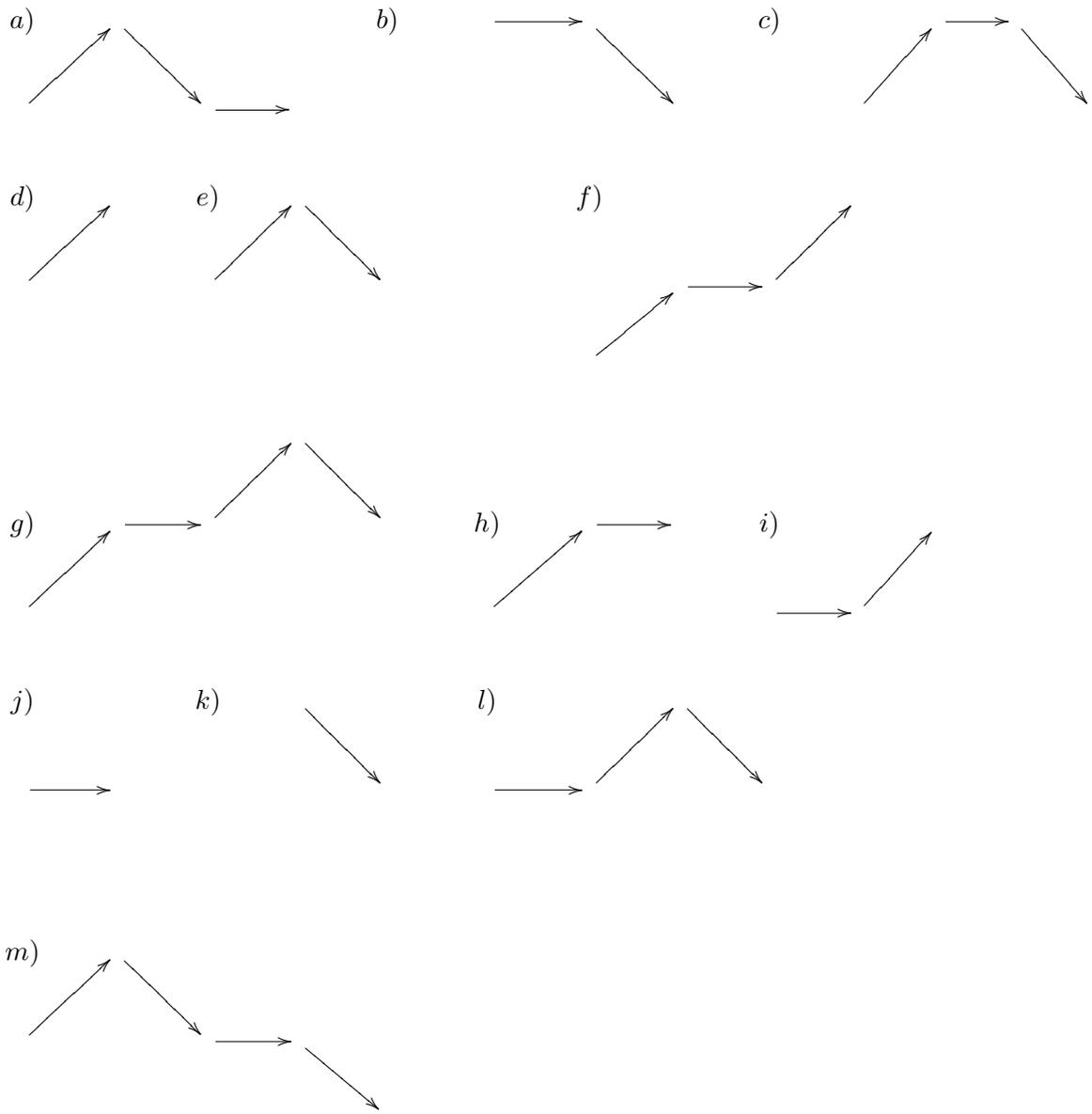


Abbildung 6.6.: Pfadmuster  $GS_{control}$

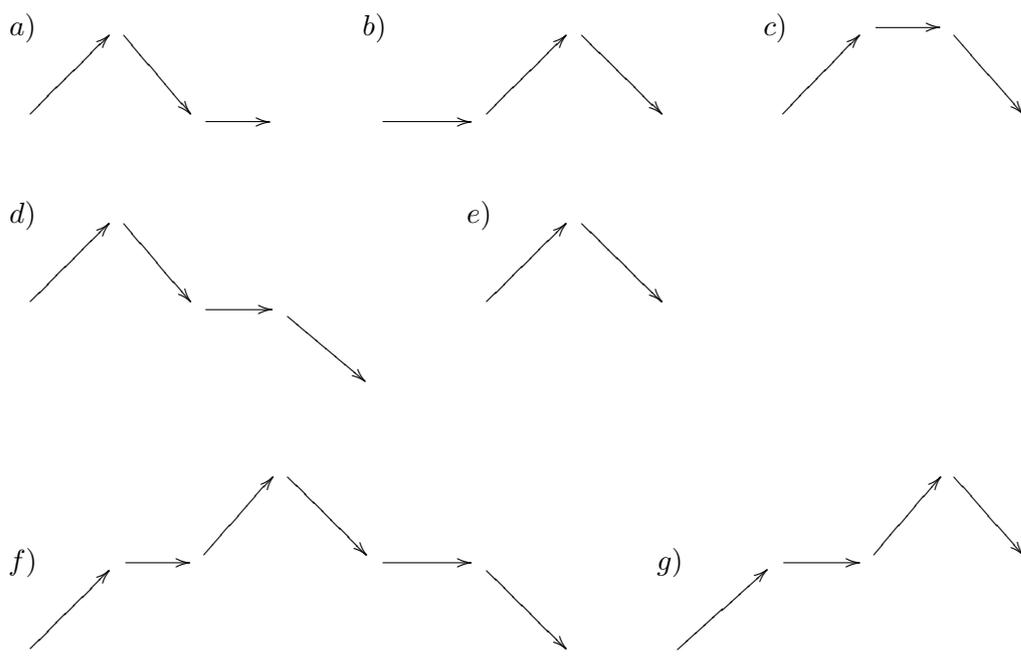


Abbildung 6.7.: Pfadmuster  $GS_{random}$

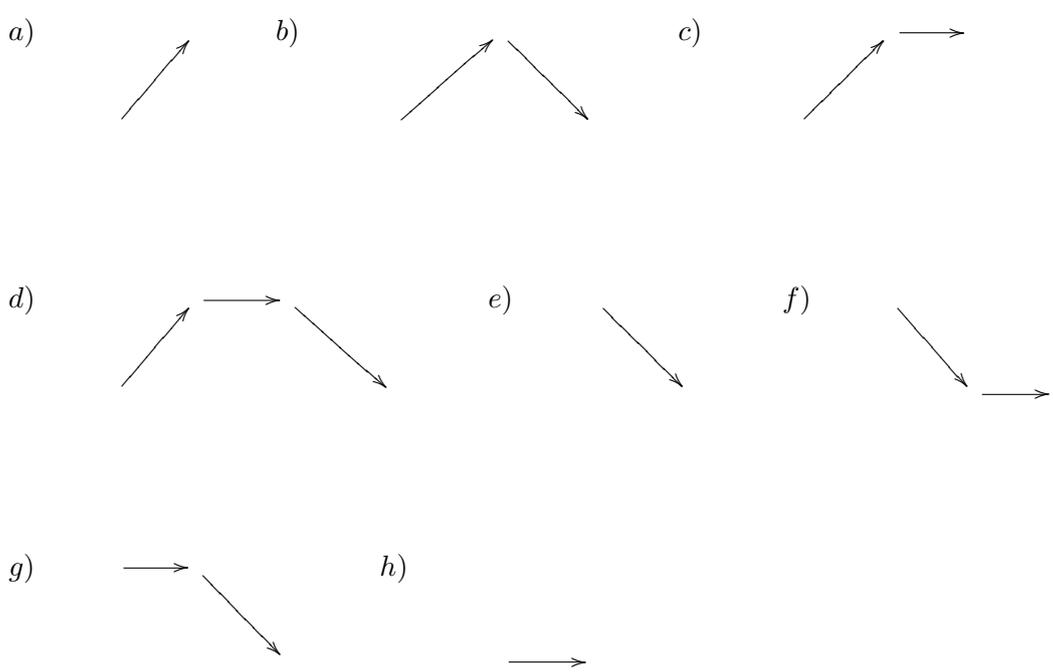


Abbildung 6.8.: Gültige Pfadmuster (Hirst und St-Onge, 1998)

### 6.2.1.2. Skalierte Pfadlänge

Im vorigen Abschnitt wurde die einfache Pfadlänge evaluiert. Es hat sich gezeigt, dass die ungewichtete Pfadlänge in lichten Teilen der FN-Hierarchie deutlich schlechtere Ergebnisse liefert, als in dichten Teilen. Das Skalieren der Pfadlänge ist eine bewährte Methode, diesem Effekt entgegen zu wirken. Zuerst werden die adaptierten WordNet-Maße von Leacock und Chodorow (1998) und Wu und Palmer (1994) besprochen. Anschließend werden die Ergebnisse des FE-Overlap-Maßes diskutiert.

**Tiefe der FN-Hierarchie** Das adaptierte Ähnlichkeitsmaß von Leacock und Chodorow (1998) benutzt die Tiefe der FN-Hierarchie, um Pfade zu skalieren. Die Korrelationswerte betragen 0,199 ( $GS_{random}$ ) und 0,473 ( $GS_{control}$ ). Damit zeigt dieses Maß die gleiche Korrelation mit den Gold-Standards, wie das „einfache Pfadlänge“-Maß. Die Skalierung mit der maximalen Tiefe der FN-Hierarchie ist also ungeeignet, den *sparse data* -Effekt zu mindern.

**Gemeinsamer Superframe** Das Ähnlichkeitsmaß von Wu und Palmer (1994) benutzt zur Skalierung der Pfadlänge die Abstände der Frames zum ersten gemeinsamen Superframe und die Tiefe des Superframes in der Hierarchie (siehe Kapitel 3.1.3.3).

Die Korrelationswerte weisen das bekannte Muster auf. Für Frame-Paare aus  $FP_{control}$  ist der Wert mit 0,320 höher als für Paare aus  $FP_{random}$  mit 0,153. Die unterschiedliche Dichte der FN-Hierarchie erklärt auch hier die Datenlage. Weiterhin kann festgestellt werden, dass beide Korrelationswerte niedriger sind als die Korrelationswerte des „einfache Pfadlänge“-Ähnlichkeitsmaßes, der Abstand zwischen beiden Werten jedoch nicht so groß ist wie beim „einfache Pfadlänge“-Ähnlichkeitsmaß.

Ein wichtiger Aspekt bei der Berechnung der Ähnlichkeit -mit Hilfe des von Wu und Palmer vorgestellten Algorithmus- ist die durchschnittliche Tiefe im Graphen. Die Pfade zwischen den Paaren aus  $FP_{random}$  sind durchschnittlich mit einer Tiefe von 1,18 eingebettet, die Pfade zwischen den Paaren aus  $FP_{control}$  weisen eine durchschnittliche Tiefe von 1,19 auf. Diese Werte machen deutlich, dass eine Gewichtung des Pfades mit der Tiefe in der FN-Hierarchie nicht sinnvoll ist. Der Unterschied ist zu gering, als dass er große Auswirkungen auf das Ergebnisse hätte.

Ein weiterer wichtiger Punkt ist nicht zuletzt, dass die FN-Hierarchie mehrere Wurzeln besitzt (siehe Kapitel 2.1.2.3). Wenn eine Wurzel identisch mit dem ersten gemeinsamen Superframe zweier Frames ist, führt dies dazu, dass der Pfad zwischen den Frames den Ähnlichkeitswert 0 erhält (siehe Formel 6.2).

$$ConSim_{Frame}(F1, F2) = \frac{2 * N3}{N1 + N2 + 2 * N3} \quad (6.2)$$

$N1$  und  $N2$  beschreiben den Abstand der Frames  $F1$  und  $F2$  zum ersten gemeinsamen Superframe.  $N3$  beschreibt den Abstand zur nächsten Wurzel.

10 Pfade zwischen Frame-Paaren aus  $FP_{random}$  und 40 Pfade zwischen Paaren aus  $FP_{control}$  werden aus diesem Grund mit Ähnlichkeit 0 bewertet. Unter anderem betrifft dies Frame-Paare wie KNOT\_CREATION - ROPE\_MANIPULATION oder RUN\_RISK - ENDANGERING, die beide zu den 30 ähnlichsten Frame-Paaren des  $GS_{control}$  gehören. Dies führt insgesamt zu einer Häufung von gebundenen Rängen am Ende der Ordnungen und insgesamt zu einem schlechteren Korrelationswert.

Zusammenfassend kann man sagen, dass die Länge des Pfades zum ersten gemeinsamen Superframe und dessen Tiefe in der FN-Hierarchie nicht geeignet sind, Pfade in der FN-Hierarchie zu skalieren und Frame-Ähnlichkeit robust zu beschreiben.

Ein weitere interessante Beobachtung ist, dass in beiden Datensätzen der Superframe im Durchschnitt in der Mitte des Pfades liegt. Die durchschnittliche Länge des Pfades bis zum ersten gemeinsamen Superframe beträgt 1,49 für Pfade von Paaren aus  $FP_{control}$  und 2,34 für Pfade von Frame-Paaren aus  $FP_{random}$ . Dies ist jeweils genau die Hälfte der durchschnittlichen Pfadlänge.

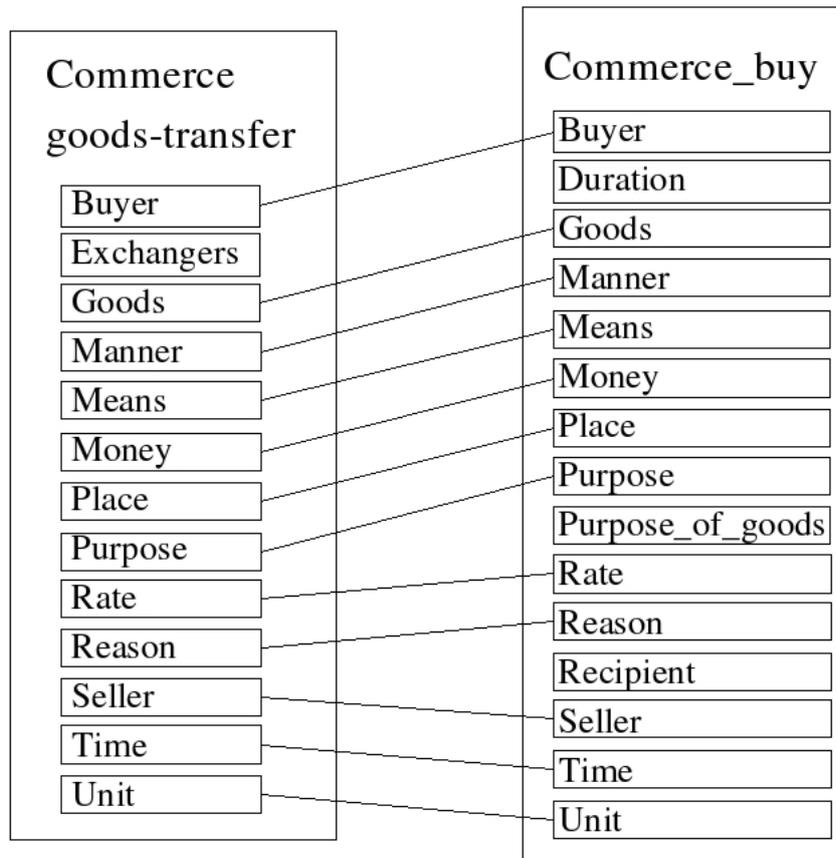
**FE-Overlap** Die FE-Overlap-Maße berechnen die Frame-Ähnlichkeit als den Quotienten der gemeinsamen FE mit der maximalen Anzahl von FE eines der Frames (siehe Kapitel 5.2.3). Die Korrelationswerte für diese beiden Ähnlichkeitsmaße sind 0,221, beziehungsweise 0,263, für die einfache FE-Überlappung, beziehungsweise die FE-Überlappung mit Namensgleichheit für den  $GS_{random}$ . Die Maße korrelieren mit Werten von 0,389, beziehungsweise 0,332 für den einfachen FE-Overlap, beziehungsweise den FE-Overlap mit Namensgleichheit mit dem  $GS_{control}$ . Der Unterschied zwischen den Korrelationswerten mit  $GS_{random}$  und  $GS_{control}$  ist geringer, als bei dem „einfache Pfadlänge“ - Ähnlichkeitsmaß, wobei insgesamt die Korrelation im Vergleich zum Bestwert am Wenigsten abnimmt.

Die Abnahme der Werte läßt sich mit Hilfe der Anzahl der verbundenen FE und deren Verteilung erklären. Tabelle 6.5 zeigt die Anzahl der verbundenen FE pro Frame-zu-Frame-Relation für die gesamte FN-Hierarchie. Durchschnittlich sind 69,92% aller

Inheritance	61,76
Using	41,37
Subframe	46,41
Perspective_on	69,92
Causative_of	51,02
Inchoative_of	45,25

Tabelle 6.5.: Anzahl verbundener FE pro Relation in Prozent für FN gesamt

FE von zwei Frames, die über die Frame-zu-Frame-Relation *Perspective\_on* verbunden sind, miteinander verknüpft. Ein gutes Beispiel dafür ist das Frame-Paar COMMERCE\_GOODS\_TRANSFER - COMMERCE\_BUY. 80% der FE sind verbunden (siehe Abbildung 6.9). Wie schon in Abschnitt (siehe Kapitel 6.2.1.1, auf Seite 49) dargelegt, bestehen die Pfade zwischen zwei Frames meist nicht nur aus Relationen eines Typs.



Commerce_goods-transfer	Commerce_buy
Buyer	Buyer
Goods	Goods
Manner	Manner
Means	Means
Money	Money
Place	Place
Purpose	Purpose
Rate	Rate
Reason	Reason
Seller	Seller
Time	Time
Unit	Unit

Abbildung 6.9.: Beispiel FE-Overlap

	$GS_{random}$		$GS_{control}$	
	FE-Overlap	FE-Overlap-Name	FE-Overlap	FE-Overlap-Name
Inheritance	69,47	72	44,68	50
Using	23,16	19	29,79	25
Subframe	0	0	0	0
Perspective_on	4,21	4	19,15	16,67
Causative_of	3,16	5	6,38	8,33
Inchoative_of	0	0	0	0
Precedes	0	0	0	0

Tabelle 6.6.: Prozentuale Verteilung der Relationen in den TOP30 Frame-Paaren

Tabelle 6.6 zeigt die durchschnittliche Verteilung der Relationen in den Pfaden zwischen den TOP30-Frame-Paaren für beide FE-Overlap-Ähnlichkeitsmaße und für beide Gold-Standards. Auffällig ist, dass nur vier verschiedene Relationen in den Pfaden vorkommen. Dies sind die *Inheritance*-, die *Using*-, die *Perspective\_on*- und die *Causative\_of*-Relation. Für Frame-Paare aus  $FP_{random}$  beträgt der Anteil der *Inheritance*-Relation für das einfache FE-Overlap-Ähnlichkeitsmaß 69,74% und für das FE-Overlap-Maß mit Namensgleichheit 72%. Die zweithäufigste vorkommende Relation ist die *Using*-Relation. Der Anteil der *Inheritance*- und *Using*-Relation liegt zusammen bei über 90% für beide FE-Overlap-Maße bei Frame-Paaren aus  $FP_{random}$ . Die für das Modellieren der Frame-Ähnlichkeit der FN-Hierarchie bedeutenden Relationen *Perspective\_on* und *Causative\_of* haben nur Anteile zwischen 3,16% und 5%.

Die Verteilung der Relationen ändern sich bei den TOP30 Frame-Paaren aus  $FP_{control}$  erwartungsgemäß. Die *Inheritance*- und *Using*-Relationen sind mit rund 75% immer noch die am häufigsten vorkommenden Relationen, aber der Anteil der *Perspective\_on*- und *Causative\_of*-Relationen ist wesentlich höher. Für das einfache FE-Overlap-Maß beträgt der Anteil der *Perspective\_on*-Relation in den Pfaden 19,15%. Damit ist der Anteil rund fünf mal größer als bei dem gleichen Maß, angewandt auf Paare aus  $FP_{random}$ . Der Anteil der *Causative\_of*-Relation ist bei Pfaden zwischen Frame-Paaren aus  $FP_{control}$  rund doppelt so hoch wie bei Pfaden zwischen Frame-Paaren aus  $FP_{random}$ . Wie schon bei den Pfaden zwischen Frame-Paaren aus  $FP_{random}$ , kommen die *Subframe*-, *Inchoative\_of*- und *Precedes*-Relationen nicht in den Pfaden vor.

Die Bedingung, dass nur Pfade zwischen zwei Frames akzeptiert werden, wenn mindestens eine Rolle der beiden Frames über den Pfad verbunden ist, hat nur minimale Auswirkung auf die Zahl der gefundenen Pfade. Die Zahl der gefundenen Pfade reduziert sich bei Anwendung des einfachen FE-Overlap-Maßes nur bei den Frame-Paaren aus  $FP_{control}$  um zwei. Die Anzahl der Pfade bei Anwendung des einfachen FE-Overlap-Maßes auf Paare aus  $FP_{random}$  bleibt gleich. Wird die Bedingung der Namensgleichheit des FE hinzugefügt, so reduziert sich die Anzahl der akzeptierten Pfade für beide Gold-Standards um 27.

Aus diesen Ergebnissen kann geschlossen werden, dass das Absinken der Korrelationswerte für die FE-Overlap-Maße, angewandt auf Frame-Paare aus  $FP_{control}$ , zwei Ursachen hat. Im Falle des einfachen FE-Overlap-Maßes führt die Verteilung der Relationen und die Verbundenheit der FE dazu, dass die Korrelation abnimmt. Der Anteil der eher schlecht verbundenen *Using*-Relation und *Causative\_of*-Relation ist bei Paaren aus  $FP_{control}$  größer als bei Frame-Paaren aus  $FP_{random}$ . Dies hat zwar auf die Zahl der gefundenen Pfade keinen Einfluß, jedoch auf die Bewertung der Ähnlichkeit. Für das FE-Overlap-Maß mit Namensgleichheit trifft dies ebenfalls zu. Zusätzlich wird die Korrelation durch die Reduktion der gefundenen Pfade negativ beeinflusst.

### 6.2.1.3. Hybrider Ansatz

Zur Berechnung der Ähnlichkeit wird die FN-Hierarchie bei dem adaptierten hybriden Ansatz von Maguitman u. a. (2005) in Adjazenzmatrizen umgeformt. Hierbei werden die Relationen in drei Kategorien aufgeteilt und jede Kategorie gewichtet (siehe Kapitel 5.1.3). Die „fuzzy membership matrix“  $W_{FN}$ , aus der die Ähnlichkeitswerte abgelesen werden, setzt sich nach folgender Formel zusammen:

$$W_{FN} = A_{RK\_1}^* \odot G_{FN} \odot A_{RK\_1}^* \quad (6.3)$$

Wichtig ist hier, dass die Matrix  $G_{FN}$  diejenige Matrix ist, bei der die *Perspective\_on*-, *Causative\_of*-, *Inchoative\_of*- und *Precedes*-Relation berücksichtigt werden. Die Matrix  $A_{RK\_1}^*$  wird nur mit Hilfe der *Inheritance*-, *Using*- und *Subframe*-Relation erstellt. Aus Formel 6.3 geht hervor, dass beliebig viele *Inheritance*-, *Using*- oder *Subframe*-Relationen in einem Pfad benutzt werden können, allerdings höchstens eine *Perspective\_on*-, *Causative\_of*-, *Inchoative\_of* oder *Precedes*-Relation.

Dies führt zu schlechten Korrelationswerten, vor allem bei der Sortierung des  $FP_{random}$ . Hier werden lediglich 13 von 155 Frame-Paaren mit Ähnlichkeitswerten versehen. Ein Korrelationswert des adaptierten hybriden, graph-basierten Ähnlichkeitsmaßes von 0,036 ist nicht signifikant. Auch für die Sortierung des  $FP_{control}$ -Datensatzes gilt, dass deutlich weniger Frame-Paare mit Ähnlichkeitswerten größer 0 versehen werden, als bei dem „einfache Pfadlänge“-Maß. Nur 102 von 155 Paaren werden mit einer Ähnlichkeit  $> 0$  bewertet. Die Korrelationswerte für die Ordnung mit dem  $GS_{control}$  sind mit 0,360 signifikant. Da aber die Korrelation zwischen der Sortierung des hybriden, graph-basierten Ähnlichkeitsmaßes und dem  $GS_{random}$  nicht signifikant ist, eignet sich dieses Ähnlichkeitsmaß nicht, um Frame-Ähnlichkeit robust zu beschreiben.

### 6.2.2. Nicht-struktur-basierte Maße

Das Definition-Overlap-, LU-Overlap- und das Frame-Name-Distanz-Ähnlichkeitsmaß werden in diesem Abschnitt untersucht. Die Evaluation wird zeigen, dass diese drei Maße nicht als robuste Frame-Ähnlichkeitsmaße geeignet sind, da mindestens eine Sortierung eines Datensatzes nicht signifikant mit dem entsprechenden Gold-Standard korreliert.

	$FP_{random}$	$FP_{control}$
Def-Overlap-einfach	25	53
Def-Overlap-erweitert	43	67

Tabelle 6.7.: Anzahl der Frame-Paare mit einer Ähnlichkeit  $> 0$

### 6.2.2.1. Definition-Overlap

Das Definition-Overlap-Maß misst Ähnlichkeit als Überschneidung im Definitionsvokabular. Je mehr Wörter in den Definitionen zweier Frames übereinstimmen, desto ähnlicher werden diese Frames bewertet. Im Folgenden präsentiere ich Experimente mit zwei verschiedenen Versionen dieses Maßes (siehe Kapitel 5.2.2). Die einfache Version (kurz: Def-Overlap-einfach) nutzt nur die in den Frame-Definitionen vorhandenen Wörter, um Überschneidungen zu bestimmen. Die zweite Version (kurz: Def-Overlap-erweitert) nutzt zusätzlich WordNet-Synonyme der vorhandenen Wörter, um den *sparse-data* Effekt zu minimieren. Aber die Sortierungen beider Versionen dieses Maßes korrelieren nicht signifikant mit dem  $GS_{random}$ . Die Korrelationswerte für die Sortierungen mit dem  $GS_{control}$  sind 0,310 für die einfache Version, beziehungsweise 0,305 für die erweiterte Version.

Das Def-Overlap-Maß ist nicht struktur-basiert. Tabelle 6.7 zeigt die Anzahl der Frame-Paare, die als ähnlich bewertet wurden. Es ist gut zu erkennen, dass die Erweiterung des einfachen Definition-Overlap-Maßes mit Synonymen zu mehr bewerteten Paaren führt. Gleichzeitig führt die Erweiterung um Synonyme dazu, dass die Ähnlichkeitswerte steigen.

<b>Def-Overlap-einfach</b>		
Frame_1	Frame_2	sim
Inspecting	Inspecting	1
Change_of_phase	Cause_change_of_phase	0,40
Shoot_projectiles	Use_firearm	0,36
Altered_phase	Change_of_phase	0,33
Employing	Hiring	0,32
<b>Def-Overlap-erweitert</b>		
Frame_1	Frame_2	sim
Inspecting	Inspecting	1
Shoot_projectiles	Use_firearm	0,46
Altered_phase	Change_of_phse	0,33
Getting	Giving	0,33
Employing	Hiring	0,32

Tabelle 6.8.: TOP 5 Frame-Paare Sortierung  $FP_{control}$

<b>Def-Overlap-einfach</b>		
Frame_1	Frame_2	sim
Strictness	Taking_time	0,08
Cause_to_be_wet	Feigning	0,08
Exertive_force	Labeling	0,06
Scrutiny	Sending	0,06
Cause_to_move_in_place	Change_of_phase	0,06
<b>Def-Overlap-erweitert</b>		
Frame_1	Frame_2	sim
Abundance	Measure_linear_extent	0,14
Strictness	Taking_time	0,13
Medical_instruments	Posing_as	0,13
Clothing	Infrastructure	0,12
Rashness	Reasoning	0,11

Tabelle 6.9.: TOP 5 Frame-Paare Sortierung  $FP_{random}$

Beispiele dafür sind die Ähnlichkeitswerte des Frame-Paares STRICTNESS - TAKING\_TIME und SHOOT\_PROJECTILES - USE\_FIREARM. Für das Paar SHOOT\_PROJECTILES - USE\_FIREARM steigt der Ähnlichkeitswert von 0,36 auf 0,46 (siehe Tabelle 6.8). Eine nicht ganz so starke Erhöhung des Wertes ist bei dem Frame-Paar STRICTNESS - TAKING\_TIME zu beobachten. Hier steigt der Ähnlichkeitswert von 0,08 auf 0,13 (siehe Tabelle 6.9).

Die Tabellen 6.8 und 6.9 zeigen die jeweils fünf bestbewerteten Frame-Paare aus den Sortierungen für beide Varianten des Maßes und für beide Datensätze. Vergleichen wir beide Tabellen, so fallen zwei Dinge auf: Erstens, die Ähnlichkeitswerte für die Frame-Paare aus dem Datensatz  $FP_{control}$  sind höher als die Werte für die Paare aus  $FP_{random}$ . Zweitens, die Daten aus den beiden Datensätzen reagieren unterschiedlich auf die Erweiterung des Definition-Overlap-Ähnlichkeitsmaßes.

Das Verhalten der Daten unterscheidet sich auf zwei Arten: Zum einen steigen die Ähnlichkeitswerte durchschnittlich um 0,058 für Frame-Paare aus  $FP_{random}$ , während für Paare aus  $FP_{control}$  sich die Werte nur um durchschnittlich 0,006 erhöhen. Andererseits ändern sich die TOP 5 Frame-Paare aus  $FP_{random}$  bis auf das Paar STRICTNESS - TAKING\_TIME bei zusätzlicher Benutzung von Synonymen, verglichen mit dem einfachen Definition-Overlap-Maß. In den TOP 5 der Frame-Paare aus  $FP_{control}$  ändert sich dagegen nur ein Frame-Paar.

Zusammenfassend kann man sagen, dass die Definitionen der Frames als Datengrundlage für ein robustes Frame-Ähnlichkeitsmaß nicht geeignet sind. Die Definitionen sind einerseits mit 74 Wörtern Länge relativ kurz, andererseits wird dieses *sparse data* - Problem noch durch ein zu uneinheitliches Definitionsvokabular verschärft. Auch eine Er-

	$FP_{random}$	$FP_{control}$
LU-Overlap	0,093	0,285
LU-Semantic-Space	<b>0,021</b>	<b>0,043</b>

Tabelle 6.10.: Korrelationswerte LU-basierte Ähnlichkeitsmaße

weiterung durch WordNet-Synonyme kann diese Lücken nicht füllen.

### 6.2.2.2. Lexical Units

In meinen Experimenten benutze ich LUs auf zwei verschiedene Arten, um Frame-Ähnlichkeit zu bestimmen. In der ersten Variante werden die LUs direkt in ein einfaches LU-Overlap-Ähnlichkeitsmaß eingebunden, in der zweiten Variante werden die LUs genutzt, um einen Vektorraum zu konstruieren, der die einzelnen Frames als Vektoren enthält. Zwischen diesen Vektoren wird die Frame-Ähnlichkeit mit Hilfe der Kosinus-Ähnlichkeit bestimmt. Tabelle 6.10 zeigt die Korrelationswerte aus dem Evaluationsexperiment<sup>6</sup>. Auch bei den LU-basierten Ähnlichkeitsmaßen sind die Korrelationswerte für die Sortierungen des  $FP_{control}$ -Datensatzes besser als die Werte für die Sortierung von  $FP_{random}$ . Desweiteren sind die Werte für das einfache LU-Overlap-Maß (kurz: LU-Overlap) höher als für das vektorraumbasierte Ähnlichkeitsmaß (kurz: LU-Semantic-Space). Die Korrelationswerte für das vektorraumbasierte Ähnlichkeitsmaß sind nicht signifikant.

Die Abdeckung der Ähnlichkeitsmaße ist ebenfalls sehr unterschiedlich. Das einfache LU-Overlap-Maß bewertet nur zwei der 155 Frame-Paare aus  $FP_{random}$ . Die zwei Frame-Paare sind RELIGIOUS\_BELIEF - CERTAINTY und ACCOUTREMENTS - CONNECTORS. Die Frames RELIGIOUS\_BELIEF und CERTAINTY besitzen beide die LU *believe.v.* Aus  $FP_{control}$  werden 18 Frame-Paare durch das LU-Overlap-Ähnlichkeitsmaß abgedeckt. Die Abdeckung durch das vektorraumbasierte Ähnlichkeitsmaß ist deutlich größer. Hier werden aus dem Datensatz  $FP_{random}$  155 Frame-Paare und aus dem Datensatz  $FP_{control}$  147 Frame-Paare bewertet.

Abschließend ist festzustellen, dass der vektorraumbasierte Ansatz nicht geeignet ist, um Frame-Ähnlichkeit zu beschreiben. Die Korrelationswerte für das einfache LU-Overlap-Maß sind zwar signifikant, aber niedrig. Wegen der geringen Abdeckung ist es jedoch als robustes Frame-Ähnlichkeitsmaß ungeeignet.

### 6.2.2.3. Frame-Name-Distanz

Das „Frame-Name-Distanz“-Maß nutzt zur Bestimmung der Frame-Ähnlichkeit den Frame-Namen. Die Hypothese besagt, je ähnlicher sich zwei Frame-Namen sind, desto ähnlicher

<sup>6</sup>Rot eingefärbte Korrelationswerte sind nicht signifikant.

$FP_{control}$		
Frame_1	Frame_2	sim
Inspecting	Inspecting	1
Commerce_collect	Commerce_sell	0,95
Cause_impact	Cause_motion	0,92
Measure_area	Measure_linear_extent	0,91
Motion_noise	Motion_directional	0,91

$FP_{random}$		
Frame_1	Frame_2	sim
Measure_volume	Measure_mass	0,92
Cause_impact	Cause_to_be_wet	0,87
Surpassing	Soaking	0,76
Subject_stimulus	Sent_items	0,75
Remembering_information	Reporting	0,71

Tabelle 6.11.: Frame-Name-Distanz-Maß - Jaro-Winkler Werte der 5 besten Frame-Paare

sind sich die Frames. Zur Bestimmung der Ähnlichkeit der Frame-Namen wird das Jaro-Winkler-Distanzmaß verwendet. Die Ergebnisse des Evaluationsexperiments zeigen, dass die Sortierung von  $FP_{random}$  mit  $GS_{random}$  nicht signifikant korreliert. Der Korrelationswert der Sortierung von  $FP_{control}$  mit  $GS_{control}$  ist dagegen sehr hoch. In der Sortierung des Datensatzes  $FP_{control}$  werden 133 Frame-Paare mit einer Ähnlichkeit größer 0 bewertet. Die Sortierung des Datensatzes  $FP_{random}$  enthält 122 Frame-Paare, deren Ähnlichkeit größer 0 ist.

Tabelle 6.11 zeigt die TOP5 Frame-Paare aus den Sortierungen der beiden Datensätze. Man erkennt, dass die Ähnlichkeitswerte für die Paare aus  $FP_{control}$  höher sind als für Frame-Paare aus  $FP_{random}$ . Dies ist darin begründet, dass die Frame-Paare aus  $FP_{control}$  die längeren Präfixe<sup>7</sup> teilen. Das Jaro-Winkler-Distanz-Maß bevorzugt Wortpaare mit gleichen Präfixen. Die Zahl der Frame-Namen mit gleichen Präfixen ist speziell unter „\_scenario“-Frames sehr hoch. Subframes des COMMERCE\_SCENARIO-Frames sind beispielsweise die Frames COMMERCIAL\_TRANSACTION, COMMERCE\_MONEY-TRANSFER oder COMMERCE\_GOODS-TRANSFER. Ebenfalls Subframes des COMMERCE

\_SCENARIO-Frames sind die Frames COMMERCE\_COLLECT und COMMERCE\_SELL aus Tabelle 6.11. Diese Frame-Namen deuten an, dass jeder dieser Frames eine „Handel“-Situation beschreibt. Dieser Zusammenhang ist für Frame-Namen, die nicht aus solch dichten Teilen von FN stammen, nicht gegeben.

<sup>7</sup>Ein Präfix bezeichnet hier eine identische wortinitiale Zeichenfolge.

Zusammenfassend ist auch das Frame-Name-Distanz-Maß als robustes Frame-Ähnlichkeitsmaß nicht geeignet. Quantitativ werden die meisten Frame-Paare mit einem Ähnlichkeitswert größer 0 bewertet, aber auf Grund ihres informellen Charakters sind Frame-Namen keine verlässliche Informationsquelle. Frame-Namen können Frame-Ähnlichkeit lediglich andeuten.

### 6.3. Zusammenfassung

Die Evaluation der Frame-Ähnlichkeitsmaße in diesem Kapitel hat gezeigt, dass nicht jede in FN zur Verfügung stehende Information geeignet ist, Frame-Ähnlichkeit robust und zuverlässig zu beschreiben. Die verschiedenen Ähnlichkeitsmaße wurden gegen zwei Gold-Standards evaluiert. So konnte das Verhalten der Ähnlichkeitsmaße in dichten und weniger dichten Teilen von FN untersucht werden.

Meine Untersuchung der Ähnlichkeitsmaße hat gezeigt, dass Ähnlichkeitsmaße, die die FN-Hierarchie nutzen, die besseren Ergebnisse liefern. Das beantwortet gleichzeitig die Frage nach der Bedeutung der Frame-Pfade. Pfade in der FN-Hierarchie zwischen zwei Frames sind zentraler Bestandteil eines robusten Frame-Ähnlichkeitsmaßes und besitzen eine große Aussagekraft. Die wichtigste Eigenschaft eines Pfades, in Bezug auf Frame-Ähnlichkeit, ist die Pfadlänge. Das „einfache Pfadlänge“-Ähnlichkeitsmaß hat sich für die Beschreibung von Frame-Ähnlichkeit als adäquat erwiesen.

Es hat sich allerdings auch gezeigt, dass hierarchie-basierte Ähnlichkeitsmaße stark vom *sparse data* -Problem betroffen sind. Die Untersuchung bereits für WordNet entwickelter Strategien zur Verminderung des *sparse data* -Problems ergab, dass weder die Skalierung der Pfadlänge mit der Tiefe der FN-Hierarchie, noch die Skalierung mit der Tiefe des ersten gemeinsamen Superkonzepts in der FN-Hierarchie, noch die Einführung von Pfadmustern, zu einem zufriedenstellenden Ergebnis führen. Bei der Betrachtung der Pfadmuster hat sich allerdings gezeigt, dass es wiederkehrende Muster gibt und es sinnvoll ist, gerade in lichten Teilen der FN-Hierarchie Pfadmuster zur Filterung von Pfaden zu benutzen. Die Evaluation hat deutlich gemacht, dass die Skalierung der Pfadlänge mit dem FE-Overlap die beste Möglichkeit zur Verminderung des *sparse data* -Problems darstellt.

Als ebenfalls unbefriedigend hat sich die Leistung des adaptierten hybriden, graph-basierten Ähnlichkeitsmaßes von Maguitman u. a. (2005) erwiesen. Zwar sind die Korrelationswerte für gut ausgearbeitete Teile der FN-Hierarchie gut, allerdings muss für weniger gut ausgearbeitete Teile der zu Grunde liegende Algorithmus angepasst werden.

Die nicht-struktur-basierten Maße sind, so hat sich weiterhin gezeigt, in viel stärkerem Maße vom Entwicklungsstand der jeweiligen Frames abhängig. Nur zwei von sechs Ähnlichkeitsmaßen erzeugen mit beiden Gold-Standards korrelierende Sortierungen der Daten. Da die Art und Weise der Gestaltung für Frame-Namen und Frame-Definitionen nicht garantiert werden kann, sind Maße, die implizit Annahmen über die innere Struktur von Frame-Namen und Frame-Definitionen machen, nicht zukunftssicher.

Abschließend kann gesagt werden, dass die Arbeitsdefinition 4.2.1 in Teilen revidiert werden muss. Eine vorläufige Definition sollte lauten:

**Zwei Frames  $F_1$  und  $F_2$  sind sich ähnlich, wenn sie über die FN-Hierarchie miteinander verbunden sind.**

Der Grad der Ähnlichkeit wird am besten durch die Zahl der verbundenen Frame-Elemente von  $F_1$  und  $F_2$  approximiert. Für eine solch einfache technische Definition spricht der hohe Korrelationswert des „einfachen Pfadlänge“-Ähnlichkeitsmaßes mit  $\tau = 0,473$  mit  $GS_{control}$ . Dieser Wert ist lediglich 0,057 kleiner, als der Korrelationswert der Versuchspersonen ( $\tau = 0,53$ ) mit dem Gold-Standard. Dies ist ein Beweis für die hohe Qualität von FrameNet. Eine zukünftige Revision der Definition sollte zu einem linguistisch informierteren Begriff von Frame-Ähnlichkeit führen.

## 7. Zusammenfassung und Ausblick

Ziel meiner Arbeit war es, ein robustes automatisches Frame-Ähnlichkeitsmaß zu entwickeln. Im ersten Teil der Diplomarbeit habe ich das Phänomen „Frame-Ähnlichkeit“ experimentell nachgewiesen. Gleichzeitig habe ich zwei Gold-Standard-Sortierungen von Frame-Paaren eingeführt. Diese Gold-Standards wurden im zweiten Teil der Arbeit zur Evaluation der zu untersuchenden Frame-Ähnlichkeitsmaße genutzt.

Im zweiten Teil der vorliegenden Studie wurden verschiedene Ähnlichkeitsmaße untersucht, um die Eigenschaften für ein robustes Frame-Ähnlichkeitsmaß aufzudecken. Hier hat sich gezeigt, dass Frame-Ähnlichkeit mit Hilfe der strukturellen Informationen von FrameNet modelliert werden kann.

Im folgenden Abschnitt will ich kurz die erreichten Ziele zusammenfassen und anschließend mögliche zukünftige Arbeit skizzieren

### 7.1. Erreichte Ziele

Im ersten Teil der Arbeit habe ich gezeigt, dass es einen intuitiven Ähnlichkeitsbegriff für Frames gibt. Mit Hilfe eines Experimentes (siehe Kapitel 4.1 auf Seite 26) wurde dieser Ähnlichkeitsbegriff untersucht. Die Interrater-Reliabilität der Versuchspersonen ist hoch signifikant<sup>1</sup>. Somit wurde der intuitive Frame-Ähnlichkeitsbegriff empirisch nachgewiesen. Des Weiteren wurden durch das Experiment zwei Gold-Standard-Sortierungen lizenziert (siehe Kapitel 4.1.2 auf Seite 27).

Im zweiten Teil der Diplomarbeit wurden mehrere Frame-Ähnlichkeitsmaße untersucht. Es wurden sowohl vorhandene WordNet-Wortähnlichkeitsmaße für FrameNet adaptiert, als auch neue Ähnlichkeitsmaße speziell für FrameNet entwickelt (siehe Kapitel 5 auf Seite 32). Die Ähnlichkeitsmaße gliedern sich dabei in struktur-basierte Ähnlichkeitsmaße, die das semantische Netz der FrameNet-Datenbank nutzen, und nicht-struktur-basierte Ähnlichkeitsmaße.

Anschließend wurden die Maße mit Hilfe der beiden Gold-Standards evaluiert (siehe Kapitel 6 auf Seite 44). Bei dieser Evaluation zeigte sich, dass das semantische Netz, indem die Frames organisiert sind, ein äußerst wichtiger Bestandteil eines robusten Frame-Ähnlichkeitsmaßes ist. Die getesteten struktur-basierten Ähnlichkeitsmaße korrelieren besser mit den beiden Gold-Standards als nicht-struktur-basierte Ähnlichkeitsmaße (siehe Kapitel 6.1.1 auf Seite 44). Hervorzuheben ist, dass die getesteten struktur-basierten Maße alle Frame-zu-Frame-Relationen nutzen, nicht nur die „is-a“-artigen Relationen (*Inheritance*, *Using*, *Subframe*). Weiterhin konnte ich zeigen, dass sich Pfadmuster dazu

---

<sup>1</sup> $W_{Kendall} = 0,789$  für  $FP_{random}$  bzw.  $W_{Kendall} = 0,697$  für  $FP_{control}$  (siehe Kapitel 4.1.2.2 auf Seite 28)

eignen, weniger aussagekräftige Pfade auszufiltern (siehe Kapitel 6.2.1.1 auf Seite 51).

Im Verlauf der Untersuchung zeigte sich, dass die struktur-basierten Ähnlichkeitsmaße auf Grund des unterschiedlichen Entwicklungsstandes von FrameNet vom *sparse data* - Problem betroffen sind. Als robustes Maß hat sich das Frame-Element-Overlap- Maß mit identischen Rollennamen erwiesen <sup>2</sup>. Dieses Maß nutzt die Übereinstimmung von Frame-Elementen von Frames, um Frame-Ähnlichkeit zu modellieren (siehe Kapitel 5.2.3 auf Seite 39). Das beste Ergebnis lieferte das „einfache Pfadlänge“- Maß in gut ausgearbeiteten Teilen von FrameNet. In Kapitel 6.3 wurde abschließend unter Berücksichtigung aller Erkenntnisse eine Definition von Frame-Ähnlichkeit gegeben.

Kernthema dieser Diplomarbeit war die Untersuchung und Modellierung von Frame-Ähnlichkeit unter Verwendung von FrameNet. Weitere, auch korpus-basierte Ähnlichkeitsmaße, werden im Artikel „Measuring frame relatedness“ (Pennacchiotti und Wirth, 2009) vorgestellt.

## 7.2. Zukünftige Arbeit

Dieser Abschnitt gibt einen Ausblick auf mögliche Anwendung des in dieser Arbeit entwickelten Ähnlichkeitsbegriffes.

Eine Reihe von Applikationen kann vom Einsatz eines robusten Frame-Ähnlichkeitsmaßes profitieren. So ist es mit dem hier entwickelten Frame-Ähnlichkeitsmaß möglich Systeme zu implementieren, die natürlichsprachliche Folgerungsbeziehungen („Textual Entailment“) modellieren. In dem Artikel „Building Text Meaning Representations from Conceptually Related Frames – A Case Study“ (Burchardt u. a., 2005) wurde solch ein System auf Basis von Frames bereits theoretisch betrachtet.

Im Rahmen der „PASCAL Recognising Textual Entailment Challenge“ (Dagan u. a., 2006) (kurz: *RTE*) wird aktiv an Textual Entailment- Systemen geforscht. Die Aufgabe bei RTE ist es, Folgebeziehungen (*Textual Entailment* (kurz: *Entailment* )) zwischen zwei Texten, genannt *Text* (Abk.: *T*) und *Hypothese* (Abk.: *H*), zu erkennen. Das Ziel ist es *T* und *H* mit Frames zu annotieren. In einem zweiten Schritt werden Frame-Paare aus *T* und *H* aligniert, wenn sie sich ähnlich sind. Die Hypothese lautet, je mehr Frames aligniert werden können, desto höher ist die Wahrscheinlichkeit, dass eine Folgebeziehung zwischen *T* und *H* besteht. Der triviale Fall ist, dass in *T* und *H* nur identische Frames vorkommen. Diese können dann aligniert werden. Interessanter sind Fälle wie das in Abbildung 7.1 gezeigte *T* - *H* -Paar Nr. 215 aus dem FATE<sup>3</sup> -Korpus.

---

<sup>2</sup> $\tau = 0,263$  mit  $GS_{random}$  und  $\tau = 0,332$  mit  $GS_{control}$  (siehe Kapitel 6.2.1.2 auf Seite 58)

<sup>3</sup>FATE („Frame Annotation for Textual Entailment“) (Burchardt und Pennacchiotti, 2008): Ein Frame-annotiertes RTE Korpus.

T	H	sim
ATTEMPT	ATTACK	0
KILLING	ATTACK	0,110
LEADERSHIP	ATTACK	0

Tabelle 7.1.: Frame-Ähnlichkeiten

(7.1) **T**: Only a few days ago, ARENA assassins tried to kill the president of the Mortgage Bank, Mr. Mason, for not following their orders.

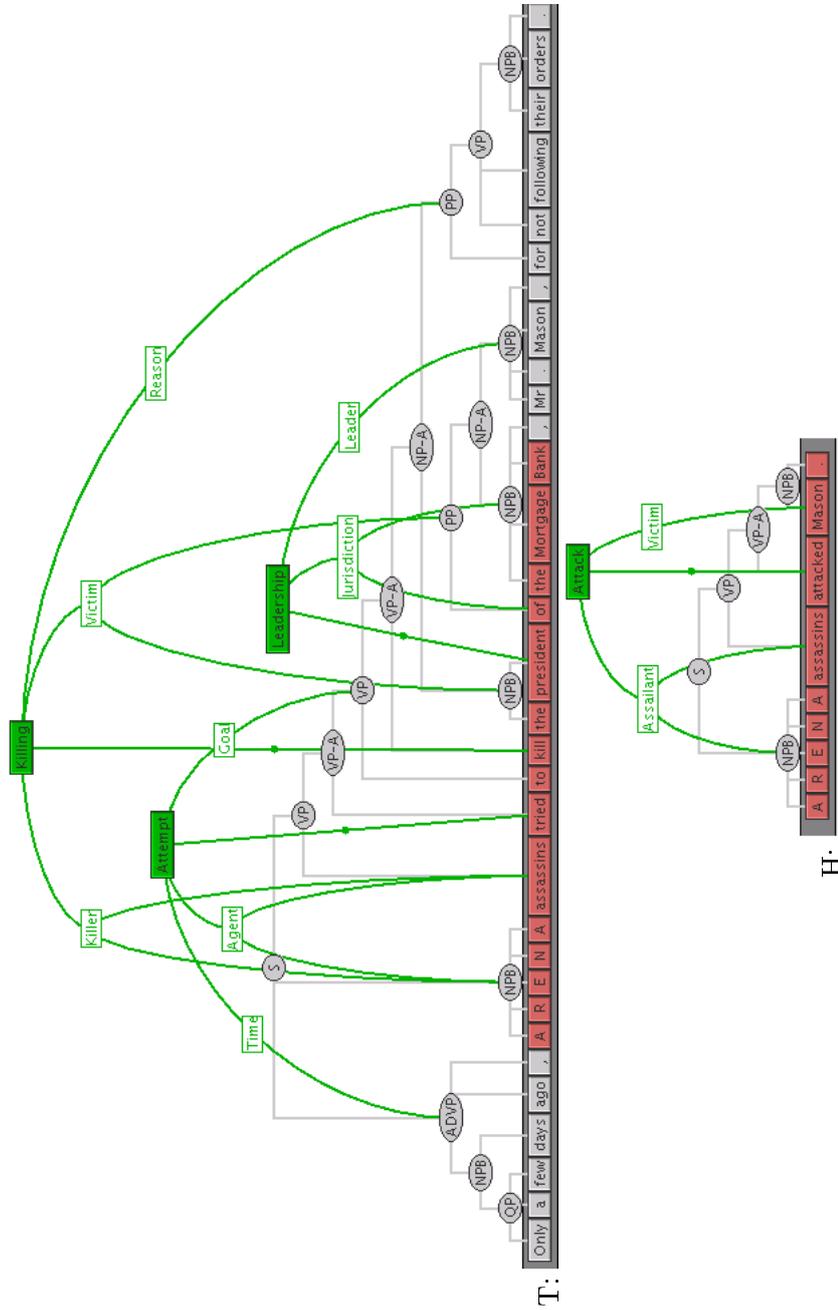
(deutsch: T: Vor wenigen Tagen versuchten ARENA Auftragsmörder den Präsidenten der Mortgage Bank zu töten, da dieser ihre Befehle nicht befolgt hatte.)

(7.2) **H**: ARENA assassins attacked Mason.

(deutsch: H: ARENA Auftragsmörder griffen Mason an.)

In diesem Beispiel ist Entailment zwischen **T** (7.1) und **H** (7.2) gegeben. Das vorgeschlagene RTE-System sollte den KILLING und den ATTACK - Frame alignieren. Hierfür wird die Ähnlichkeiten der Frame-Paare ATTEMPT - ATTACK, KILLING - ATTACK und LEADERSHIP - ATTACK berechnet. Die Frame-Ähnlichkeit für das Frame-Paar KILLING - ATTACK ist mit  $sim = 0,110$  am höchsten (siehe Tabelle 7.1). Das Paar wird aligniert. Da Identische Füller der FEs ein weiteres Indiz für eine bestehende Entailment-Beziehung zwischen **T** und **H** sind, werden in einem weiteren Schritt verbundene FE gesucht. Das FE *Killer* des KILLING -Frame ist mit dem *Assailant* -FE des ATTACK -Frame verbunden, und das FE *Victim* des KILLING -Frame ist mit dem FE *Victim* des ATTACK -Frames verbunden. Der Füller für das *Killer* -FE und das *Assailant* -FE ist „ARENA assassins“. Der Füller für das *Victim* -FE des KILLING -Frames ist „the president of the Mortgage Bank, Mr. Mason“, der Füller für das *Victim* -FE des ATTACK -Frames ist „Mason“. Die FE- Füller sind also identisch und eine Entailment -Beziehung zwischen **T** und **H** sehr wahrscheinlich.

Sowohl die vielfältigen Anwendungsmöglichkeiten eines robusten und linguistisch informierten Frame-Ähnlichkeitsmaßes, als auch die mit einfachen Methoden erzielten guten Ergebnisse, sollten zu weiterer Forschung auf diesem Gebiet ermutigen.



H:

Abbildung 7.1.: FATE Text-Hypothese-Paar (FATE T/H-Paar 215)

# A. Gold Standards

## A.1. $GS_{control}$

Die Tabelle zeigt die komplette Gold-Standard Sortierung  $GS_{control}$ .

Frame_1	Frame_2	Rang
Inspecting	Inspecting	1
Knot_creation	Rope_manipulation	2,5
Shoot_projectiles	Use_firearm	2,5
Scouring	Scrutiny	4
Ambient_temperature	Temperature	5
Fleeing	Escaping	6
Employing	Hiring	7
Change_of_phase	Cause_change_of_phase	8
Altered_phase	Change_of_phase	9
Cause_change_of_phase	Altered_phse	10
Cause_impact	Hit_target	11
Cause_impact	Cause_motion	12
Rape	Abusing	13
Getting	Giving	14
Being_necessary	Reason	15
Being_necessary	Needing	16
Scouring	Inspecting	17
Possession	Amassing	18
Hit_target	Use_firearm	19
Use_firearm	Cause_impact	20
Sentencing	Jury_deliberation	21
Rape	Cause_harm	22
Rape	Arson	23
Getting	Theft	24
Dimension	Quantity	25
Reliance	Desiring	26
Measure_area	Measure_linear_extent	27
Run_risk	Endangering	28
Safe_situation	Security	29
Reciprocity	Exchange	30
Commerce_sell	Reciprocity	31

Frame_1	Frame_2	Rang
Being_employed	Quitting	32
Being_employed	Employment_end	33
Finish_competition	Success_or_failure	34
Fields	Craft	35
Hit_target	Cause_motion	36
Protecting	Run_risk	37
Intentionally_affect	Rope_manipulation	38,5
Knot_creation	Intentionally_affect	38,5
Intentionally_affect	Use_firearm	40
Suspicion	Criminal_investigation	41
Dimension	Unemployment_rate	42
Age	Measure_duration	43
Contingency	Being_necessary	44
Have_as_requirement	Deserving	45
Body_movement	Motion	46
Terms_of_agreement	Being_obligatory	47
Appeal	Arraignment	48
Thwarting	Subversion	49
Commerce_collect	Commerce_sell	50
Success_or_failure	Thwarting	51
Run_risk	Safe_situation	52
Damaging	Preventing	53
Product_delivery	Expensiveness	54
Businesses	Commerce_pay	55
Motion_noise	Motion_directional	56
Measure_duration	People_by_age	57
Measure_by_action	Quantity	58,5
Measure_duration	Quantity	58,5
Measure_area	Temperature	60,5
Temperature	Measure_duration	60,5
Successful_action	Thwarting	62
Endangering	Safe_situation	63
Giving	Expensiveness	64
Amassing	Giving	65
Arrest	Suspicion	66
Suspicion	Scrutiny	67
Criminal_investigation	Seeking_to_achieve	68
Criminal_investigation	Inspecting	69
Unemployment_rate	Measure_area	70
Notification_of_charges	Verdict	71
Criminal_investigation	Scouring	72,5

Frame_1	Frame_2	Rang
Locating	Criminal_investigation	72,5
Questioning	Verdict	74
Surrendering_possession	Fining	75
Choosing	Forgoing	76
Rewards_and_punishments	Getting	77
Suspicion	Bail_setting	78
Success_or_failure	Expertise	79,5
Successful_action	Expertise	79,5
Expertise	Thwarting	81
Reliance	Reason	82
Finish_competition	Attempt	83
Surrendering	Rape	84
Redirecting	Travel	85
Appeal	Try_defendant	86
Locating	Success_or_failure	87
Criminal_investigation	Success_or_failure	88
Having_commercial_agreement	Product_delivery	89
Exchange	Sending	90
Employing	Intentionally_affect	92,5
Hit_target	Intentionally_affect	92,5
Intentionally_affect	Cause_impact	92,5
Intentionally_affect	Getting	92,5
Arrest	Arson	95
Smuggling	Notification_of_charges	96
Kidnapping	Court_examination	97
Entering_of_plea	Theft	99
Robbery	Entering_of_plea	99
Suspicion	Robbery	99
Trial	Piracy	101
Delivery	Removing	102,5
Emptying	Delivery	102,5
Submitting_documents	Taking	104
Expensiveness	Exporting	105
Expensiveness	Reciprocity	106
Jury_deliberation	Offenses	107
People	Custom	108
Fields	People	109
Reliance	Contingency	110
Subversion	Success_or_failure	111
Motion_directional	Location_of_light	112
Exchange_currency	Fining	113

Frame_1	Frame_2	Rang
Age	Speed	114
Fluidic_motion	Cause_begin_motion	115
Fining	Importing	116
Renting	Exporting	117,5
Renting_out	Importing	117,5
Bringing	Excreting	119
Bringing	Operate_vehicle	120
Path_shape	Getting_underway	121
Beyond_compare	Severity_of_offense	122
Suspicion	Attempt	123
Committing_crime	Operate_vehicle	124
Submitting_documents	Attention	125
Communication	Legality	126
Committing_crime	Taking	127
Rape	Possession	128
Amassing	Kidnapping	142
Being_in_operation	Medical_conditions	142
Bringing	Cause_harm	142
Clothing	Being_named	142
Criminal_investigation	Finish_competition	142
Endangering	Likelihood	142
Evoking	Setting_fire	142
Finish_competition	Suspicion	142
Giving	Rape	142
Measure_by_action	People	142
Motion	Aggregate	142
Motion_directional	Clothing	142
Needing	Taking_time	142
Operate_vehicle	Terms_of_agreement	142
Pardon	Intentional_traversing	142
People	Measure_duration	142
Quitting	Fining	142
Reason	Taking_time	142
Rejuvenation	Physical_artwork	142
Revenge	Bungling	142
Security	Likelihood	142
Sidereal_appearance	Aggregate	142
Simple_name	Proliferating_in_number	142
Source_of_getting	Setting_out	142
Subjective_influence	Infrastructure	142
Success_or_failure	People	142

Frame_1	Frame_2	Rang
Success_or_failure	Suspicion	142

## A.2. $GS_{random}$

Die Tabelle präsentiert die Gold-Standard Sortierung  $GS_{random}$ .

Frame_1	Frame_2	Rang
Measure_volume	Measure_mass	1
Communication_manner	Statement	2
Giving	Sent_items	3
Abundance	Measure_linear_extent	4
Remembering_information	Reporting	5
Religious_belief	Certainty	6
Cause_change_of_consistency	Create_representation	7
Gathering_up	Amalgamation	8
Abounding_with	Scouring	11
Being_detached	Filling	11
Judicial_body	Misdeed	11
Measure_mass	Likelihood	11
Prominence	Frugality	11
Delivery	Reparation	16
Fleeing	Sending	16
Giving	Communication	16
Operational_testing	Reason	16
Sending	Reparation	16
Suitability	Completeness	20
Willingness	Difficulty	20
Willingness	Regard	20
Cause_to_fragment	Connectors	26
Expectation	Estimated_value	26
Experiencer_subj	Objective_influence	26
Grasp	Terms_of_agreement	26
Rashness	Reasoning	26
Resolve_problem	Forging	26
Separation	Giving	26
Social_interaction_evaluation	Request	26
Strictness	Taking_time	26
Be_in_agreement_on_assessment	Surpassing	36,5
Cause_impact	Cause_to_be_wet	36,5
Comitting_crime	Create_physical_artwork	36,5
Court_examination	Reveal_secret	36,5

Frame_1	Frame_2	Rang
Encoding	Desiring	36,5
Excreting	Being_attached	36,5
Hiring	Occupy_rank	36,5
Intentional_traversing	Attack	36,5
Intentionally_affect	Birth	36,5
Intentionally_affect	Judgment_communication	36,5
Participation	Suspicion	36,5
Surrendering_possession	Importing	36,5
Becoming	Abusing	52,5
Becoming	Being_employed	52,5
Becoming	Inhibit_movement	52,5
Cause_to_fragment	Inhibit_movement	52,5
Comitting_crime	Safe_situation	52,5
Create_physical_artwork	Left_to_do	52,5
Damaging	Craft	52,5
Emotion_active	Being_obligated	52,5
Exchange	Misdeed	52,5
Existence	Being_detached	52,5
Filling	Difficulty	52,5
Getting	Opinion	52,5
Giving_in	Detaining	52,5
Judgment_communication	Preserving	52,5
Make_noise	Safe_situation	52,5
Manipulation	Suspiciousness	52,5
Measure_mass	Fairness_evaluation	52,5
Resurrection	Daring	52,5
Social_event	Disgraceful_situation	52,5
Wearing	Craft	52,5
Amalgamation	Gizmo	66,5
Emanating	Operating_a_system	66,5
Giving	Instance	66,5
Judicial_body	Shoot_projectiles	66,5
Quantity	Adjusting	66,5
Remainder	Attention_getting	66,5
Scrutiny	Sending	66,5
Subject_stimulus	Sent_items	66,5
Accoutrements	Connectors	85
Adorning	Perception_active	85
Becoming_a_member	Remembering_to_do	85
Behind_the_scenes	Imprisonment	85
Birth	Reparation	85

Frame_1	Frame_2	Rang
Body_description_holistic	Abundance	85
Buildings	Congregating	85
Cause_to_move_in_place	Change_of_phase	85
Change_of_leadership	Resolve_problem	85
Contribution	Predicting	85
Deserving	Buildings	85
Domain	Misdeed	85
Expected_location_of_person	Opinion	85
Influence_of_event_on_cognizer	Social_connection	85
Manufacturing	Ammunition	85
Operational_testing	Defend	85
Preserving	Proliferating_in_number	85
Public_services	Cause_to_make_noise	85
Remembering_to_do	Cause_to_be_dry	85
Secrecy_status	Estimated_value	85
Sent_items	Evoking	85
Shoot_projectiles	Social_event	85
Taking_time	Contingency	85
Abusing	Network	126
Adducing	Cause_to_be_dry	126
Ammunition	Cause_change_of_phase	126
Apply_heat	Forming_relationships	126
Attack	Quitting	126
Attention_getting	Fields	126
Attention_getting	Scouring	126
Attributed_information	Direction	126
Becoming_a_member	Temperature	126
Being_active	Measure_duration	126
Being_named	Expectation	126
Body_description_holistic	Translating	126
Body_mark	Supply	126
Bragging	Importing	126
Candidness	Forging	126
Cause_change_of_consistency	Grasp	126
Cause_temperature_change	Travel	126
Cause_to_be_wet	Feigning	126
Change_of_leadership	Perception_active	126
Clemency	Gizmo	126
Clothing	Indigenous_origin	126
Clothing	Infrastructure	126
Containers	Idiosyncrasy	126

Frame_1	Frame_2	Rang
Craft	Institutionalization	126
Deny_permission	Friction	126
Detaining	Apply_heat	126
Dodging	Building	126
Eclipse	Statement	126
Emotion_directed	People_by_residence	126
Escaping	Containers	126
Exertive_force	Labelling	126
Exporting	Become_silent	126
Expressing_publicly	Motion_directional	126
Food	Topic	126
Frequency	Medical_instruments	126
Ground_up	Closure	126
Hiding_objects	Forming_relationships	126
Hiring	Clothing	126
Immobilization	Sent_items	126
Jugdment	Cutting	126
Kinship	Exporting	126
Labelling	Buildings	126
Legality	Fastener	126
Measure_by_action	Social_connection	126
Medical_instruments	Posing_as	126
Medical_professionals	Respond_to_proposal	126
Ordinal_numbers	Activity_ready_state	126
Preventing	Being_named	126
Prominence	Delivery	126
Public_services	Being_in_operation	126
Redirecting	Measure_by_action	126
Religious_belief	Cutting	126
Research	Immobilization	126
Resurrection	Strictness	126
Social_event	Rope_manipulation	126
Social_event	Word_realtions	126
Sole_instance	Chatting	126
Sufficiency	Becoming_a_member	126
Surpassing	Soaking	126
Theft	Getting_up	126
Use_firearm	Namesake	126
Withdraw_from_participation	Cause_to_be_dry	126

# Literaturverzeichnis

- [Baker u. a. 1998] BAKER, Collin F. ; FILLMORE, Charles J. ; LOWE, John B.: The Berkeley FrameNet Project. In: *Proceedings of the 17th international conference on Computational linguistics-Volume 1* Association for Computational Linguistics Morristown, NJ, USA (Veranst.), 1998, S. 86–90
- [Bernard 1986] BERNARD, John R. L.: *The Macquarie Thesaurus*. Macquarie Library, 1986
- [Budanitsky 1999] BUDANITSKY, Alexander: Lexical Semantic Relatedness And Its Application In Natural Language Processing / Computer Systems Research Group, University of Toronto. August 1999. – Forschungsbericht
- [Budanitsky und Hirst 2001] BUDANITSKY, Alexander ; HIRST, Graeme: Semantic distance in WordNet: An experimental, application-oriented evaluation of five measures. In: *Workshop on WordNet and Other Lexical Resources* Bd. 2, 2001
- [Budanitsky und Hirst 2006] BUDANITSKY, Alexander ; HIRST, Graeme: Evaluating WordNet-based Measures of Lexical Semantic Relatedness. In: *Computational Linguistics* 32 (2006), Nr. 1, S. 13–47
- [Burchardt 2008] BURCHARDT, Aljoscha: *Saarbrücken dissertation series in Computational Linguistics*. Bd. 29: *Modelling Textual Entailment with Role-Semantic Information*. Saarbrücken : DFKI GmbH and Saarland University, 2008
- [Burchardt und Frank 2006] BURCHARDT, Aljoscha ; FRANK, Anette: Approaching Textual Entailment with LFG and FrameNet Frames. In: *2nd PASCAL RTE Workshop*, 2006, S. 92–97
- [Burchardt u. a. 2005] BURCHARDT, Aljoscha ; FRANK, Anette ; PINKAL, Manfred: Building text meaning representations from contextually related frames—a case study. In: *Proceedings of the Sixth International Workshop on Computational Semantics, IWCS-6*. Tilburg, Netherlands, 2005
- [Burchardt und Pennachioti 2008] BURCHARDT, Aljoscha ; PENNACHIOTTI, Marco: FATE: a FrameNet-Annotated corpus for Textual Entailment. In: *Proceedings of LREC 2008*, 2008. – to appear
- [Burchardt u. a. 2007] BURCHARDT, Aljoscha ; REITER, Nils ; THATER, Stefan ; FRANK, Anette: A Semantic Approach To Textual Entailment: System Evaluation and Task Analysis. In: *Proceedings of the Workshop on Textual Entailment and Paraphrasing*. Prague, June 2007, S. 10–15

- [Chang u. a. 2002] CHANG, Nancy ; NARAYANAN, Srinu ; PETRUCK, Miriam R.: From frames to inference. In: *Proceedings of the First International Workshop on Scalable Natural Language Understanding, Heidelberg, Germany, 2002*
- [Charles 2001] CHARLES, Walter G.: Contextual correlates of meaning. In: *Applied Psycholinguistics* 21 (2001), Nr. 04, S. 505–524
- [Cohen u. a. 2003] COHEN, William W. ; RAVIKUMAR, Pradeep ; FIENBERG, Stephen E.: A comparison of string distance metrics for name-matching tasks. In: *Proceedings of the IJCAI-2003 Workshop on Information Integration on the Web (IIWeb-03)*, 2003
- [Collins und Loftus 1975] COLLINS, Allan M. ; LOFTUS, Elizabeth F.: A spreading-activation theory of semantic processing. In: *Psychological Review* 82 (1975), Nr. 6, S. 407–428
- [Cover und Thomas 1991] COVER, Thoms M. ; THOMAS, Joy A.: *Elements of Information Theory*. John Wiley & Sons, 1991
- [Dagan u. a. 2006] DAGAN, Ido ; GLICKMAN, Oren ; MAGNINI, Bernardo: The PASCAL recognising textual entailment challenge. In: *Lecture Notes in Computer Science* 3944 (2006), S. 177–190
- [Erk u. a. 2003] ERK, Katrin ; KOWALSKI, Andrea ; PADÓ, Sebastian ; PINKAL, Manfred: Towards a resource for lexical semantics: a large german corpus with extensive semantic annotation. In: *Proceedings of ACL-03*. Morristown, NJ, USA, 2003
- [Fellbaum 1998] FELLBAUM, Christiane (Hrsg.): *WordNet. An electronic lexical database. Language, Speech and Communication*. Cambridge, MA : MIT Press, 1998
- [Fellbaum und Miller 1990] FELLBAUM, Christiane ; MILLER, George A.: Folk Psychology or Semantic Entailment? Comment on Rips and Conrad (1989). In: *Psychological Review* 97 (1990), Nr. 4, S. 565–70
- [Fillmore 1977] FILLMORE, Charles: *Fundamental Studies in Computer Sciences*. Bd. 5. Kap. Scenes-and-frames, S. 55–88, North Holland Publishing, 1977
- [Fillmore u. a. 2003] FILLMORE, C.J. ; PETRUCK, Miriam R. ; RUPPENHOFER, Joseph ; WRIGHT, Abby: Framenet in action: The case of attaching. In: *International Journal of Lexicography* 16 (2003), Nr. 3, S. 297–332
- [Finkelstein u. a. 2002] FINKELSTEIN, Lev ; GABRILOVICH, Evgeniy ; MATIAS, Yossi ; RIVLIN, Ehud ; SOLAN, Zach ; WOLFMAN, Gadi ; RUPPIN, Eytan: Placing Search in Context: The Concept Revisited. In: *ACM Transactions on Information Systems* 20 (2002), Nr. 1, S. 116–131
- [Francis u. a. 1982] FRANCIS, W.N. ; KUČERA, H. ; MACKIE, A.W.: *Frequency Analysis of English Usage: Lexicon and Grammar*. Houghton Mifflin Boston, MA:, 1982

- [Frank u. a. 2007] FRANK, Anette ; KRIEGER, Hans-Ulrich ; XU, Feiyu ; USZKOREIT, Hans ; CRYSMANN, Berthold ; JÖRG, Brigitte ; SCHÄFER, Ulrich: Question answering from structured knowledge sources. In: *Journal of Applied Logic* 5 (2007), Nr. 1, S. 20–48
- [Golub und Kahan 2007] GOLUB, Gene ; KAHAN, William: Calculating the singular values and pseudo-inverse of a matrix. In: *Milestones in Matrix Computation: The Selected Works of Gene H. Golub with Commentaries* (2007)
- [Gurevych 2006] GUREVYCH, Iryna: Computing Semantic Relatedness Across Parts of Speech / University of Technology Darmstadt, Department of Computer Science. 2006. – Forschungsbericht
- [Hirst und St-Onge 1998] HIRST, Graeme ; ST-ONGE, David: Lexical chains as representations of context for the detection and correction of malapropisms. In: *WordNet: An Electronic Lexical Database* (1998), S. 305–332
- [Hornby 1995] HORNBY, Albert S.: *Oxford advanced learner's dictionary*. Oxford University Press, 1995
- [Jaro 1995] JARO, Matthew A.: Probabilistic linkage of large public health data files. In: *Statistics in Medicine* 14 (1995), Nr. 5-7, S. 491–8
- [Jiang und Conrath 1997] JIANG, Jay J. ; CONRATH, David W.: Semantic similarity based on corpus statistics and lexical taxonomy. In: *Proceedings of International Conference on Research in Computational Linguistics* Bd. 33 Taiwan (Veranst.), 1997
- [Kandel 1986] KANDEL, Abraham: *Fuzzy mathematical techniques with applications*. Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA, 1986
- [Katz 1987] KATZ, S.: Estimation of probabilities from sparse data for the language model component of a speech recognizer. In: *Acoustics, Speech, and Signal Processing [see also IEEE Transactions on Signal Processing]*, *IEEE Transactions on* 35 (1987), Nr. 3, S. 400–401
- [Kendall 1938] KENDALL, Maurice: A new measure of rank correlation. In: *Biometrika* 30 (1938), S. 81–93
- [Kim und Candan 2006] KIM, Jong W. ; CANDAN, K. S.: CP/CV: concept similarity mining without frequency information from domain describing taxonomies. In: *Proceedings of the 15th ACM international conference on Information and knowledge management* ACM Press New York, NY, USA (Veranst.), 2006, S. 483–492
- [Kingsbury und Palmer 2002] KINGSBURY, Paul ; PALMER, Martha: From treebank to propbank. In: *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC-2002)*, 2002, S. 1989–1993
- [Kipfer 2001] KIPFER, Babara A. (Hrsg.): *Roget's International Thesaurus*. Bd. 6. Harper Resource, 2001

- [Kozima und Furugori 1993] KOZIMA, Hideki ; FURUGORI, Teiji: Similarity between words computed by spreading activation on an English dictionary. In: *Proceedings of the sixth conference on European chapter of the Association for Computational Linguistics, April, 1993*, S. 21–23
- [Lapata 2006] LAPATA, Mirella: Automatic Evaluation of Information Ordering: Kendall's Tau. In: *Computational Linguistics* 32 (2006), Nr. 4, S. 471–484
- [Leacock und Chodorow 1998] LEACOCK, Claudia ; CHODOROW, Martin: Combining local context and WordNet similarity for word sense identification. In: *WordNet: An Electronic Lexical Database* 49 (1998), Nr. 2, S. 265–283
- [Lee u. a. 1993] LEE, Yoon J. ; JOON, Ho L. ; MYOUNG, Ho K.: Information Retrieval Based On Conceptual Distance In IS-A Hierachies. In: *Journal of Documentation* 49 (1993), Nr. 2, S. 188–207
- [Lin 1998] LIN, Dekang: An information-theoretic definition of similarity. In: *Proceedings of the Fifteenth International Conference on Machine Learning, 1998*, S. 296–304
- [Maguitman u. a. 2006] MAGUITMAN, Ana G. ; MENCZER, Filippo ; ERDINC, Fulya ; ROINESTAD, Heather ; VESPIGNANI, Alessandro: Algorithmic Computation and Approximation of Semantic Similarity. In: *World Wide Web* 9 (2006), Nr. 4, S. 431–456
- [Maguitman u. a. 2005] MAGUITMAN, Ana G. ; MENCZER, Filippo ; ROINESTAD, Heather ; VESPIGNANI, Alessandro: Algorithmic detection of semantic similarity. In: *Proceedings of the 14th international conference on World Wide Web* ACM New York, NY, USA (Veranst.), 2005, S. 107–116
- [Miller u. a. 1990] MILLER, George A. ; BECKWITH, Richard ; FELLBAUM, Christiane ; GROSS, Derek ; MILLER, Katherine J.: Introduction to WordNet: An On-line Lexical Database. In: *International Journal of Lexicography* 3 (1990), Nr. 4, S. 235–244
- [Miller und Charles 1991] MILLER, George A. ; CHARLES, Walter G.: Contextual correlates of semantic similarity. In: *Language and Cognitive Processes* 6 (1991), Nr. 1, S. 1–28
- [Morris 1988] MORRIS, Jane: Lexical Cohesion, the Thesaurus, and the Structure of Text / Computer Systems Research Institute, University of Toronto. 1988. – Forschungsbericht
- [Morris und Hirst 1991] MORRIS, Jane ; HIRST, Graeme: Lexical cohesion computed by thesaural relations as an indicator of the structure of text. In: *Computational Linguistics* 17 (1991), Nr. 1, S. 21–48
- [Morris und Hirst 2004] MORRIS, Jane ; HIRST, Graeme: Non-classical lexical semantic relations. In: *Workshop on Computational Lexical Semantics, Human Language Technology Conference of the North American Chapter of the ACL, Boston, 2004*

- [Narayanan und Harabagiu 2004] NARAYANAN, Srinu ; HARABAGIU, Sandra: Answering questions using advanced semantics and probabilistic inference. In: HARABAGIU, Sandra (Hrsg.) ; LACATUSU, Finley (Hrsg.): *HLT-NAACL 2004: Workshop on Pragmatics of Question Answering*. Boston, Massachusetts, USA, May 2004, S. 10–16
- [Ohara u. a. 2004] OHARA, K. ; FUJII, S. ; OHORI, T. ; SUZUKI, R. ; SAITO, H. ; ISHIZAK, S.: The Japanese FrameNet project: An introduction. In: *Proceedings of the Workshop on Building Lexical Resources from Semantically Annotated Corpora at LREC 2004*, 2004
- [Pennacchiotti u. a. 2008] PENNACCHIOTTI, Marco ; DE CAO, Diego ; MAROCCO, Paolo ; BASILI, Roberto: Towards a Vector Space Model for FrameNet-like Resources. In: *Proceedings of LREC* Bd. 2008, 2008
- [Pennacchiotti und Wirth 2009] PENNACCHIOTTI, Marco ; WIRTH, Michael: Measuring frame relatedness. In: *Proceeding of the 12th Conference of the European Chapter of the Association of Computational Linguistics*. Athen, Greece, 2009. – to appear
- [Quillian 1966] QUILLIAN, M. Ross: *Semantic Memory*. Air Force Cambridge Research Laboratories, Office of Aerospace Research, United States Air Force, 1966
- [Rada u. a. 1989] RADA, Roy ; MILI, Hafedh ; BICKNELL, Ellen ; BLETNER, Maria: Development and application of a metric on semantic nets. In: *IEEE Transactions on Systems, Man, and Cybernetics* 19 (1989), Nr. 1, S. 17–30
- [Resnik 1995] RESNIK, Philip: Using Information Content to Evaluate Semantic Similarity in a Taxonomy. In: *International Joint Conference On Artificial Intelligence* Bd. 14 Lawrence Erlbaum Associates Ltd (Veranst.), 1995, S. 448–453
- [Ross 1976] ROSS, Sheldon M.: *A First Course in Probability*. Macmillan New York, 1976
- [Rubenstein und Goodenough 1965] RUBENSTEIN, Herbert ; GOODENOUGH, John B.: Contextual correlates of synonymy. In: *Communications of the ACM* 8 (1965), Nr. 10, S. 627–633
- [Ruppenhofer u. a. 2006] RUPPENHOFER, Joseph ; ELLSWORTH, Michael ; PETRUCK, Miriam R. ; JOHNSON, Christopher R. ; SCHEFFCZYK, Jan: FrameNet II: Extended Theory and Practice. In: *International Computer Science Institute, Berkeley, USA* (2006)
- [Shen und Lapata 2007] SHEN, Dan ; LAPATA, Mirella: Using Semantic Roles to Improve Question Answering. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing and on Computational Natural Language Learning*, 2007, S. 12–21
- [Siegel und Castellan 1988] SIEGEL, Sidney ; CASTELLAN, N. J.: *Nonparametric statistics for the behavioral sciences*. McGraw-Hill New York, 1988

- [Stevenson und Wilks 2001] STEVENSON, Mark ; WILKS, Yorick: The Interaction of Knowledge Sources in Word Sense Disambiguation. In: *Computational Linguistics* 27 (2001), Nr. 3, S. 321–349
- [Subirats und Petruck 2003] SUBIRATS, Carlos ; PETRUCK, Miriam: Surprise! Spanish FrameNet! In: *Proceedings of the Workshop on Frame Semantics at the XVII. International Congress of Linguistics*. Prague, 2003
- [Summers 2003] SUMMERS, Della ; PROCTER, Paul (Hrsg.): *Longman Dictionary of Contemporary English*. 2003
- [Sussna 1993] SUSSNA, Michael J.: Word sense disambiguation for free-text indexing using a massive semantic network. In: *Proceedings of the second international conference on Information and knowledge management* ACM New York, NY, USA (Veranst.), 1993, S. 67–74
- [Sussna 1997] SUSSNA, Michael J.: *Text retrieval using inference in semantic metanetworks*. San Diego La Jolla, CA, USA, University of California, Dissertation, 1997
- [Tonelli und Pianta 2008] TONELLI, Sara ; PIANTA, Emanuele: Frame Information Transfer from English to Italian. In: *6th International Conference on Language Resources and Evaluation (LREC 2008)*, 2008
- [Winkler 1999] WINKLER, William E.: The state of record linkage and current research problems. In: *Statistics of Income Division, Internal Revenue Service Publication R 4* (1999)
- [Wu und Palmer 1994] WU, Zhibiao ; PALMER, Martha: Verbs semantics and lexical selection. In: *Proceedings of the 32nd annual meeting on Association for Computational Linguistics* Association for Computational Linguistics Morristown, NJ, USA (Veranst.), 1994, S. 133–138