

Kontext- und Korreferenzanalyse für gesprochene Sprache

Masterarbeit
von

Tobias Hey

An der Fakultät für Informatik
Institut für Programmstrukturen
und Datenorganisation (IPD)

Erstgutachter:	Prof. Dr. Walter F. Tichy
Zweitgutachter:	Prof. Dr. Ralf H. Reussner
Betreuender Mitarbeiter:	Dipl.-Inform. Sebastian Weigelt

Bearbeitungszeit: 01.04.2016 – 30.09.2016

Ich versichere wahrheitsgemäß, die Arbeit selbstständig angefertigt, alle benutzten Hilfsmittel vollständig und genau angegeben und alles kenntlich gemacht zu haben, was aus Arbeiten anderer unverändert oder mit Abänderungen entnommen wurde.

Die Regeln zur Sicherung guter wissenschaftlicher Praxis im Karlsruher Institut für Technologie (KIT) habe ich befolgt.

Karlsruhe, 30.09.2016

.....
(**Tobias Hey**)

Kurzfassung

Menschen nutzen zum Verständnis ihres Gegenüber den Kontext der Konversation und ihr erlerntes Wissen über Zusammenhänge. Deshalb sollte auch Rechnern zum Verständnis von gesprochener Sprache ein Kontextwissen zur Verfügung stehen. Hierfür wird im Rahmen des *PARSE*-Projektes ein Kontextanalysewerkzeug vorgestellt, welches ein Kontextwissen aufbaut. Dies schließt den Aufbau einer Repräsentation des zuletzt Gesagten sowie eine Konzeptbildung mit ein. Dies ermöglicht es, die gebildeten Konzepte um Beziehungen untereinander anzureichern und schafft es somit ein umfassenderes Abbild der Zusammenhänge zu bieten. Um eine der vielfältigen Anwendungsmöglichkeiten des Kontextwissens zu verdeutlichen, wird eine kontextbasierte Auflösung von Referenzen umgesetzt. Diese setzt auf eine Siebarchitektur, welche in der Lage ist, je nach gegebenen Informationen die Kandidaten einer Referenz einzugrenzen. Die Funktionalität wird anschließend auf händischen Transkriptionen von Anweisungen an einen Roboter evaluiert. Die Auflösung von Referenzen erreicht dabei ein F1-Maß von 0,88. Das Referenzsystem *Stanford Deterministic Coreference Resolution System*, welches den Stand der Technik darstellt, erreicht hingegen nur einen Wert von 0,64. Kontextwissen kann also maßgeblich zum Verständnis gesprochener Sprache beitragen. Dieses Verständnis könnte zukünftig mit einer vom Kontext unterstützten Wortbedeutungsanalyse noch weiter ausgebaut werden.

Inhaltsverzeichnis

1. Einleitung	1
1.1. Natürlichsprachliche Programmierung	2
1.2. Kontext- und Korreferenzanalyse	2
1.3. Aufbau der Arbeit	2
2. Grundlagen	3
2.1. Sprachwissenschaftliche Grundbegriffe	3
2.1.1. Natürliche Sprache	3
2.1.2. Syntax	3
2.1.3. Semantik	4
2.1.4. Pragmatik	4
2.1.5. Entität	4
2.1.6. Diskurs	4
2.1.7. Kohärenz	5
2.1.8. Lexikalische Semantik	5
2.2. Kontext in der Sprachwissenschaft	6
2.2.1. Kotext	6
2.2.2. Kontextualisierung	7
2.3. Natural Language Processing	7
2.3.1. Wortart-Markierung	7
2.3.2. Phrasenerkennung	8
2.3.3. Eigennamenerkennung	9
2.3.4. Markieren semantischer Rollen	9
2.4. Korreferenzanalyse	10
2.4.1. Natürlichsprachliche Referenzen	10
2.4.2. Stanford Deterministic Coreference Resolution System	12
2.5. Lexikalische Datenbanken	12
2.5.1. WordNet	13
2.5.2. Proposition Bank	13
2.5.3. VerbNet	14
2.5.4. FrameNet	15
2.6. Ontologien	15
3. PARSE: Programming ARchitecture for Spoken Explanations	17
3.1. Besonderheiten der gesprochenen Sprache	17
3.2. Projektarchitektur	18
3.3. Bisheriger Projektverlauf	18
3.4. Domänenontologie	19
4. Verwandte Arbeiten	21
4.1. Programmieren in natürlicher Sprache	21
4.2. Wissens- und Informationsextraktion	24

4.3. Korreferenzanalyse unterstützt durch Kontextwissen	25
5. Analyse und Entwurf	27
5.1. Diskussion der Verwandten Arbeiten	28
5.2. Kontextanalyse	29
5.2.1. Kontextdefinition	30
5.2.2. Kontextinformationen im sprachlichen Kontext	30
5.2.3. Kontextgenerierung	33
5.2.3.1. Entitätenerkennung	34
5.2.3.2. Aktionserkennung	39
5.2.3.3. Konzeptualisierung	41
5.2.3.4. Zustandsanalyse	45
5.2.4. Repräsentation des Kontextwissens	48
5.3. Korreferenzanalyse	49
5.3.1. Stand der Forschung	50
5.3.2. Kontextbasierte Korreferenzanalyse	51
5.3.3. Pronominale Anapherauflösung	52
5.3.3.1. Vorangehende Entitäten	52
5.3.3.2. Numerus	53
5.3.3.3. Bezugsrahmen	53
5.3.3.4. Bedingungen	54
5.3.3.5. Gleiche Aktion	54
5.3.3.6. Typübereinstimmung	54
5.3.3.7. Geschlecht	55
5.3.3.8. Ausführbarkeit von Aktionen	56
5.3.3.9. Ortsbeziehungen	56
5.3.3.10. Übereinstimmung semantischer Rollen	57
5.3.3.11. Abstand zum referierenden Ausdruck	57
5.3.3.12. Reflexiv- und Possessivpronomen	58
5.3.4. Entwurf der pronominalen Anapherauflösung	58
5.3.5. Identitätsauflösung	60
5.3.5.1. Konzeptübereinstimmung	61
5.3.5.2. Subjektidentitätsauflösung	61
5.3.5.3. Objektidentitätsauflösung	62
6. Implementierung	65
6.1. Anpassungen an der Vorverarbeitung	65
6.1.1. Eigennamenerkennung	65
6.1.2. Semantische-Rollen-Markierer	66
6.2. Kontextanalyse	68
6.2.1. Entitätenerkennung	69
6.2.2. Aktionserkennung	71
6.2.3. Ortsbeziehungserkennung	73
6.2.4. Welt- und Domänenwissenanreicherung	74
6.2.5. Zustandsbestimmer	78
6.3. Korreferenzanalyse	79
6.3.1. Pronominale Anapherauflösung	84
6.3.2. Subjektidentitätsauflösung	87
6.3.3. Objektidentitätsauflösung	87
7. Evaluation	89
7.1. Evaluationskorporus	89

7.2.	Vorgehensweise	90
7.3.	Kontextanalyse	91
7.3.1.	Entitätenerkennung	92
7.3.2.	Aktionserkennung	93
7.3.3.	Ortsbeziehungserkennung	94
7.3.4.	Welt- und Domänenwissenanreicherung	95
7.3.5.	Zustandsbestimmung	97
7.4.	Korreferenzanalyse	98
7.4.1.	Referenzkorreferenzer	98
7.4.2.	Basis-Korreferenzanalyse	99
7.4.3.	Kontextbasierte Korreferenzanalyse	100
7.5.	Bewertung der Ergebnisse	103
8.	Zusammenfassung und Ausblick	105
	Literaturverzeichnis	109
	Anhang	115
A.	Tagsets und Rolesets	115
A.1.	Penn Treebank Tagset	115
A.2.	VerbNet Thematic Roles	116
B.	PARSE	118
B.1.	Domänenontologie-Schnittstelle	118
B.2.	Domänenontologie-Abbild	119
C.	Konfigurationsdatei der Korreferenzimplementierung	123
D.	Erweiterung des Korpus	124
D.1.	Szenario 1: Fill the cup	124
D.2.	Szenario 2: Prepare a meal	124
E.	Musterlösung der Korreferenzanalyse	126

Abbildungsverzeichnis

3.1. Vereinfachte Projekt-Architektur (basierend auf [WT15])	18
4.1. Beispiel für den Einsatz von situativem Kontext zur Verb-Konzept Abbildung [MTLS15]	22
4.2. Ablauf des Inferenzprozesses [BUCW10]	23
4.3. Verbtyp und Gliedsatztyp-Erkennung [DG13]	24
5.1. Beispiel für natürlichsprachliche Referenzen	28
5.2. Hierarchieebenen des Kontextwissens	32
5.3. Schritte der Entitätenerkennung	39
5.4. Ergebnisse des SRL in Verbindung mit Aktionen und Entitäten	40
5.5. Schritte der Aktionserkennung	41
5.6. Abstraktion eines Konzeptes aus Objektentitäten	42
5.7. Mit Domänenontologie erweiterte Konzeptualisierungshierarchie	44
5.8. Übersicht über den Ablauf der Konzeptualisierung	45
5.9. Repräsentationsebenen des Kontextes	49
5.10. Übersicht über den Ablauf der pronominalen Anapherauflösung	59
5.11. Schritte der Identitätsauflösung	63
6.1. Vorverarbeitungsfließband-Ausgabe nach Ausführung des Eigennamenerkenners	66
6.2. Vorverarbeitungsfließband-Ausgabe nach Ausführung des Semantische-Rollen-Markierers	67
6.3. Aufbau des Kontextanalysierers mit angedeutetem Kontext	68
6.4. Repräsentation des Kontextwissens	70
6.5. Ergebnis des Entitätenerkenners	72
6.6. Ergebnis des Kontext nach dem Aktionserkennung	73
6.7. Schematischer Ablauf der Konzeptbildung	74
6.8. Ergebnis des Kontext nach der Domänenanbindung	76
6.9. Iterationsprozess der Überkonzeptbildung	77
6.10. Ablauf der Kontextanalyse	79
6.11. Aufbau des Korreferenzanalysierers	81
6.12. Übersicht der Siebe und ihrer Reihenfolge für die einzelnen Analyse-Systeme	82
6.13. Beispielhafte Repräsentation der Referenzen im Graph	84
6.14. Beispiel einer Objektpronomen-Anapherauflösung mit den entsprechenden Sieben	85
7.1. Übersicht der Ergebnisse der Korreferenzvarianten	102
B.1. Ausschnitt aus der Domänenontologie-Schnittstelle	118
D.2. Pictures of the fridge	125
D.3. Table and dishwasher	125
D.4. View at the kitchen with marked microwave M	125

Tabellenverzeichnis

2.1. Auszug aus dem Penn-Tagset	7
2.2. Phrasen-Markierungen	8
2.3. Eigennamen-Markierungen	9
2.4. Argumentübersetzung nach PropBank Eintrag zu Prädikat „bring“	10
2.5. Einige häufig verwendete thematische Rollen	14
3.1. Die Struktur der Domänenontologie	19
5.1. Präpositionalphrasen mit Ortsbestimmung	38
5.2. Typen von Personalpronomen und ihre möglichen Bezugsobjekte	55
6.1. Umgesetzte Eigenschaften und die entsprechenden Siebe	83
7.1. Informationen über neue Sprachaufnahmen	90
7.2. Ergebnisse der Teilschritte der Kontextanalyse	92
7.3. Ergebnisse der Evaluation der Aktionserkennung	93
7.4. Ergebnisse der Evaluation des Stanford Korreferenzers	99
7.5. Ergebnisse der Evaluation der Basis-Variante der Korreferenzanalyse	100
7.6. Ergebnisse der Evaluation der kontextbasierten Korreferenzanalyse	100

1. Einleitung

In der heutigen Zeit ist unser Alltag in nahezu allen Bereichen von Software durchdrungen, denn sie steuert unsere Geräte vom Wecker bis hin zum Smartphone. Doch nur wenige Personen sind in der Lage Software mithilfe von Programmiersprachen zu erstellen oder zu verstehen. Bedingt durch die fortwährende Automatisierung und den vermehrten Einsatz von Robotern wird diese Fähigkeit allerdings in Zukunft noch mehr an Bedeutung gewinnen. Es liegt also nahe das Programmieren von Software mithilfe der eigenen Ausdrucksweise des Menschen zu ermöglichen, der natürlichen Sprache. Gerade der Umgang mit interaktiven Maschinen wie z.B. Haushaltsrobotern wäre so in einer natürlichen und dem Menschen leicht zugänglichen Art und Weise möglich, die es selbst nicht technik-affinen Personen erlaubt diesen Maschinen neue Anweisungen und Aufgaben beizubringen.

Die Herausforderungen, welche bei der Umsetzung von natürlicher Sprache in Quelltext entstehen, sind allerdings vielfältig. Die natürliche Sprache lässt viele Formulierungen zu, welche für einen Rechner nicht ohne Weiteres eindeutig auflösbar sind.

- (1) Armar, please go to the dishwasher and open it.

Ein Beispiel hierfür sind Referenzen wie z.B. die Anapher *it* in Beispiel 1. Würde man diese Referenzen nicht auflösen, wäre bei vielen Anweisungen nicht klar, auf welche Objekte oder welche Aktionen sie sich beziehen, wodurch es nicht möglich wäre eine korrekte und vollständige Abbildung der Anweisungen in Quelltext zu gewährleisten.

Das menschliche Gehirn löst diese Referenzen indem es auf die semantischen Zusammenhänge und das erlernte Wissen über die strukturellen Gegebenheiten in der natürlichen Sprache zurückgreift [Fel13]. Gerade die semantischen Zusammenhänge hängen jedoch stark vom Kontext ab, in dem die Konversation stattfindet. Dieser umfasst unter anderem Informationen wie das zuletzt Gesagte, das Thema der Konversation, den Gesprächspartner oder aber den Zustand der Umwelt in der man sich gerade befindet.

Möchte man also eine vollständige Abbildung der Anweisungen in Quelltext durchführen, ist es nötig auch eine möglichst vollständige Vorstellung des Kontextes zu besitzen, in dem die Aussagen getroffen werden. Denn hierdurch bietet sich die Möglichkeit implizite Verweise und Zusammenhänge zu erkennen und dadurch Herausforderungen wie das Auflösen von Referenzen besser behandeln zu können.

1.1. Natürlichsprachliche Programmierung

Diese Arbeit ist Teil des Projektes *Programming Architecture for Spoken Explanations*, kurz *PARSE* (siehe Kapitel 3), des IPD Tichy am KIT [Kar]. Das Projekt beschäftigt sich mit der Aufgabe in gesprochener natürlicher Sprache formulierte Anweisungen in, von einem Zielsystem, verständlichen Quelltext zu übersetzen. Somit ist die Aufgabe im Feld der natürlichsprachlichen Programmierung angesiedelt.

Um diese Abbildung durchführen zu können, werden verschiedene Verarbeitungsschritte der Rechnerlinguistik angewandt, wie die Syntax- und Semantikanalyse. Gerade die Semantikanalyse muss hierbei für eine korrekte Abbildung ein möglichst korrektes Verständnis des Gesagten schaffen, um zu verhindern, dass der generierte Quelltext Fehler enthält. Daher ist es nötig Mehrdeutigkeiten aufzulösen und auch Unklarheiten wie Referenzen zu beseitigen.

Daher wird im Rahmen dieser Arbeit als weiterer Schritt der Semantikanalyse, eine Kontextanalyse umgesetzt. Diese versucht ein möglichst umfassendes Abbild des Gesagten aufzubauen und dieses mit Kontextwissen anzureichern. Hierdurch sollen die weiteren Schritte der Umsetzung in Quelltext unterstützt werden, welches am Beispiel des Auflöserns von Referenzen evaluiert wird.

1.2. Kontext- und Korreferenzanalyse

In der Arbeit werden zunächst die im *PARSE*-Projekt vorkommenden natürlichsprachlichen Anweisungen hinsichtlich eines in ihnen enthaltenen Kontextwissen analysiert, eine Repräsentation dieses Wissens geschaffen und dieses wiederum mit Weltwissen angereichert. Die Referenzauflösung wird hierbei als Anwendungsbeispiel verwendet, da es hier viele Mehrdeutigkeiten gibt, die ohne dieses Wissen nur schwierig aufgelöst werden können und somit beurteilt werden kann, welches Potenzial Kontextwissen bietet.

In dieser Arbeit wird also ein Analyse-Werkzeug erstellt, welches die Eingabe auf Kontextinformationen hin analysiert und die gewonnenen Informationen geeignet in den im Projekt als geteilten Datenspeicher genutzten Graph einfließen lässt. Mit Kontextinformationen sind in diesem Fall Informationen gemeint, welche die Zusammenhänge des Gesagten darstellen. In erster Linie beinhaltet dies die Informationen aus den zuletzt gegebenen Anweisungen, wie z.B. die existenten Objekte, deren Zustände und durchgeführten Aktionen, aber auch allgemeines Wissen über Zusammenhänge zwischen beschriebenen Objekten und Aktionen, wie z.B. mögliche Teil-Ganzes Beziehungen zwischen Entitäten.

Als zweiten Schritt nach dem Aufbau des Kontextwissens soll ein weiteres Werkzeug erstellt werden, dass mithilfe der gewonnenen Kontextinformationen die Auflösung von Referenzen durchführt.

1.3. Aufbau der Arbeit

Diese Arbeit ist wie folgt strukturiert: Zunächst werden in Kapitel 2 die für das Verständnis des Vorgehens nötigen Grundlagen und verwendeten Werkzeuge eingeführt. Daraufhin wird in Kapitel 3 ein kurzer Überblick zum *PARSE*-Projekt und seinem aktuellen Stand gegeben. In Kapitel 4 werden verschiedene mit dieser Arbeit verwandte Arbeiten vorgestellt und erläutert. Das Analyse-Kapitel (Kapitel 5) beschreibt die Schlüsse, die für die Kontext- und Korreferenzanalyse gezogen werden können und gibt ein Konzept für die entstehenden Werkzeuge. Die eigentliche Implementierung der Werkzeuge wird daraufhin in Kapitel 6 erörtert. Ob die erwarteten Ergebnisse mittels der Werkzeuge erreicht wurden, wird in Kapitel 7 diskutiert und Kapitel 8 schließt die Arbeit mit einer Zusammenfassung der Ergebnisse und einem Ausblick ab.

2. Grundlagen

Dieses Kapitel gibt einen Einblick in die für das Verständnis dieser Arbeit essentiellen Konzepte. Hierzu gehören Grundlagen im Bereich der Sprachwissenschaften (Abschnitt 2.1), insbesondere der Kontextanalyse (Abschnitt 2.2) sowie der Korreferenz (Abschnitt 2.4). Des Weiteren werden verschiedene in dieser Arbeit genutzte Werkzeuge, wie Werkzeuge der computergestützten Verarbeitung natürlicher Sprache (Abschnitt 2.3) und lexikalische Datenbanken (Abschnitt 2.5) eingeführt. Außerdem wird eine kurze Einführung in das Konzept der Ontologie (Abschnitt 2.6) gegeben.

2.1. Sprachwissenschaftliche Grundbegriffe

Um die in dieser Arbeit behandelten sprachwissenschaftlichen Konzepte und Zusammenhänge verstehen zu können, werden in diesem Abschnitt erste Grundbegriffe aus der Sprachwissenschaft eingeführt.

2.1.1. Natürliche Sprache

Mit natürlicher Sprache wird in der Sprachwissenschaft die gesprochene oder anderweitig geäußerte Form der Kommunikation des Menschen verstanden, welche regional und sozial geschichtet aus einer historischen Entwicklung einer Verständigungsweise entstanden ist. Sie steht im Kontrast zur künstlichen Sprache, welche alle Formen von Sprachen beinhaltet, die gezielt für einen Verwendungszweck und ohne diese Entwicklung geschaffen wurden, wie z.B. Programmiersprachen. Natürliche Sprachen unterscheiden sich hauptsächlich dadurch von künstlichen Sprachen, dass sie lexikalische und strukturelle Mehrdeutigkeiten besitzen bzw. eine Vagheit oder Bedeutungsvielfalt ihrer Ausdrücke existiert [Buß02]. Natürliche Sprachen besitzen ein Regelwerk, welches vorgibt, welche Zeichenfolgen zugelassen sind (Syntax) und was deren Bedeutung ist (Semantik), die sogenannte Grammatik.

2.1.2. Syntax

In der Sprachwissenschaft bezeichnet der Begriff Syntax die Lehre vom Bau einer Texteinheit und ist damit Teil der Grammatik natürlicher Sprachen. Sie definiert dabei wie und in welcher Form Zeichen und Wörter zu übergeordneten Einheiten wie Phrasen-, Teilsätzen oder ganzen Sätzen zusammengesetzt werden dürfen und bildet damit ein Regelwerk ab, welches die Struktur eines Satzes vorgibt [Buß02]. In der englischen Sprache ist eines der wichtigsten Konzepte dieses Regelwerks, die festgelegte Satzgliedreihenfolge *Subjekt-Prädikat-Objekt*, welche jedem satzwertigen Teil der Sprache zugrunde liegen muss um

ihn als syntaktisch korrekt bezeichnen zu können. Außerdem bildet die Syntax zusammen mit der Semantik und Pragmatik einen der Teilbereiche der Zeichentheorie, der Semiotik [Mor38]. Hierin beschäftigt sie sich mit der Anordnung und Beziehung von Zeichen untereinander [Buß02].

2.1.3. Semantik

Mit dem Begriff der Semantik wird im sprachwissenschaftlichen Umfeld die Lehre von der Beziehung zwischen Zeichen und ihrer Bedeutung (in der Semiotik) sowie in der Linguistik die Lehre von der wörtlichen Bedeutung von sprachlichen Ausdrücken allgemein verstanden [Buß02]. Sie beschäftigt sich also mit die Bedeutung betreffenden Beziehungen zwischen Zeichen, Wörtern oder Phrasen. So besteht eines der Felder der Semantik darin zu verstehen, wie die Bedeutung komplexer Begriffe aus einfacheren Begriffen zusammengesetzt werden kann, also z.B. die Bedeutung einer Phrase aus den Teilbedeutungen seiner bildenden Elemente abgeleitet werden kann.

So sind Formulierungen immer dann semantisch korrekt, wenn diese eine sinnhafte Bedeutung besitzen [PK08].

2.1.4. Pragmatik

Die Pragmatik beschäftigt sich hingegen mit der Bedeutung und dem Zweck einer Formulierung in einer bestimmten Situation. Diese Situation kann vom Sprecher, der Äußerungsform und den Gegebenheiten zum Zeitpunkt der Äußerung abhängen.

(2) You have to read this book!

So kann der Satz in Beispiel 2 sowohl als Befehl gemeint sein, als auch eine Empfehlung darstellen, je nachdem ob er von einem Freund oder einem Lehrer an einen Schüler geäußert wurde [PK08].

2.1.5. Entität

Ein weiterer für das Verständnis dieser Arbeit essentieller Begriff ist der Begriff der Entität, welcher ursprünglich aus der Philosophie und der dortigen Lehre des Seins (Ontologie) entstammt. Hier bezeichnet eine Entität einen Sammelbegriff für alle konkret oder abstrakt existierenden, also *Seienden* Dinge. Zumeist sind dies real oder fiktiv existierende Personen oder Objekte, welche sich dadurch auszeichnen, dass sie existieren können.

2.1.6. Diskurs

Der Begriff Diskurs beschreibt in der Sprachwissenschaft unterschiedliche Aspekte. Die für diese Arbeit passendsten beschriebenen Aspekte sind Diskurs als Begriff für eine konversationelle Interaktion [CC14] sowie als Begriff für geäußerten Text [Van76]. Somit besteht ein Diskurs aus einer Menge von Aussagen, die zusammen eine Interaktion aufweisen und thematisch verbunden sind. So stellen zusammengehörige Texte, wie z.B. Nachrichtenartikel einen Diskurs dar, einzeln geäußerte Sätze ohne Zusammenhang hingegen nicht.

Vor allem werden Diskurse hinsichtlich ihrer Semantik betrachtet. Hiermit ist die Untersuchung von Diskurs-bildenden Strukturen gemeint, also die Untersuchung der Teile des Textes, welche dafür sorgen, dass eine Aneinanderreihung von Aussagen als zusammengehörig wahrgenommen wird. Die Theorie der Diskurssemantik beschäftigt sich dabei hauptsächlich mit satzübergreifenden Zusammenhängen, wie z.B. Anaphern [Buß02].

Um die gewonnenen Ergebnisse darzustellen hat sich in der Sprachwissenschaft eine formale Darstellung von Texten ausgeprägt, die sogenannte Diskursrepräsentation, welche die

mentalen Modelle darstellt, die der interpretierende Akteur bei der Aufnahme des Textes in dynamischer Weise erstellt [Buß02]. Die vorzugsweise genutzte Diskursrepräsentation basiert auf der Diskursrepräsentationstheorie entwickelt von Kamp [Kam81] und ordnet Diskursen sogenannte **Diskursrepräsentationstrukturen** (DRS) und diesen wiederum Wahrheitsbedingungen zu. Ein wichtiger Aspekt hierbei ist das Konzept des **Diskursreferenten** „[...] eine Art Platzhalter für Objekte, auf die die verschiedenen Textprädikationen, auch in verschiedenen Sätzen (Textanaphern), bezogen werden [...]“ [Buß02], also konzeptionelle Entitäten, welche die eigentlichen Entitäten in der beschriebenen Welt repräsentieren.

2.1.7. Kohärenz

Kohärenz bezeichnet in der Sprachwissenschaft den semantisch-kognitiven Sinnzusammenhang eines Textes. Also den Grad zu dem die Teile eines Textes zusammenhängen. Kohärenz wird hierbei hauptsächlich durch zwei Aspekte von Referenz gebildet: Zum einen durch das wiederholte Bezug nehmen auf die gleiche Entität, auch **Korreferenz** genannt (siehe Abschnitt 2.4) und zum anderen durch die „strukturierte textuelle Entwicklung der Informationen in den verschiedenen Referenzbereichen (z.B. Erhalt, Verschiebung, Neueinführung)“ [Buß02]. Kohärenz ist somit eine wichtige Voraussetzung für die Wahrnehmung von Ausdrücken und Sätzen als Text bzw. Diskurs.

2.1.8. Lexikalische Semantik

Lexikalische Semantik ist eine Forschungsrichtung innerhalb der Sprachwissenschaften, welche sich mit den verschiedenen Bedeutungen von Wörtern und ihren bedeutungsgebundenen Beziehungen beschäftigt. Wörter werden in diesem Feld über zwei Begriffe definiert. Zum einen bezeichnet man mit **Lexem** eine Paarung einer bestimmten grammatikalischen Wortform mit ihrer Bedeutung. Zum anderen wird ein solches Lexem in der lexikalischen Semantik mittels seines **Lemma** repräsentiert. Das Lemma eines Lexemes ist seine Grundform. Beispielsweise wäre das Lemma von *Türen* die Grundform *Tür*. Die Umwandlung eines Lexems in sein Lemma wird in der Sprachwissenschaft als **Lemmatisierung** bezeichnet. Der Prozess der Lemmatisierung ist hierbei im Allgemeinen nicht deterministisch, da das Lemma eines Lexems vom Kontext und der sich aus dem Kontext ergebenden Wortart des Lexems abhängt [JM09, Kap. 19].

Im Feld der lexikalischen Semantik werden Beziehungen zwischen den Bedeutungen von natürlichsprachlichen Ausdrücken als semantische Relationen bezeichnet. Im Folgenden wird ein kurzer Überblick über die für diese Arbeit relevanten semantischen Relationen gegeben.

Synonymie

Zwei sprachliche Ausdrücke werden als **synonym** bezeichnet, wenn eine semantische Relation zwischen ihnen besteht, die eine **Bedeutungsgleichheit** dieser Ausdrücke darstellt. Synonyme können hierbei aus einfachen Variationsmöglichkeiten in der Sprache resultieren, wie z.B. bei „*erhalten* und *bekommen*“, oder aber ihren Ursprung in regionalen (*Pilz* und *Schwammerl*), soziodialektalen (*Geld* und *Moneten*), stilistischen (*Raum* und *Gemach*) oder fachsprachlichen (*bedeutungsgleich* und *synonym*) Gegebenheiten haben [Buß02]. Außerdem wird zwischen perfekten und nicht-perfekten Synonymen unterschieden. Zwei sprachliche Ausdrücke sind nur dann perfekte Synonyme, wenn sie in jedem Kontext erfolgreich bzw. sinnerhaltend gegeneinander ausgetauscht werden können.

Hyponymie und Hyperonymie

Mit Hyponymie wird die semantische Relation der **Unterordnung** bezeichnet. Diese Unterordnung bezieht sich auf eine inhaltliche Spezifizierung. Somit ist z.B. *Apfel* **hyponym** zu *Frucht*, weil die Bedeutung von *Apfel* spezifischer ist als die von *Frucht*. Umgekehrt ist *Frucht* ein **Hyperonym** von *Apfel*. Somit lassen sich mittels dieser semantischen Relationen hierarchieähnliche Gliederungen von Ausdrücken aufbauen [Buß02].

Meronymie und Holonymie

Die Meronymie ist eine weitere hierarchiebildende semantische Relation, die eine **Teil-Ganzes-Beziehung** ausdrückt. Hierbei ist ein Ausdruck ein **Meronym** eines anderen, falls er Teil des anderen Ausdrucks ist. Ein Beispiel hierfür wäre, dass *Tür* ein Meronym von *Haus* ist. Umgekehrt ist hierbei *Haus* ein **Holonym** von *Tür*, weil es die Entität darstellt, welche die *Tür* beinhaltet. Im Allgemeinen sind diese Relationen asymmetrisch und nicht transitiv (Das *Haus* hat eine *Tür*. Die *Tür* hat einen *Griff*. Aber das *Haus* hat keinen *Griff*) [Buß02].

2.2. Kontext in der Sprachwissenschaft

In der Sprachwissenschaft existieren verschiedene Definitionen des Kontextbegriffs, welche je nach Untersuchungsfeld einen leicht anderen Fokus setzen. Die für diese Arbeit passendste Definition stammt aus Bußmanns *Lexikon der Sprachwissenschaft* [Buß02] und definiert den Begriff wie folgt:

Als umfassender Begriff der Kommunikationstheorie bezeichnet Kontext alle Elemente einer Kommunikationssituation, die systematisch die Produktion und das Verständnis einer Äußerung bestimmen. [...] Erst im Zusammenwirken mit allen K.-Faktoren ergibt sich aus dem Sinnpotential eines Satzes der aktuelle Sinn einer Äußerung bzw. aus einer Folge von Äußerungen ein Text.

Hierbei unterscheidet Bußmann zwischen drei elementaren Typen von Kontext: Dem allgemeinen Kontext der Sprechsituation, wie z.B. dem Ort, der Zeit und dem Handlungszusammenhang der Äußerung, dem **persönlichen** und **sozialen Kontext**, welcher die Beziehung zwischen den Gesprächspartnern, ihre Einstellungen und Interessen sowie ihr Wissen bzw. ihre wechselseitigen Wissensannahmen beinhaltet und dem **sprachlichen Kontext**, der die Ausdrücke grammatikalisch und semantisch verknüpft. Die beiden Ersteren werden auch zusammenfassend als **situativer Kontext** bezeichnet und beziehen sich auf alle Zusammenhänge, die über die Beziehungen gegeben durch die Syntax und Semantik hinausgehen. Letzterer lässt sich wiederum in den satzinternen Kontext, also die syntaktischen Strukturen mittels denen der Ausdruck in den Satz eingebettet ist und dem Text bzw. Diskurs in dem die Äußerung getroffen wird, dem sogenannten **Kotext** unterscheiden [Buß02].

2.2.1. Kotext

Kotext bezeichnet die Teile des sprachlichen Kontextes, die über die Zusammenhänge im aktuellen Satz hinausgehen. Also alle Zusammenhänge, die sich auf die Teile des Diskurses beziehen, die vor dem aktuellen Satz geäußert wurden. Dies sind also alle Informationen die aus den satzinternen Kontexten der vorangegangenen Sätze ableitbar sind sowie Beziehungen zwischen diesen abgeleiteten Informationen. Somit ist der Kotext gerade bei der Auflösung von Referenzen von großer Bedeutung, denn ohne diese Informationen ist das Erkennen von Ausdrücken, die referentielle Kohärenz stiften, nicht möglich [Fra96].

Tabelle 2.1.: Auszug aus dem Penn-Tagset

Markierung	Bedeutung
NNP	Eigennamen, Singular
VB	Verb, Infinitiv
TO	Wort <i>to</i>
DT	Determinativ
JJ	Adjective
NN	Substantiv, Singular
CC	Koordinierende Konjunktion
PRP	Personalpronomen

2.2.2. Kontextualisierung

Kontextualisierung bezeichnet die Hypothese, dass Kontexte nicht immer gegeben sind sondern im Verlauf einer verbalen Konversation von den Sprechern aufgebaut werden können. Hierbei werden Kontextualisierungs-Hinweise (engl. *contextualization cues*) mit dem vorhandenen Hintergrundwissen in Verbindung gebracht. Die Informationen aus dem Hintergrundwissen werden dabei auf die wahrscheinlichste Domäne eingeschränkt und als zusammenhängenden Rahmen (engl. *Frames*) repräsentiert. Die Bedeutung eines Hinweises hängt daraufhin also von der Interaktion mit weiteren Hinweisen, die auf denselben oder andere Rahmen verweisen, ab. Das Wahrnehmen verschiedener Hinweise sorgt dafür, dass auch dann eine Interpretation des Gesagten möglich ist, wenn nicht alle Hinweise erfasst wurden.

Es kann allerdings bei interkultureller Kommunikation zu Missverständnissen kommen, da die Rahmen soziokulturell gebunden sind [Buß02].

2.3. Natural Language Processing

Mit dem Begriff *Natural Language Processing* wird in der Forschung die computergestützte Verarbeitung und damit die Analyse, das Verstehen oder das Produzieren von natürlicher Sprache bezeichnet. Die Sprache kann hierbei durch Text oder aber als gesprochener Sprache betrachtet werden. Die Aufgaben in diesem Forschungsfeld sind vielfältig und reichen vom Übersetzen von einer Sprache in eine andere über den Aufbau und die Generierung von Wissen aus Sprachquellen bis hin zum Verstehen von Aussagen [All03]. Da nicht alle dieser Aspekte relevant für diese Arbeit sind wird nur ein kurzer Überblick über die relevanten Aufgaben gegeben.

2.3.1. Wortart-Markierung

Mit Wortart-Markierung wird die Aufgabe der Zuordnung von Wortarten zu jedem Einzelwort eines Textes bezeichnet. Diese Aufgabe, im Englischen mit *Part-Of-Speech-Tagging* bezeichnet, ist essentiell für ein grundlegendes Verständnis der Zusammenhänge der Wörter eines Satzes. Viele Zusammenhänge lassen sich bereits direkt aus den Wortarten ableiten. Um eine einheitliche Darstellung der Wortarten zu ermöglichen, haben sich verschiedene Sätze von Markierungen, sogenannte *Tagsets*, herausgebildet. Das meist verwendete dieser Tagsets ist das Penn-Tagset [MMS93], welches auch in dieser Arbeit verwendet wird.

In Beispiel 3 ist der Satz „John go to the green cupboard and open it“ mit den entsprechenden Penn-Tagset Wortart Markierungen versehen, welche tiefer gestellt den Wörtern vorangehen.

- (3) [NNP John] [VB go] [TO to] [DT the] [JJ green] [NN cupboard] [CC and] [VB open]
[PRP it]

Die Bedeutungen der einzelnen Markierungen sind in Tabelle 2.1 erläutert. Hierbei lässt sich z.B. bereits aus der Wortart-Markierung *NNP* von „John“ ablesen, dass es sich hierbei um einen Eigennamen im Singular handelt, was eine nützliche Information für ein Verständnis der beschriebenen Gegebenheiten ist. Eine vollständige Übersicht über die enthaltenen Markierungen und ihre jeweiligen Bedeutungen ist in Anhang A.1 gegeben.

2.3.2. Phrasenerkennung

Die Phrasenerkennung (engl. *Chunking*) beschäftigt sich mit der Erkennung von neben geordneten sich nicht überlappenden Teilstrukturen eines Satzes, welche daraufhin als Phrasen markiert werden. Hierbei werden Phrasen, wie z.B. Nominalphrasen oder Verbalphrasen, unterschieden, wodurch die Phrasenerkennung weitere syntaktische Zusammenhänge eines Satzes abbildet und damit genauere Informationen über die zusammenhängenden Wörter eines Satzes bietet [TB00].

- (4) [NP John] [VP go] [PP to] [NP the green cupboard] and [VP open] [NP it]

In Beispiel 4 ist der Satz aus dem vorherigen Beispiel mit dem Ergebnis der Phrasenerkennung versehen. Hierbei stellen die tiefgestellten Zeichen den jeweiligen Phrasentyp dar. Eine Auflistung der existierenden Markierungen mit ihren zugehörigen Phrasentypen findet sich in Tabelle 2.2. Häufig werden die Informationen der Phrasenerkennung pro Wort gespeichert. Um dabei trotzdem Zusammengehörigkeiten zwischen verschiedenen Wörtern abbilden zu können, hat sich das *IOBES*-Format als Standard herausgebildet. Hierbei werden die pro Wort Informationen zusätzlich um die Angabe erweitert, ob sich das Wort am Anfang (*IOBES*-Tag: B), am Ende (*IOBES*-Tag: E), innerhalb (*IOBES*-Tag: I) oder außerhalb (*IOBES*-Tag: O) der angegebenen Phrase befindet. Zusätzlich gibt es den *IOBES*-Tag S, welcher bei einer Phrase die nur aus diesem Wort besteht eingesetzt wird.

Tabelle 2.2.: Phrasen-Markierungen

Markierung	Bedeutung
NP	Nominalphrase
VP	Verbalphrase
PP	Präpositionalphrase
ADVP	Adverbialphrase
ADJP	Adjektivphrase
SBAR	Phrase eingeleitet von einer unterordnenden Konjunktion
PRT	Partikel
INTJ	Einwurf

- (5) [S-NP John] [S-VP go] [S-PP to] [B-NP the] [I-NP green] [E-NP cupboard] [O and] [S-VP open] [S-NP it]

In Beispiel 5 ist der Satz mit den Informationen im *IOBES*-Format versehen.

2.3.3. Eigennamenerkennung

Bei der Eigennamenerkennung (engl. *Named Entity Recognition* oder kurz *NER*) wird der übergebene Text auf Namen von Personen, Organisationen oder Orte untersucht und die gefundenen Namen in eine dieser Kategorien eingeordnet. Hierzu werden sowohl regel- und lexikonbasierte Verfahren als auch maschinelle Lernverfahren eingesetzt. [TD03]

- (6) [PER Wolf], currently a journalist in [LOC Argentina], played with [PER Del Bosque] in the final years of the seventies in [ORG Real Madrid].

Tabelle 2.3.: Eigennamen-Markierungen

Markierung	Bedeutung
PER	Person
LOC	Ort
ORG	Organisation

In Beispiel 6 ist der Satz „Wolf, currently a journalist in Argentina, played with Del Bosque in the final years of the seventies in Real Madrid.“ mit den entsprechenden Eigennamenmarkierungen versehen. Die Übersetzung der Markierung zur eigentlichen Bedeutung lässt sich in Tabelle 2.3 ablesen.

2.3.4. Markieren semantischer Rollen

Eine weitere wichtige Aufgabe für das Verständnis von natürlicher Sprache stellt das Verstehen von Ereignissen und den teilnehmenden Entitäten dar. Dies bedeutet herauszufinden *Wer*, *Was*, zu *Wem*, *Wann* und *Wo* getan hat. Da diese Zusammenhänge in einem Satz durch ganz unterschiedliche Konstruktionen abgebildet werden können benötigt man eine Abstraktion dieser Konstruktionen.

Die in der Forschung verwendete Repräsentation dieser Zusammenhänge sind *semantische Rollen*, welche die abstrakten Rollen der Argumente eines Prädikats, die sie innerhalb eines Ereignisses einnehmen können, darstellen. Hierbei lassen sich unterschiedliche Abstraktionsgrade darstellen: Von einer sehr spezifischen semantischen Rolle wie *BUYER*, dem Käufer, über eine abstrakte Rolle wie *AGENT*, also dem Handelnden, bis hin zu einer super-abstrakten Darstellung wie *PROTO-AGENT*. Unter einem *PROTO-AGENT* werden alle Rollen zusammengefasst, die mehr *PROTO-AGENT* Eigenschaften, also Eigenschaften, die einem Handelnden zuzuordnen sind, besitzen als *PROTO-PATIENT* Eigenschaften (siehe [Dow91]). Die Beziehungen, welche diese Rollen einem Prädikat und dessen Argumenten zuordnen, variieren je nach gewünschtem Abstraktionsgrad und werden daher in unterschiedlichen Datenbanken wie FrameNet oder PropBank (siehe Abschnitt 2.5) gespeichert und gepflegt [JM09].

Die Aufgabe dieses Schrittes des *Natural Language Processing* ist es also den Argumenten eines Prädikats ihre semantischen Rollen zuzuweisen (im Englischen mit *Semantic Role Labeling* oder kurz *SRL* bezeichnet).

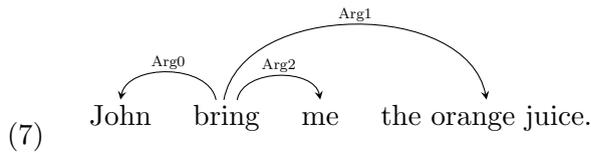


Tabelle 2.4.: Argumentübersetzung nach PropBank Eintrag zu Prädikat „bring“

Argument	spezifische Rolle (PropBank)	abstrakte Rolle (VerbNet ¹)
Arg0	Überbringer	AGENT
Arg1	überbrachter Gegenstand	THEME
Arg2	Ziel der Überbringung	DESTINATION

In Beispiel 7 ist der Satz „John bring me the orange juice“ mit einer solchen Zuordnung versehen dargestellt. Die Übersetzung der Abstrakten Argumentnummern auf die eigentlichen Rollen hängt dabei vom betrachteten Prädikat und den genutzten Datenbanken ab. In Tabelle 2.4 sind die Rollenzuordnungen aus PropBank und VerbNet abzulesen.

Um diese Zuweisung computergestützt durchführen zu können benötigt man zunächst die Informationen über die Wortarten sowie aus der Syntaxanalyse. Mithilfe dieser Informationen lassen sich dann zumeist mittels maschineller Lernverfahren Werkzeuge erstellen, die diese Aufgabe umsetzen. Da es sich hierbei allerdings um eine komplexe Aufgabe handelt, erreichen auch die besten Werkzeuge nur eine F1-Wert von um die 78% auf den als Referenz genutzten Daten der CoNLL-2005 Konferenz (engl. Conference on Natural Language Learning) bei welcher die Zuordnung semantischer Rollen die Aufgabe war [cona].

2.4. Korreferenzanalyse

Um die Probleme bei der Auflösung von natürlichsprachlichen Referenzen genauer betrachten und die Aussagen im Verlauf dieser Arbeit verständlicher darstellen zu können, wird eine kurze Einführung in die linguistischen Hintergründe zum Thema Korreferenz geboten, sowie ein kurzer Einblick in den aktuellen Stand der rechnergestützten Korreferenzanalyse gegeben. Die Informationen in diesem Abschnitt basieren größtenteils auf den Ausführungen von Daniel Jurafsky und James H. Martin in Kapitel 21 ihres Buches *Speech and Language Processing* [JM09].

2.4.1. Natürlichsprachliche Referenzen

In der Sprachwissenschaft bezeichnet man natürlichsprachliche Ausdrücke als Referenz, wenn diese auf eine Entität verweisen. Unterschieden wird hierbei zwischen dem linguistischen Ausdruck, der die Referenz ausführt, im Folgenden **referierender Ausdruck** genannt, und der Entität auf die die Referenz zeigt, der **Referent**.

(8) John, please go to the door and open it.

(9) When he arrived home, John went to sleep.

So stellt der Ausdruck „it“ in Beispiel 8 den referierender Ausdruck dar, welcher auf den Referenten *door* verweist. Aber auch der Ausdruck „door“ an sich ist eine referierender Ausdruck, denn dies ist die lexikalische Repräsentation der Entität *door*. Zwei referierende Ausdrücke werden als **korreferierend** bezeichnet, wenn sie die gleiche Entität als

¹Für Übersetzungen der thematischen Rolle s. Tabelle 2.5

Referenten besitzen. Somit sind die beiden Ausdrücke „door“ und „it“ korreferierend. Im Allgemeinen bezeichnet man Referenzen auf Entitäten, welche bereits zuvor in den Diskurs eingeführt wurden, als **Anaphern**. Außerdem besteht die Möglichkeit sich mittels eines referierenden Ausdrucks auf einen Referenten zu beziehen, der erst nach diesem Ausdruck in den Diskurs eingeführt wird, wie es beim referierenden Ausdruck *he* und dem Referenten *John* in Beispiel 9 der Fall ist. Eine solche Referenz wird als **Katapher** bezeichnet.

Die Möglichkeiten, welche die natürliche Sprache an Referenzen bietet, sind sehr vielfältig. Um einen Überblick über die grundlegenden Phänomene in der englischen Sprache zu geben betrachten wir die von Jurafsky und Martin hierbei unterschiedenen fünf Arten von referierenden Ausdrücken:

Unbestimmte Nominalphrasen

Als unbestimmte Nominalphrasen werden Phrasen, also in sich abgeschlossene syntaktische Einheiten, bezeichnet, die eine bisher nicht erwähnte Entität in den Diskurs einführen. Also eine Referenz auf eine neue Entität darstellen. Dies sind zumeist Nominalphrasen mit dem unbestimmten Artikel *a* (oder *an*), wie z.B. „*a cup of water*“.

Bestimmte Nominalphrasen

Bestimmte Referenzen hingegen beziehen sich auf eine Entität, die dem Diskurs-Teilnehmer bereits bekannt ist. Dies kann über eine bereits erfolgte Erwähnung der Entität im Diskurs erfolgt sein oder aber durch den Kontext als eindeutig und bekannt vorausgesetzt werden. Diese Referenzen können über bestimmte Nominalphrasen mit dem bestimmten Artikel *the* gebildet werden.

Pronomen

Eine weitere Form von bestimmten Referenzen sind Referenzen gebildet durch Pronomen, wie es in Beispiel 8 der Fall war. Hierbei können die Pronomen sowohl in einer Anapher verwendet werden, als auch Teil einer Katapher sein, also in einer Aussage auftreten in der sie vor dem Referenten erwähnt werden.

Ein Unterschied zwischen einer bestimmten Referenz beschrieben durch Nominalphrasen und derjenigen durch Pronomen ist, dass eine pronominale Referenz einen deutlich geringeren Bezugsbereich besitzt als die nominale Referenz. Denn Pronomen beziehen sich zumeist auf Entitäten, die in direkter Nähe beschrieben werden (zumeist nicht weiter als 1-2 Sätze entfernt) [JM09, S. 733]. Bestimmte Referenzen gebildet durch bestimmte Nominalphrasen hingegen können sich auch auf deutlich entferntere Referenten beziehen.

Demonstrativpronomen

Referenzen gebildet mit Demonstrativpronomen, wie *this* und *that*, unterscheiden sich von den Referenzen gebildet mit einfachen bestimmten Pronomen, wie im vorherigen Paragraph betrachtet. Sie können sowohl alleine eine Referenz darstellen, als auch als Artikel in einer Nominalphrase auftreten. Ihre lexikalische Bedeutung gibt außerdem einen Hinweis auf die wörtliche oder metaphorische Distanz zum Referenten. So bezieht sich *this* auf eine, räumlich oder zeitlich, nahe Entität, *that* hingegen auf eine weiter entfernte.

Namen

Namen stellen eine weitere häufige Form von referierenden Ausdrücken dar. Gemeint sind hiermit Namen von Personen, Organisation oder auch Orten. Sie können dabei sowohl auf neue, als auch auf bereits bekannte Entitäten verweisen und bieten somit sowohl die Möglichkeit bisher nicht erwähnte Entitäten in den Diskurs einzuführen, als auch auf bereits erwähnte Entitäten zu referieren.

Sonderformen

Zusätzlich zu den fünf genannten Phänomenen verweisen Jurafsky und Martin noch auf drei weitere Phänomene, welche die Auflösung von Referenzen verkomplizieren. Zum einen sind dies Referenzen gebildet über implizite Zusammenhänge, wie es bei *door* und *dishwasher* in Beispiel 10 der Fall ist

- (10) a. John, please go to the dishwasher and open it.
 b. Then close the door again.

Hierbei deutet die bestimmte Nominalphrase „the door“ darauf hin, dass eine *door* Entität bereits erwähnt wurde. Doch dadurch, dass aus einem Wissen über Zusammenhänge in der Welt bekannt ist, dass eine *dishwasher* Entität eine *door* besitzt lässt sich ableiten, dass sich „the door“ auf die Entität *dishwasher* bezieht.

Ein weiteres Phänomen sind die sogenannten **Generics**, welche sich nicht auf eine bestimmte Entität sondern die gesamte Klasse dieser Entitäten beziehen, wie es bei *they* in Beispiel 11 der Fall ist.

- (11) I'm interested in buying a Mac laptop. They are very stylish.
 (12) It was raining lava after Frodo had thrown the ring into the fires of Mount Doom.

Außerdem gibt es noch Formulierungen in denen referierende Ausdrücke keine Referenz erzeugen. Wie es z.B. bei *It* in Beispiel 12 auftritt.

2.4.2. Stanford Deterministic Coreference Resolution System

Das *Stanford Deterministic Coreference Resolution System* [LPC⁺11] ist eines der besten Korreferenzanalyse-Werkzeuge in der aktuellen Forschung. Es handelt sich hierbei um eine Sammlung von deterministischen Korreferenzmodellen die zusammenwirkend sowohl lexikalische, syntaktische, semantische als auch Diskursinformationen in die Analyse einfließen lassen. Außerdem bezieht das System dokumentweite Informationen wie das Geschlecht und Singular-/Plural-Informationen zu den gefundenen Erwähnungen mit ein. Das Werkzeug ist enthalten in der Werkzeugkette *Stanford CoreNLP* [sta]. Für das Erkennen von Korreferenzrelationen benötigt es einige Vorverarbeitungsschritte, die entweder durch die Werkzeuge des Stanford CoreNLP zur Verfügung gestellt werden können oder per Markierungen von anderen Werkzeugen übergeben werden können. Die benötigten Vorverarbeitungsschritte sind hierbei das Segmentieren des Textes in Wörter und Sätze (engl. *Tokenization*), die Markierung des Textes mit Wortarten(s. Abschnitt 2.3.1), die Lemmatisierung der Wörter, die Eigennamenerkennung(s. Abschnitt 2.3.3) und die Ergebnisse einer Syntaxanalyse. Der verwendete Ansatz setzt auf ein Mehrwege-Sieb aus deterministischen Sieben, welche unterschiedliche Genauigkeiten des Ergebnisses repräsentieren. Mit diesem Ansatz erreicht das Werkzeug auf den CoNLL-2011 [conb] Daten einen F1-Wert von 57.79%.

2.5. Lexikalische Datenbanken

Um die in Abschnitt 2.1.8 vorgestellten lexikalischen Beziehungen zwischen Wörtern in einer einheitlichen und frei zugänglichen Art und Weise zur Verfügung zu stellen existieren verschiedene lexikalische Datenbanken. Die verschiedenen Datenbanken decken unterschiedliche Teilbereiche der Beziehungen ab und versuchen hierbei ein möglichst umfassendes Angebot von Informationen zu diesem Bereich anzubieten.

2.5.1. WordNet

Eine in der englischsprachigen Rechnerlinguistik häufig verwendete lexikalische Datenbank ist *WordNet* [Fel98]. Sie enthält Informationen über semantische Relationen zwischen Wörtern und baut aus diesen Beziehungen ein lexikalisches Netz auf. Hierzu werden die Lemmata der betrachteten englischen Wörter unterschieden in Substantive, Verben, Adjektive und Adverbien und anhand ihrer Bedeutung in *Synsets* organisiert. Lemmata die sich im selben Synset befinden haben eine gleiche oder ähnliche Bedeutung, können aber ebenso in mehreren Synsets vorkommen. So ist das Wort *bass* sowohl im Synset [bass, bass voice, basso] und beschreibt die Bassstimme als auch im Synset [sea bass, bass] und beschreibt einen Fisch.

Außerdem bietet WordNet Informationen über weitere semantische Relationen. So werden Hyperonym- oder Hyponym-Beziehungen zwischen den Synsets mittels Relationen abgebildet, welches zu einer hierarchischen Ordnung der Synsets führt. Ebenso werden Teil-Ganzes Beziehungen (Meronymie) zwischen den Synsets dargestellt und somit eine weitere Ebene von Hierarchien eingefügt.

2.5.2. Proposition Bank

Die Proposition Bank [PGK05], zumeist kurz *PropBank* genannt, ist eine Datenbank bestehend aus zwei Teilen. Zum einen beinhaltet sie die Sätze aus dem Penn TreeBank Corpus [MMS93], mit semantischen Rollen versehen und zum anderen stellt sie ein Lexikon von Rahmen zur Verfügung, welche aus einer Menge von semantischen Rollen pro Prädikat bestehen (siehe Abschnitt 2.3.4). Das Rahmenlexikon von PropBank ordnet jedem Prädikat eine Rahmendatei (engl. *frame file*) zu, welche pro Bedeutung des Verbs eine Menge von semantischen Rollen spezifisch für diese Bedeutung zur Verfügung stellt. Hierbei werden den Argumenten Nummern anstatt Rollennamen zugeordnet um eine gewisse Generalisierung durchzuführen. So steht **Arg0** normalerweise für den *PROTO-AGENT* und **Arg1** für den *PROTO-PATIENT*, also den Handelnden bzw. Behandelten des Ereignisses. Zusätzlich dazu beinhalten die Rahmendateien aber auch eine spezifische Rollenbezeichnung. So enthält die Rahmendatei des Verbs „bring“ für die Bedeutung **bring.01**, welche der Bedeutung „carry along with, move literally or metaphorically“ entspricht, die in Beispiel 13 dargestellten Rollen.

(13) **bring.01**

Arg0: Bringer

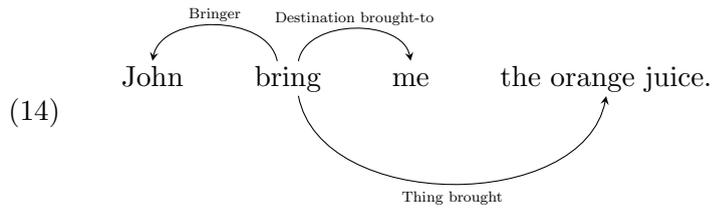
Arg1: Thing brought

Arg2: Benefactive or destination brought-for, brought-to

Arg3: Attribute, state after bringing, secondary action

Arg4: Ablative, brought-from

Das Wissen, welches in diesem Rahmenlexikon gespeichert ist, wird daraufhin für die Aufgabe des Markierens semantischer Rollen (s. Abschnitt 2.3.4) verwendet. Hierbei wird die Auflösung der Argumentnummern auf die konkreten Rollen anhand des erkannten Prädikats auf den Prädikat-bezogenen Rahmen und daraufhin anhand der Anzahl und der Ausprägung der erkannten Argumente auf die entsprechende Bedeutung durchgeführt.



In Beispiel 14 ist die Übersetzung der Argumentrollen aus Beispiel 7 anhand der Bedeutung **bring.01** dargestellt.

2.5.3. VerbNet

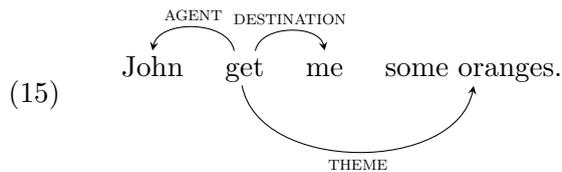
Bei *VerbNet* [KKRP06] handelt es sich um eine lexikalische Datenbank mit einer etwas anderen Ausrichtung als die Anderen. VerbNet fokussiert sich auf Informationen über Verben bzw. Verbklassen. Eine Verbklasse beinhaltet die Verben, welche eine gleiche Bedeutung und syntaktische Struktur aufweisen. Außerdem enthält sie thematische Rollen für die Prädikat-Argument Struktur dieser Verben, Einschränkungen bei der Auswahl der Argumente sowie Rahmen bestehend aus einer syntaktischen Beschreibung und semantischen Eigenschaften wie z.B. „in Bewegung“ basierend auf einer zeitlichen Funktion, die angibt zu welchem Zeitpunkt (Start, während, Ende) des Ereignisses die Eigenschaft gilt.

Da in dieser Arbeit hauptsächlich die thematischen Rollen aus VerbNet Verwendung finden werden diese näher betrachtet. Thematische Rollen sind hierbei Abstraktionen von semantischen Rollen, die gewisse Gemeinsamkeiten der spezifischen Rollen zusammenfassen. So existieren für die Verben *break* und *open* die spezifischen semantischen Rollen *Breaker* und *Opener*, aber beide Rollen haben gemeinsam, dass sie den Handelnden bzw. Verursachenden des Ereignisses beschreiben. Die thematische Rolle des Handelnden wird mit *AGENT* bezeichnet. Einen Überblick über die am häufigsten vorkommenden thematischen Rollen und ihre Bedeutung wird in Tabelle 2.5 gegeben.

VerbNet definiert 23 verschiedene thematische Rollen, welche zum Teil nur von bestimmten Verbklassen genutzt werden aber größtenteils über alle Klassen hinweg generalisieren. Eine Übersicht aller in VerbNet vorhandenen thematischen Rollen kann in Anhang A.2 nachvollzogen werden.

Tabelle 2.5.: Einige häufig verwendete thematische Rollen

Thematische Rolle	Definition
AGENT	Der willensmäßige Verursacher eines Ereignisses
THEME	Der Teilnehmer/Gegenstand, welcher direkt von dem Ereignis beeinflusst wird
BENEFICIARY	Der Teilnehmer, welcher direkt von dem Ereignis profitiert
EXPERIENCER	Der Teilnehmer dem etwas widerfährt oder dem etwas bewusst wird
SOURCE	Der Ausgangspunkt eines Gegenstands in einem Transfer
DESTINATION	Der Endpunkt eines Gegenstands in einem Transfer



Somit ergeben sich für das Beispiel „John get me some oranges.“ die thematischen Rollen, welche in Beispiel 15 dargestellt sind.

2.5.4. FrameNet

Bei *FrameNet* [BFL98] handelt es sich um eine weitere lexikalische Datenbank mit einer anderen Struktur und Ausrichtung als WordNet. Bei FrameNet werden die Lemmata der repräsentierten Wörter in lexikalischen Einheiten (*Lexical Units*) dargestellt. Eine lexikalische Einheit besteht aus einem Lemma und dem Rahmen (*Frame*) dem es zugeordnet ist. Ein Rahmen stellt hierbei „eine schematische Repräsentation einer Sammlung von charakteristischen Merkmalen, Attributen und Funktionen einer Situation und der charakteristischen Interaktionen mit Dingen typischerweise assoziiert mit ihr dar“ [All01, S. 251]. Jeder Rahmen besteht aus einer Anzahl von semantischen Rollen, die spezifisch für diesen Rahmen sind, die sogenannten Rahmenelemente (*frame elements*). Die verschiedenen Rahmenelemente werden in Kern- und Nicht-Kern-Rahmenelemente unterschieden, wobei Kern-Rahmenelemente alle Rahmenelemente sind die essentiell für die Bedeutung des Rahmens sind und Nicht-Kern-Rahmenelemente zumeist zusätzliche optionale Informationen bieten. So beinhaltet der Rahmen *Openness* die Kern-Rahmenelemente: *Barrier*, *Theme* und *UsefulLocation* und die Nicht-Kern-Rahmenelemente *Degree* und *Time* und ist wie folgt definiert:

A *UsefulLocation* is accessible to some (possibly generic) *Theme* despite a potential *Barrier*, or is not accessible because of the *Barrier*.

Außerdem enthält dieser Rahmen die lexikalischen Einheiten *closed.a*, *dark.a* und *open.a*. Das *.a bezeichnet, dass es sich hierbei um die Adverbien dieser Lemmata handelt. Zusätzlich zu der Zuordnung von Lemmata und Rahmenelementen zu Rahmen modelliert FrameNet auch Beziehungen zwischen Rahmen sowie Rahmenelementen. Hierbei können Rahmen unter anderem voneinander erben oder Generalisierungen von Rahmenelementen über Vererbung abgedeckt werden. Somit bietet FrameNet eine weitere Möglichkeit semantische Rollen für Prädikate zu ermitteln sowie darüber hinaus Informationen über inwiefern die enthaltenen Wörter normalerweise in Texten verwendet werden.

2.6. Ontologien

Mit dem Begriff *Ontologie* wird in der elektronischen Datenverarbeitung eine an den aus der philosophischen Lehre stammenden gleichnamigen Begriff, welcher die „Lehre vom Sein“ beschreibt, angelehnte Form der Wissensrepräsentation beschrieben. Die Verbindung zu dieser Lehre besteht darin, dass als Betrachtungsgegenstand konkret oder abstrakt existierenden Dinge, genannt Entitäten (s. Abschnitt 2.1.5), sowie ihre strukturellen und semantischen Beziehungen untereinander betrachtet werden.

Guarino et al. [GOS09] definieren Ontologien in der elektronischen Datenverarbeitung als ein Informationsobjekt, welches die Struktur eines Systems formal darstellt. Ein solches System ist hierbei zumeist eine begrenzte Domäne oder ein Teilsystem, dessen Wissen möglichst vollständig abgebildet wird. Der größte Vorzug von Ontologien gegenüber herkömmlichen Datenbanken ist die strukturierte zumeist hierarchische Darstellung des Wissens. In einer Ontologie werden einzelne Wissens-elemente (**Individuen**) unter einem

gemeinsamen Begriff (**Klasse**) zusammengefasst. Diese Begriffe können wiederum unter anderen Begriffen zusammengefasst werden, sodass eine hierarchische Struktur entsteht. Diesen Prozess bezeichnet man als Begriffsbildung (engl. *conceptualization*). Gruber definiert Ontologien über diesen Begriff wie folgt: „An ontology is an explicit specification of a shared conceptualization.“ (*Eine Ontologie ist die explizite Spezifikation gemeinsamer Begriffsbildung.*) [Gru93].

Zusätzlich zu den einzelnen Wissens-elementen und Begriffen (oder auch Konzepten) können in Ontologien auch Beziehungen (**Relationen**) zwischen den Elementen abgebildet werden, die über die hierarchische Struktur der Begriffsbildung hinausgehen. Somit sind die Kernelemente einer Ontologie Klassen und Individuen sowie Relationen.

Klassen bilden hierbei die Struktur der Ontologie ab und fassen gemeinsame Eigenschaften der ihnen untergeordneten Konzepte zusammen. Dadurch, dass sich Klassen wiederum anderen Klassen unter- und überordnen lassen, bilden sie die hierarchische Struktur, die sogenannte **Taxonomie** der Ontologie ab.

Individuen hingegen sind konkrete Ausprägungen (Instanzen) der Klassen und bilden somit die konkreten Wissens-elemente der Ontologie ab. Hierbei besteht die Möglichkeit, dass ein Individuum mehreren Klassen angehört.

Es gibt zwei verschiedene Arten von **Relationen** in Ontologien, die Beziehungen oder auch Eigenschaften von Elementen einer Ontologie abbilden. Zum einen gibt es Objekt-Relationen, die zwischen Individuen oder Klassen beschrieben werden und sowohl uni- als auch bidirektional sein können. Zum anderen Datentyp-Relationen, welche Beziehungen zwischen Klassen oder Individuen und einem Datentyp wie z.B. **Integer** abbilden und somit den Datentyp einer Eigenschaft darstellen.

3. PARSE: Programming ARchitecture for Spoken Explanations

Wie bereits in Kapitel 1 beschrieben, ist diese Arbeit Bestandteil des Projektes *PARSE*. Ziel des Projektes ist es, Anweisungen, gegeben in gesprochener natürlicher Sprache, in für ein Zielsystem interpretierbaren Quelltext umzuwandeln. Somit ist das Projekt im Feld der natürlichsprachlichen Programmierung (engl. *Natural Language Programming*) angesiedelt und beschäftigt sich mit den diversen Problemen, die bei dieser Abbildung entstehen. Dies beinhaltet die lexikalische, syntaktische und semantische Analyse der Anweisungen, um ein Verständnis des Gesagten zu erhalten sowie die Anbindung an verschiedene Wissensquellen zur korrekten Umsetzung in Quelltext.

Das Projekt beschränkt sich hierbei zunächst auf die englische Sprache. Ein Aspekt, der das Projekt von vielen anderen Projekten im Bereich der natürlich sprachlichen Programmierung abhebt, ist der explizite Fokus auf Anweisungen formuliert in gesprochener Sprache.

3.1. Besonderheiten der gesprochenen Sprache

Die Betrachtung von gesprochener natürlicher Sprache im Gegensatz zu geschriebener natürlicher Sprache sorgt für einige zusätzlich zu betrachtende Phänomene im Bereich der natürlichsprachlichen Programmierung: Zum einen formuliert der Mensch seine gesprochenen Aussagen anders als bei Texten die er verfasst. So treten in gesprochener Sprache vermehrt Ellipsen und Pronomen, sowie Phänomene wie Füllwörter oder Wortwiederholungen auf. Zum anderen kann bedingt durch die vorgeschaltete automatische Spracherkennung (kurz *ASR*) nicht davon ausgegangen werden, dass Satzgliederungseinheiten, wie Satzzeichen, in der Eingabe vorhanden sind und alle Wörter korrekt erkannt werden.

(16) Armar eh please go to the the dishwasher and burn it on

So könnte ein beispielhafter Eingabetext nach der Erkennung durch den ASR wie in Beispiel 16 das Füllwort „*eh*“, die Wortwiederholung „*the the*“ und den Wortfehler „*burn*“ statt *turn* enthalten.

Da allerdings viele Werkzeuge der Sprachverarbeitung auf geschriebener Sprache basieren, ist eine der Herausforderungen in diesem Projekt, die Besonderheiten der gesprochenen Sprache durch verschiedene Herangehensweisen, wie Satzteilerkennung und dem Anbieten von Wortalternativen, abzufangen und durch vermehrten Einsatz von Wissensquellen den Einsatz dieser Werkzeuge zu ermöglichen oder neue Lösungen für die Probleme zu finden.

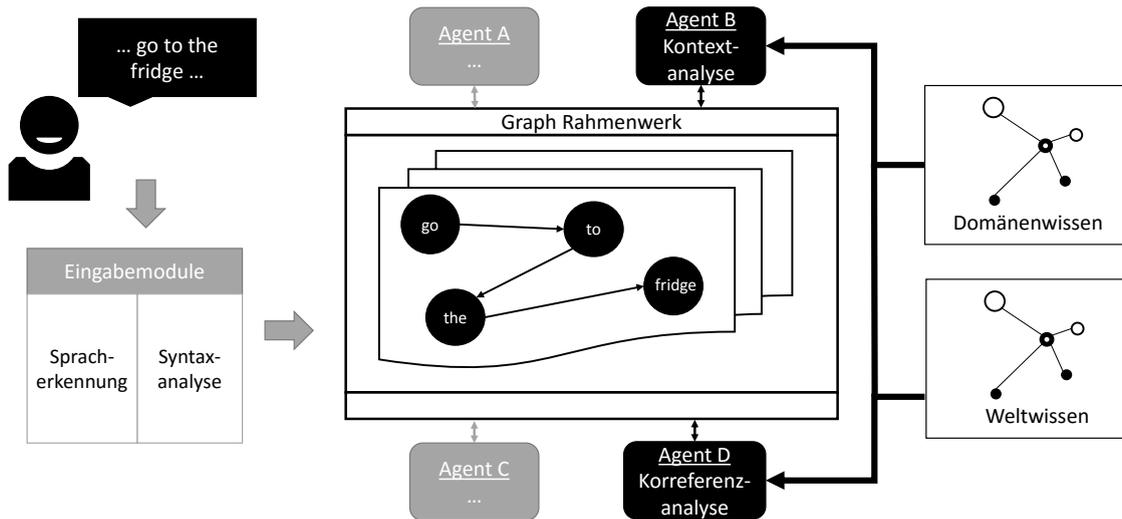


Abbildung 3.1.: Vereinfachte Projekt-Architektur (basierend auf [WT15])

3.2. Projektarchitektur

Um das Ziel des Projektes, in gesprochener Sprache formulierte Anweisungen in von einem Zielsystem verständlichen Quelltext umzuwandeln, zu realisieren, wird die Agentenbasierte Architektur *ProNat* (siehe Abbildung 3.1) eingesetzt. Diese ermöglicht es verschiedene Schritte der Semantikanalyse parallel durchzuführen. Vorgeschaltet sind die Eingabemodule, welche sich um die Umsetzung der gesprochenen Sprache in den als geteilten Datenspeicher genutzten Graph kümmern und variabel ausgetauscht werden können.

Die Agenten, welche die verschiedenen Schritte der Semantikanalyse durchführen, arbeiten daraufhin auf diesem Graph und lassen ihre Ergebnisse mittels Graphtransformationen transparent in den geteilten Datenspeicher einfließen. Die Semantikanalyse umfasst unter anderem Aufgaben wie das Auflösen von sprachlichen Mehrdeutigkeiten, das Erkennen von Kontrollstrukturen und Aktionen sowie die in dieser Arbeit durchgeführten Schritte der Kontext- und Korreferenzanalyse.

Außerdem sollen unterschiedliche Domänen als Zielsystem zulässig sein, weshalb das Domänenwissen, repräsentiert in Domänen-Ontologien, über eine einheitliche Schnittstelle eingebunden wird. Diese Architektur erlaubt es somit sowohl die Eingabequellen als auch die Zielsysteme auszutauschen und durch den Graph als einheitlichen Datenspeicher die Agenten flexibel und transparent zu gestalten.

3.3. Bisheriger Projektverlauf

Zu Beginn des *PARSE*-Projektes wurde untersucht, wie Anweisungen in natürlicher Sprache in der Domäne von Haushaltsrobotern wie ARMAR-III [ARA⁺06] gegeben werden. Hierzu wurde im Rahmen einer Bachelorarbeit [Gün15] ein Sprachkorpus von 66 Anweisungstexten für ARMAR-III, beschrieben von 22 unterschiedlichen Probanden, aufgebaut.

Daraufhin wurde eine logische Repräsentation für die als Eingabe erwartete natürliche Sprache gesucht. Das Ergebnis dieser Bachelorarbeit [Sch15] war der Aufbau einer Rahmenarchitektur für eine Repräsentation der natürlichen Sprache und aller weiteren im Projekt geteilten Informationen als Graph. Um die Überführung in diese Repräsentation zu ermöglichen wurden verschiedene ASR evaluiert und ein Analyse-Werkzeug erstellt, dass die gesprochene Eingabe mittels ASR in den Korpus einfließen lässt [Pas15].

Tabelle 3.1.: Die Struktur der Domänenontologie

Konzeptname	Beispiel	Beschreibung
Thing		Oberkonzept der Ontologie
System	Armar, Roboter	Repräsentation eines ansprechbaren Systems
Object	Fridge, Fridge.Door	auftretende Objekte
Method	open, take	ausführbare Methoden
State	opened, full	annehmbare Zustände
DataType	boolean, String	Datentypen
Parameter		Parameter der Methoden
Value		Werte der Parameter

Die eigentliche Abbildung des mittels der ASR erzeugten Textes auf die Graphrepräsentation wurde in einer weiteren Bachelorarbeit [Koc15] durchgeführt und um erste Ergebnisse von Rechnerlinguistik-Verarbeitungsschritten, wie der Wortart-Markierung, ergänzt.

Als nächster Schritt in der Entwicklung des Projektes wurden sowohl die Ergebnisse des ASR [Sch16] als auch die Ergebnisse der Wortart-Markierung [Kie16] optimiert sowie der erste Semantikanalyse-Agent implementiert. Dieser beschäftigt sich mit der Erkennung von Kontrollstrukturen in den gegebenen Anweisungen [Ste16].

3.4. Domänenontologie

Mittels der Domänenontologie werden im *PARSE*-Projekt die Gegebenheiten der Domäne repräsentiert. Sie beinhaltet Informationen über die in der Domäne auftretenden Objekte, die angesprochenen Systeme, deren Funktionen sowie Zustände und Zustandsübergänge. In Tabelle 3.1 ist die Struktur der Ontologie mit Beispielindividuen dargestellt. Außerdem enthält die Ontologie Informationen über Beziehungen zwischen den Individuen wie Teil-Ganzes-Beziehungen zwischen Objekten, mögliche Zustände die ein Objekt annehmen kann oder auch durch eine Methode ausgelöste Zustandsübergänge.

Die in der Domänenontologie enthaltenen Informationen werden durch eine einheitliche Schnittstelle zur Verfügung gestellt, sodass die dahinterstehende Ontologie und damit die Domäne jederzeit ausgetauscht werden kann. Dies ermöglicht es, das Zielsystem sowie das Einsatzgebiet variabel auszutauschen und die Quelltextabbildung unabhängig von einer speziellen Domäne durchführen zu können.

Andererseits können über die einheitliche Schnittstelle, die in der Ontologie beschriebenen Systeme, Objekte, Methoden, Parameter, Datentypen, Werte und Zustände ausgelesen sowie Beziehungen zwischen diesen Individuen ermittelt werden. Einen Überblick über die für diese Arbeit relevanten Teile der Schnittstelle wird in Abschnitt B des Anhangs gegeben.

4. Verwandte Arbeiten

Die Aufgabe Anweisungen in natürlicher Sprache in Quelltext umzuwandeln gewinnt in der aktuellen Forschung immer mehr an Relevanz. Hierbei haben sich zwei unterschiedliche Betätigungsfelder gebildet. Zum einen ist dies die Forschung zur Umwandlung von natürlicher Sprache in Programmstrukturen, wie Klassen, Instanzen, Methoden usw. wie es z.B. bei Metafor [LL05] gemacht wird. Zum anderen Arbeiten zur Abbildung von natürlicher Sprache auf bereits existierende Programmkonstrukte in einer gewissen Domäne. Also der Assoziation von Informationen gegeben durch natürlich sprachlichen Text mit bereits existenten Programmbausteinen.

Da letzterer Bereich dem Ansatz, welcher im *PARSE*-Projekt verfolgt wird, näher kommt, beschränkt sich die Betrachtung der verwandten Arbeiten aus dem Bereich des *Programmieren in natürlicher Sprache* auf dieses Feld. Insbesondere liegt bei der Analyse der Arbeiten der Fokus auf Ansätzen, die Kontextwissen in ihre Betrachtung einbeziehen oder explizit Kontextwissen aufbauen, wie es auch Zielsetzung in dieser Arbeit ist.

Aber auch Arbeiten aus anderen Bereichen können interessante Ansätze für das Vorgehen in dieser Arbeit liefern. So sind einige Arbeiten aus den Bereichen der Wissensrepräsentation und Informationsextraktion zumindest in Teilaspekten relevant für diese Arbeit. Weiterhin sind Arbeiten interessant, die Kontextwissen und Domänenwissen für die Korreferenzanalyse nutzen, was ebenso ein Ansatz dieser Arbeit ist.

4.1. Programmieren in natürlicher Sprache

Misra et al. beschäftigen sich in ihrer Arbeit **Environment-Driven Lexicon Induction for High-Level Instructions** [MTLS15] mit der Aufgabe Anweisungen beschrieben durch Verben, die komplexe mehrschrittige Aktionen beschreiben, auf mögliche Einzelaktionen abzubilden. Die betrachtete Domäne ist in diesem Fall das Geben von Anweisungen an einen Roboter. Hierbei setzen sie auf das Einbinden von Wissen über die aktuelle Umgebung des Roboters, um die möglichen Aktionen für die Abbildung einzugrenzen und somit das Lernen von neuen Anweisungen zu erleichtern.

Der verwendete Algorithmus baut ein Lexikon auf, welches die Zuordnung von nicht bekannten Verben zu neuen Konzepten beinhaltet. Diese Zuordnungen werden durch Kontextwissen unterstützt. Der in dieser Arbeit verwendete Kontext bezieht sich auf Wissen über die Umgebung in der sich der Roboter befindet, also Wissen wie, dass es eine Mikrowelle in dem Raum gibt und dass das Erhitzen eines leeren Bechers in der Mikrowelle

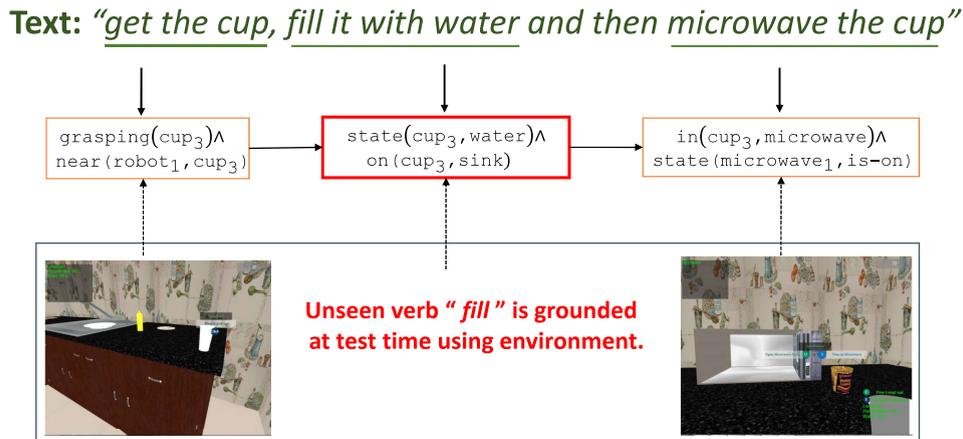


Abbildung 4.1.: Beispiel für den Einsatz von situativem Kontext zur Verb-Konzept Abbildung [MTLS15]

keinen Sinn ergibt. Daher kann der Suchraum für mögliche Einzelaktionen, welche die komplexe neue Aktion erfüllen, durch diese Informationen eingeschränkt werden und z.B. das Verb „fill“ in Abbildung 4.1 auf eine Aktion zurückgeführt werden, welche das Befüllen des Bechers beinhaltet.

Zur Repräsentation der Aktionen nutzen Misra et al. nicht spezifische textuell dargestellte Aktionen, sondern setzen auf Nachbedingungen. Außerdem nutzen sie Bedingungen ebenfalls um die Umgebung zu repräsentieren. Die Bedingungen bestehen aus Konjunktionen entweder gebildet aus räumlichen Beziehungen zwischen Objekten oder aus Zuständen der Objekte und können je nach Umgebung in der sie betrachtet werden als *wahr* oder *falsch* ausgewertet werden. Die Umgebung selbst wird als Graph dargestellt, welcher die vorhandenen Objekte als Knoten enthält und Relationen zwischen diesen Objekten über Kanten repräsentiert. Somit lässt sich durch Auswerten der Nachbedingungen und Interpretation der Zustände in der Umgebung eine effiziente Suche über alle infrage kommenden Aktionen durchführen.

Mit diesem Ansatz erreichen sie deutlich höhere Genauigkeiten bei der Zuordnung von mehrschrittigen Aktionen auf diese Aktionen repräsentierende Aktionssequenzen als vergleichbare Ansätze, die kein Kontextwissen in ihre Betrachtung mit einfließen lassen oder das Kontextwissen nur während der Lernphase des Werkzeugs miteinbeziehen. Außerdem haben sie gezeigt, dass gerade die Kombination des Benutzens des Lexikons sowohl beim Lernen als auch während der Testausführung die Genauigkeit der Zuordnung noch einmal erhöht. Aus diesen Ergebnissen folgern sie, dass die Benutzung von Umgebungswissen für die lexikalische Abdeckung von komplexen Aktionen ein vielversprechender Ansatz ist.

Die Arbeit **Intentional Context in Situated Natural Language Learning** [FR05] von Fleischmann und Roy setzt an einem noch früheren Punkt in der Betrachtung von Anweisungen in natürlicher Sprache an. Sie versuchen die Abbildung von in gesprochener natürlicher Sprache gegebenen Anweisungen auf Aktionen zu lernen. Hierzu betrachten sie zwei Spieler eines Computerspiels, wobei der eine Spieler als Experte fungiert und dem anderen Spieler Anweisungen gibt und der ausführende Spieler nichts Anderes macht, als die Anweisungen umzusetzen. Die gesprochene Sprache und die zugehörigen Aktionen werden hierbei aufgenommen und dann mittels einer Abbildung in Verbindung gesetzt.

Das Interessante bei ihrem Lernverfahren ist die Aufteilung in zwei Phasen. In der ersten Phase wird ein beabsichtigter Kontext für Aktionen aufgebaut. Hiermit ist der Aufbau einer Repräsentation der möglichen gemeinten Aktionen in einer gegebenen Situation als

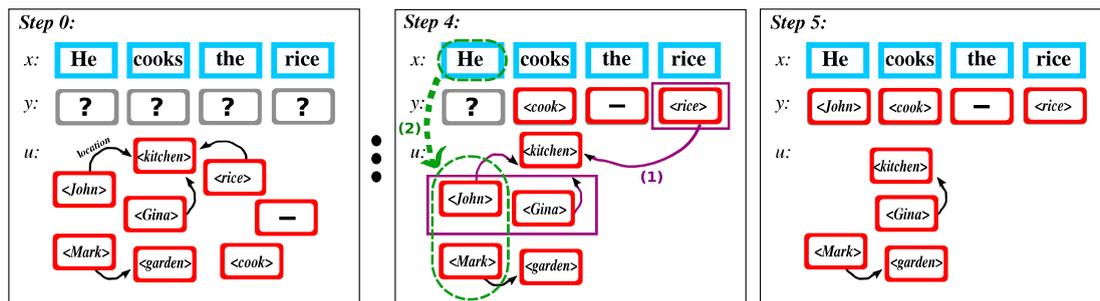


Abbildung 4.2.: Ablauf des Inferenzprozesses [BUCW10]

kontextfreie Grammatik gemeint. Der beabsichtigte Kontext ist somit eine Art von Weltwissen über die üblichen Aktionen und ihre Zusammensetzung aus anderen Aktionen in einer gegebenen Situation. Dieser wird bisher von Hand aufgebaut, soll aber in Zukunft ebenfalls automatisch gelernt werden. Mittels dieses beabsichtigten Kontextes wird dann in der zweiten Phase eine Abbildung der geäußerten Anweisungen auf die wahrscheinlichste Aktion und den daraus resultierenden Unteraktionen durchgeführt.

Es wird also eine Form von Kontext, genauer eine Form des situativen Kontextes, zur Verbesserung einer Semantikanalyse herangezogen. Allerdings wird keine genauere Aussage über das geplante Vorgehen zur Erstellung dieses Kontextes ohne manuelle Eingriffe gegeben. Die Evaluation des Ansatzes hat ergeben, dass der Aufbau einer Wissensrepräsentation generell aber insbesondere auch mit ableitbaren möglichen Intentionen für Aktionen dabei hilft die Semantik des Gesagten zu ermitteln und zu verstehen.

Die Arbeit **Towards Understanding Situated Natural Language** [BUCW10] von Bordes et al. beschäftigt sich mit der Aufgabe natürliche Sprache auf physikalische Entitäten abzubilden, dem sogenannten *language grounding*. Ihr Ansatz beruht dabei darauf, jedem Wort in einem Satz die Entität oder das abstrakte Konzept aus einem bekannten Domänenwissen zuzuordnen, auf welches es sich bezieht.

Hierbei beziehen sie Weltwissen und linguistische Informationen in die Betrachtung zur Abbildung mit ein. Sie nutzen eine Datenbank, welche Informationen über die existierenden physikalischen Entitäten und ihre Beziehungen zueinander enthält. Während des Lernens der Abbildung wird dieses Wissen durch einen Abgleich zwischen der natürlichen Sprache und den gegebenen Informationen in der Datenbank miteinbezogen, um das Ergebnis der Konzeptzuordnung zu verbessern. Um die Informationen in einem Satz für die Zuordnung nutzen zu können, verwenden sie einen Inferenzprozess. Dieser besteht darin, dass zunächst das Konzept zugeordnet wird, welches am wenigstens Mehrdeutigkeiten zulässt und diese Wahl sowie die daraus resultierenden ableitbaren Informationen bereits in das Zuordnen der anderen Konzepte im Satz miteinbezogen werden (vgl. Abbildung 4.2).

Auch diese Arbeit betrachtet also eine Form des situativen Kontextes, repräsentiert durch die von Hand erzeugte Datenbank. Doch durch die Verwendung des Inferenzprozesses und dem Miteinbeziehen von linguistischen Informationen, wie z.B. Geschlechtsmerkmalen, werden auch Informationen aus dem aktuellen Satz und dem sprachlichen Kontext für die Abbildung verwendet. Außerdem zeigt ihr Vorgehen, dass durch gewisse Kontextinformationen, vor allem Informationen über Beziehungen zwischen Entitäten sowie die Auflösung von Mehrdeutigkeiten leichter durchgeführt werden kann. Denn in ihrer Evaluation erzielten sie mit Kontextwissen deutlich niedrigere Fehlerraten in der Zuordnung. Das zur Verfügung gestellte Kontextwissen haben sie so gewählt, dass sich alle Mehrdeutigkeiten auch damit auflösen lassen. In einer Versuchsreihe ohne vollständige Kontextin-

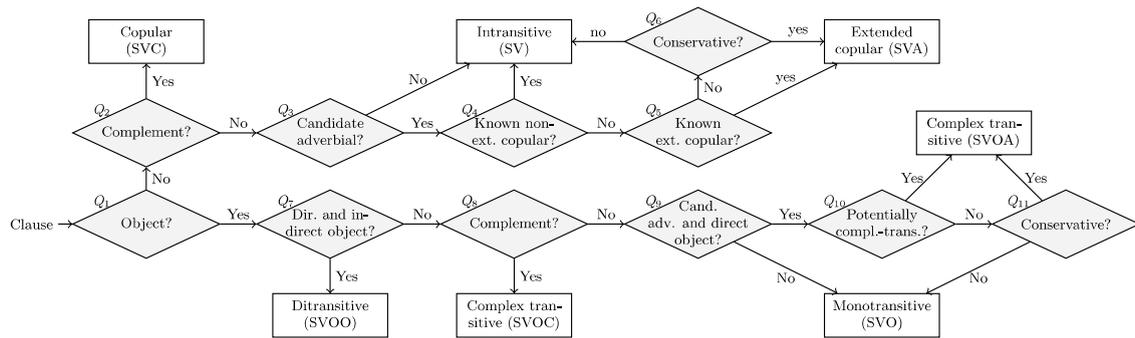


Abbildung 4.3.: Verbtyp und Gliedsatztyp-Erkennung [DG13]

formationen lagen die Fehlerraten zwischen den Ergebnissen erzielt mit vollständigem und komplett ohne Kontextwissen.

4.2. Wissens- und Informationsextraktion

Die Arbeit **Knowledge Extraction Based on Discourse Representation Theory and Linguistic Frames** [PDG12] von Presutti et al. beschäftigt sich mit der Extraktion von Wissen aus natürlicher Sprache. Ziel ist es aus in natürlicher Sprache gegebenen Aussagen eine Ontologie aufzubauen, welche in einem Text enthaltene Informationen repräsentiert.

Interessant hierbei ist der Weg, den sie bei der Erstellung der Ontologie wählen. Denn um eine langwierige Lernphase des Werkzeugs zu umgehen versuchen sie die linguistischen Rahmen zu erkennen. Hierzu werden die Aussagen in natürlicher Sprache zunächst in eine der Diskursrepräsentationstheorie (vgl. Abschnitt 2.1.6) entsprechende logische Form umgewandelt. Daraufhin nutzen sie eine Erweiterung des zum Erstellen dieser Repräsentation genutzten Werkzeugs Boxer [Bos08] um gleichzeitig die enthaltenen linguistischen Rahmen zu erkennen. Hierdurch kann direkt eine Anbindung an FrameNet (vgl. Abschnitt 2.5.4) durchgeführt werden, woraus Informationen über die Beziehungen der Entitäten und Aktionen im bearbeiteten natürlich sprachlichen Text gewonnen werden können.

Die Evaluation des Ansatzes hat ergeben, dass die Benutzung von Boxer als Rahmenerkennung zwar geringere Genauigkeit als spezialisierte Rahmenerkennungswerkzeuge wie SEMAFOR [CSDS10] bietet, aber die Ausführungszeit um den Faktor zehn schneller ist. Außerdem ist Boxer deutlich flexibler als SEMAFOR, da es nicht auf einem Korpus und für eine bestimmte Domäne gelernt werden muss. Außerdem konnten Presutti et al. zeigen, dass die aus den Informationen erstellten Ontologien stark zusammenhängend und damit geeignet für die schnelle Extraktion von Informationen aus Texten sind.

Ebenfalls aus dem Bereich der Informationsextraktion, insbesondere der *Open Information Extraction*, welche sich mit der Extraktion von nicht spezifiziertem relationalem Wissen aus unstrukturierten Textquellen beschäftigt, stammt die Arbeit **ClauseIE: Clause-Based Open Information Extraction** [DG13] von Del Corro und Gemulla. Ihr Ansatz besteht darin zunächst die im Satz enthaltenen Gliedsätze zu betrachten und sie in eine von sieben Kategorien einzuordnen (vgl. Abbildung 4.3). Ein Gliedsatz ist dabei eine Gruppe von Wörtern eines Satzes, welche durch ein Verb miteinander in Beziehung stehen, also zumeist eine Kombination aus Subjekt, Verb und als Argument fungierende Objekte. Sie erkennen die Gliedsätze über Muster im von ihnen verwendeten Abhängigkeitsgraph des Stanford Dependency Parsers [KM03]. Aus den kategorisierten Gliedsätzen werden daraufhin je nach Domäne verschiedene Vorschläge für mögliche ableitbare Wissensbeziehungen gefolgert.

Das Interessante an dem hier verwendeten Ansatz ist das verwendete Vorgehen zur Erkennung der Gliedsätze und deren Kategorisierung. Die Informationen über die Struktur der Gliedsätze kann dabei helfen Zusammenhänge zwischen Verben und Entitäten zu erkennen und somit Beziehungen zwischen ihnen abzuleiten.

Allerdings ist das verwendete Vorgehen stark abhängig von der Güte des Abhängigkeitsgraphen, welche z.B. bei Aussagen in gesprochener Sprache und damit ohne Satzzeichen deutlich schlechter ausfällt als bei geschriebenem Text. Mit diesem Ansatz erreichen Del Corro und Gemulla im Vergleich zu anderen Open-Information-Extraction-Werkzeugen eine höhere Präzision und sind flexibler durch die anpassbare Generierung von Vorschlägen.

4.3. Korreferenzanalyse unterstützt durch Kontextwissen

Gerade für die Korreferenzanalyse spielt das Kontextwissen eine entscheidende Rolle, denn ohne Wissen über zuletzt genannte Entitäten lassen sich keine Referenzen zuverlässig auflösen. Daher beschäftigen sich einige Arbeiten in diesem Bereich mit dem expliziten Einbinden von Kontextwissen in die Korreferenzanalyse. Ein erster früherer Ansatz zu Korreferenzanalyse, welchen Allen in seinem Buch *Natural Language Understanding* [All95] einführt, stützt sich bereits explizit auf Kontextwissen. Er benutzt hierzu sogenannte *History Lists*, eine simple Datenstruktur, welche die Diskursentitäten der letzten teilsatzlokalen Kontexte beinhaltet. Hierbei stehen die zuletzt hinzugefügten Entitäten als Erstes (zumeist die Entitäten aus dem gegenwärtig betrachteten Satz) und Entitäten aus weiter entfernten Kontexten später in der Liste.

Somit stellen die *History Lists* eine Art mit der Entfernung gewichtetes Gedächtnis der auftretenden Entitäten dar. Allen nutzt dieses Gedächtnis daraufhin, um eine simple Auflösung von Referenzen anhand der Entfernung und einiger Korreferenzeinschränkungen durchzuführen.

Dieser Ansatz basiert auf einem simplen Vorgehen und bietet in komplexeren Kontexten mit vielen infrage kommenden Entitäten keine ausreichende Lösung. Dies liegt daran, dass nur die Entfernung als Maß betrachtet wird und somit immer die nächstgelegene auf die Referenz passende Entität als Referent gewählt wird. Dieses Vorgehen funktioniert somit nur in Kontexten in denen wenige Entitäten auftreten, die als Referent für einen referierenden Ausdruck infrage kommen. Allerdings können die gewonnenen Informationen über zuletzt genannte Entitäten und ihre Entfernung gerade bei gesprochener Sprache trotzdem bei der Auflösung von Referenzen und der Extraktion von Beziehungen zwischen Entitäten und Aktionen helfen.

Die Arbeit **Using domain knowledge and domain-inspired discourse model for coreference resolution for clinical narratives** [JR13] von Jindal und Roth beschäftigt sich mit Korreferenzanalyse in der Domäne von Patientenberichten in Krankenhäusern. Da die gängigen Korreferenzanalyse-Werkzeuge auf Nachrichtentexten gelernt wurden, nutzen sie für eine Umsetzung in dieser Domäne einen kombinierten Ansatz aus umfassendem Domänenwissen und einem auf die Domäne eingeschränkten Diskursmodell. Hierbei setzen sie voraus, dass die Referenten im Dokument bereits erkannt und in eine von 5 Kategorien (Test, Treatment, Problem, Person und Pronoun) eingeordnet sind.

Das Domänenwissen benutzen sie für die Auflösung der Referenzen aus den Kategorien Test, Treatment und Problem, welches bedingt durch die vielen zusätzlichen medizinischen Fachbegriffe und deren Synonyme und Hyponyme in medizinischen Berichten nötig ist, um Begriffe mit möglicherweise gleicher Bedeutung identifizieren zu können. Das aufgebaute Domänenwissen besteht aus Informationen über äquivalente Bezeichnungen für verschiedene medizinische Fachbegriffe, Abbildungen von Abkürzungen auf ihre ausgeschriebene Form, sowie Wissen über Hyponymie-Beziehungen zwischen diesen Fachbegriffen. Mithilfe dieses Wissens bilden sie die im Text auftretenden Begriffe auf sogenannte semantische

Repräsentationen ab, welche eine abstrahierte Repräsentation der einzelnen Ausprägungen darstellen. Eine solche semantische Repräsentation besteht aus einer Anzahl von Modifizierern, die weitere Informationen über das Konzept geben, wie dem Körperteil auf das sich bezogen wird, einer genaueren anatomischen Positionsangabe wie z.B. „linker oberer“, der Äquivalenzklasse zu der das beschriebene Krankheitsbild passt und den restlichen Teilen der Beschreibung. Die Korreferenzanalyse dieser Begriffe wird daraufhin anhand eines Abgleichs dieser semantischen Repräsentationen durchgeführt.

Für die Korreferenzauflösung der Erwähnungen aus der Kategorie Person wird das eingeführte Diskursmodell dazu verwendet, die in der Domäne vorkommenden und somit als Referenten infrage kommenden Personen auf genau einen Patienten, mehrere Doktoren und einige Familienmitglieder einzuschränken. Außerdem gibt das Modell übliche Kontexte an, in welchen Referenzen auf bestimmte Gruppen von Personen vorkommen. Durch dieses Vorgehen können alle Erwähnungen von Personen in eine dieser Kategorien eingeordnet werden sowie die Korreferenzauflösung typischer Aussagen dieser Kategorien auf Personen dieser Kategorie eingeschränkt werden.

Die Evaluation des Ansatzes in der *2011 i2b2/VA coreference resolution challenge* [i2b] hat ergeben, dass das System zwar die Einreichung mit der höchsten Präzision von 91,7% war, aber dafür die Ausbeute um 10% niedriger lag als beim bestplatzierten Ansatz. Jindal und Roth führen diese Diskrepanz auf das konservative Verhalten ihres Ansatzes zurück und sehen die entwickelten Konzepte, wie das Diskursmodell, als weiteren Mehrwert ihrer Arbeit, welche auch auf andere Domänen übertragen werden können.

5. Analyse und Entwurf

Diese Arbeit ist Teil des Projektes *PARSE* (siehe Kapitel 3) und somit auch Teil der Umsetzung von in gesprochener natürlicher Sprache gegebenen Anweisungen in für ein Zielsystem interpretierbaren Quelltext. Eine der größten Herausforderungen bei der Umsetzung von natürlicher Sprache in Quelltext besteht darin, dass natürliche Sprache im Allgemeinen nicht eindeutig ist. Sie lässt Formulierungen zu, die für einen Rechner ohne gesondertes Wissen nicht eindeutig aufzulösen sind. Menschen nutzen zur Auflösung solcher nicht eindeutigen Formulierungen erlerntes Wissen über die strukturellen Gegebenheiten in der Sprache (syntaktische Zusammenhänge), Wissen über generelle semantischen Zusammenhänge, wie dass ein Kühlschrank eine Tür besitzt, sowie Wissen über die semantischen Zusammenhänge in denen die Äußerung getroffen wird, den sogenannten Kontext. Gerade der erste Punkt, das Wissen über syntaktische Zusammenhänge, wird häufig für das maschinelle Verständnis von natürlicher Sprache herangezogen, da diese Zusammenhänge einer gewissen Logik folgen und sich dadurch vom Rechner leichter erkennen lassen.

Betrachtet man nun aber gesprochene natürliche Sprache im Gegensatz zu geschriebener natürlicher Sprache entstehen weitere Schwierigkeiten bei der Deutung. So weist gesprochene Sprache zumeist weniger Struktur als geschriebene Sprache auf, denn in gesprochenen Kommunikationen wird selten durchgängig in vollständigen Sätzen gesprochen. Hieraus folgt, dass Äußerungen in gesprochener Sprache häufig fehlerhafte Grammatik oder unvollständige Syntax aufweisen. Zusätzlich dazu enthält gesprochene Sprache durch die ad-hoc Formulierung in geschriebener Sprache nicht auftretende Phänomene wie Hesitationswörter oder Wortwiederholungen. Diese Gegebenheiten beeinträchtigen die automatisierte Erkennung von syntaktischen Zusammenhängen immens, welche grammatikalisch und syntaktisch korrekte Äußerungen als Eingabe benötigt, was dazu führt, dass die Relevanz der beiden anderen Wissensaspekte bei der Auflösung deutlich steigt.

Gerade der Kontext bietet hierbei viel Potenzial, da er viele nützliche Informationen für die Deutung enthält. So bietet er die Möglichkeit Unklarheiten, wie unklare Bezüge durch Wortwiederholungen, aufzulösen oder Wortfehler durch nicht in den Kontext passende Formulierungen zu erkennen. Daher besteht der Ansatz dieser Arbeit darin den Kontext der in gesprochener Sprache gegebenen Anweisungen zu analysieren und in eine geeignete Repräsentation zu überführen. Um zu zeigen, dass der Kontext ein valides Hilfsmittel zur Umsetzung von gesprochener Sprache in Quelltext darstellt, wird außerdem das Auflösen von Referenzen als Anwendungsbeispiel eingesetzt, also eine kontextbasierte Korreferenzanalyse umgesetzt.

1. Armar, please go to the dishwasher and open it.
 2. Then go to the table, take the glass and put it into the dishwasher.
 3. After you have closed the door again, go to the fridge.
-

Abbildung 5.1.: Beispiel für natürlichsprachliche Referenzen

Die Korreferenzanalyse von geschriebener Sprache ist ein in der Computerlinguistik viel behandeltes Feld, welches wie bereits in Kapitel 1 motiviert ein essentieller Bestandteil eines Systems, welches Sprache in Quelltext übersetzen möchte, darstellt. Daher wäre ein plausibler Ansatz für die Umsetzung der Korreferenzanalyse für gesprochene Sprache, der Einsatz von existierenden Korreferenzanalyse Werkzeugen auf dem Text, welcher durch die Vorverarbeitungsschritte aus den gesprochenen Anweisungen erzeugt wird, und einer Verbesserung dieser Ergebnisse mit dem aufgebauten Kontextwissen. Doch gerade Korreferenzanalyse-Werkzeuge sind anfällig für die Gegebenheiten gesprochener natürlicher Sprache, denn sie setzen auf den Ergebnissen einer Syntaxanalyse auf, welche ohne gegebene Satzzeichen nur bedingt korrekte Ergebnisse liefert. Ebenso beeinflussen die Wortfehler und Wortwiederholungen die Ergebnisse einer Syntaxanalyse stark. Daher sind die Korreferenzanalyse-Ergebnisse dieser Werkzeuge auf dem *PARSE*-Korpus in vielen Fällen inkorrekt.

Um die Aufgabe der Korreferenzanalyse genauer erläutern zu können wird das Beispiel in Abbildung 5.1 herangezogen. In diesem Beispiel sind verschiedene Formen von Korreferenz dargestellt: Zum einen Anaphern gebildet durch das Pronomen *it* und zum anderen Referenzen gebildet durch Entitäten, die bereits in den Diskurs eingeführt wurden. So bezieht sich der referierende Ausdruck *dishwasher* im zweiten Satz auf die gleiche *dishwasher*-Entität wie der Ausdruck *dishwasher* im ersten Satz. Eine weitere komplexere Referenz bildet der referierende Ausdruck *the door* im dritten Satz, welcher nur durch ein Wissen über den Kontext und die Zustände der Entitäten als auf den *dishwasher* beziehend identifiziert werden kann.

Die im Projekt betrachteten Äußerungen sind Anweisungen eines Nutzers an ein System (z.B. einen Haushaltsroboter), zum Zweck der Ausführung der geäußerten Anweisungen. Diese Anweisungen werden in gesprochener englischer Sprache geäußert und daraufhin durch mehrere Vorverarbeitungsschritte (vgl. Kapitel 3) in Text überführt, welcher in einem als geteilten Datenspeicher genutzten Graph gespeichert wird. Hierdurch ist die betrachtete Sprache auf Englisch eingeschränkt und es werden Formulierungen als Eingabe erwartet, die eine imperative Satzstruktur aufweisen.

5.1. Diskussion der Verwandten Arbeiten

Nun lassen sich die in Kapitel 4 identifizierten Ansätze der verwandten Arbeiten hinsichtlich der soeben definierten Aufgabe dieser Arbeit analysieren und daraufhin überprüfen, ob die in ihnen gewählten Ansätze auf das in dieser Arbeit behandelte Problem übertragbar sind.

Zunächst lässt sich festhalten, dass die Ansätze für das Nutzen von Kontextwissen für verschiedene Aufgaben der natürlichsprachlichen Programmierung (siehe Abschnitt 4.1) alle auf dem situativen Kontext und dabei von Hand gebauten Datenbasen beruhen. Dies ist somit in dieser Form nicht auf die automatisierte Generierung von Kontextwissen dieser Arbeit übertragbar. Allerdings lassen sich Teilaspekte der Arbeiten als Hinweise für

ein mögliches Vorgehen für die durchgeführte Kontextanalyse nutzen. So bietet die Arbeit **Environment-Driven Lexicon Induction for High-Level Instructions** [MTLS15] durch die Betrachtung von Nachbedingungen und der dadurch durchgeführten Eingrenzung von möglichen Aktionen den Hinweis, dass eine solche Betrachtung von Nachbedingungen oder zumindest ein Abbild von Zuständen auch für weitere Probleme, wie z.B. das Auflösen von Korreferenzen, interessant sein könnte.

Die Arbeit **Towards Understanding Situated Natural Language** [BUCW10] hingegen deutet darauf hin, dass ein umfassendes Kontextwissen bei einer Zuordnung von sprachlichen Beschreibungen auf Entitäten einer Wissensbasis relevant sein können. Dieser Ansatz validiert also die Idee, das Kontextwissen als Hilfsmittel für verschiedene Abbildungen der sprachlichen Beschreibungen zu nutzen. Außerdem wurde gezeigt, dass durch das Kontextwissen und einen kombinierten Einsatz von händisch erzeugtem situativen Kontext und sprachlichem Kontext, wie z.B. Geschlechtsinformationen, die Kandidaten solcher Abbildungen weiter eingegrenzt werden können. Eine Abbildung, die direkt aus diesem Ansatz in das Kontextwissen übertragbar wäre, ist eine Abbildung der sprachlichen Beschreibungen auf die in der Domänenontologie beschriebenen Konzepte.

Die Arbeiten des Bereichs Wissen- und Informationsextraktion (siehe Abschnitt 4.2) geben hingegen Hinweise auf weitere Datenquellen, die für eine umfassende Generierung des Kontextes relevant sein können. So bietet die Arbeit **Knowledge Extraction Based on Discourse Representation Theory and Linguistic Frames** [PDG12] einen Hinweis auf das automatisierte Einbinden von Informationen aus der lexikalischen Datenbank FrameNet (siehe Abschnitt 2.5.4). Die Autoren nutzen dabei aber einen für diese Arbeit zu umfassenden Ansatz über die Diskursrepräsentationstheorie, welches für eine einfache Erkennung von Aktionen und ihren semantischen Rollen einen zu großen Mehraufwand bedeuten würde.

Der Ansatz aus der Arbeit **ClausIE: Clause-Based Open Information Extraction** [DG13] basiert auf Abhängigkeitsgraphen, die wie bereits in der Einleitung von Kapitel 5 beschrieben durch die gesprochene natürliche Sprache und die dadurch nicht korrekten Syntaxbäume nicht zur Verfügung stehen.

Aus dem Bereich der mittels Kontextwissen unterstützten Korreferenzanalyse (siehe Abschnitt 4.3) bieten die **History Lists** [All95] einen ersten validen Ansatz auf dem die Korreferenzanalyse dieser Arbeit aufsetzen kann. Da es sich hierbei allerdings nur um eine simple Heuristik handelt, kann es sich bei dem hier hauptsächlich behandelten Aspekt des Abstands zum referierenden Ausdruck nur um eine der möglichen Bewertungen die für potenzielle Referenten-Kandidaten durchgeführt werden kann handeln.

Aus der Arbeit **Using domain knowledge and domain-inspired discourse model for coreference resolution for clinical narratives** [JR13] lässt sich vor allem schließen, dass der in dieser Arbeit gewählte Ansatz aus extensivem Kontextwissen, gebildet mit Welt- und Domänenwissen, eine valide Möglichkeit zur Verbesserung einer Korreferenzanalyse darstellt und eine Einteilung von referierenden Ausdrücken in unterschiedliche unabhängig voneinander zu betrachtende Kategorien sinnvoll sein kann. Die Autoren benutzen ausschließlich Wissen über semantische Relationen zwischen den in ihrem Umfeld relevanten Fachbegriffen als Kontextwissen, welches vielleicht im Rahmen von Patientenberichten ausreicht um viele Mehrdeutigkeiten aufzulösen, aber für eine möglichst umfassende Ersetzung des Kontextwissens im Fall einer offenen Domäne wie in dieser Arbeit nicht ausreichend genug sein kann.

5.2. Kontextanalyse

Um die Aufgabe des Erstellens einer Repräsentation des Kontextes durchführen zu können, muss zunächst analysiert werden, was Kontext unter den gegebenen Bedingungen bedeutet,

welche Informationen gewonnen werden können, wie diese geeignet repräsentiert werden können und welche Möglichkeiten zur Gewinnung dieser Informationen bestehen.

Zunächst ist hierbei zu klären, welche Kontextinformationen interessant sein könnten. Um ein Bild über die Gegebenheiten zum Zeitpunkt einer Äußerung zu erhalten wäre es von Vorteil ermitteln zu können, welche Geschehnisse direkt vorausgegangen sind und welche Situation diese hinterlassen haben. Dies würde ermöglichen alle folgenden Aussagen in diesen Kontext einzuordnen und damit ein klareres Bild des Umfeldes der Konversation zu erhalten. Außerdem wären die Geschehnisse betreffende zusätzliche Informationen nützlich sofern sie Aufschluss über mögliche Zusammenhänge zwischen aufgetretenen und eventuell folgenden Äußerungen bieten. Es wäre also von Vorteil, ein möglichst umfassendes Wissen über Zusammenhänge und Gegebenheiten zwischen vorherigen Äußerungen und folgenden Äußerungen zu besitzen, denn dies wäre hilfreich für eine möglichst präzise Einordnung jedweder weiteren betrachteten Äußerung.

5.2.1. Kontextdefinition

In Abschnitt 2.2 wurde bereits beschrieben, was in der Sprachwissenschaft unter Kontext verstanden wird. Doch, bedingt durch die Aufgabenstellung dieser Arbeit und den Gegebenheiten des *PARSE*-Projektes sowie den gerade beschriebenen wünschenswerten Kontextinformationen, unterscheidet sich der in dieser Arbeit verwendete Kontextbegriff leicht von dem in den Sprachwissenschaften. Dies liegt daran, dass der situative Kontext, also alle Zusammenhänge die nicht aus Syntax und Semantik ableitbar sind, in dieser Arbeit nur bedingt betrachtet werden kann, da als Bearbeitungsgrundlage nur die Anweisungen in textueller Form zur Verfügung stehen. Hierdurch existieren keine Informationen über die Wahrnehmung der Umgebung zum Äußerungszeitpunkt und ebenso keine Informationen über das Wissen über Zusammenhänge der die Äußerung treffenden Person. Diese Anforderung ergibt sich aus der Domänen- und Zielsystem-unabhängigen Zielsetzung des *PARSE*-Projektes, denn somit können ebenso Systeme ohne funktionierende eigene Perzeption der Umwelt betrachtet werden. Um das Fehlen der Perzeption und eines Zusammenhangswissen in der Kontextanalyse zumindest zum Teil ausgleichen zu können werden generelle Informationen aus verschiedenen Wissensquellen, wie Weltwissenontologien oder der Domänenontologie, in den betrachteten Kontext miteinbezogen. Dieses zusätzliche Wissen wird in dieser Arbeit also ebenso unter dem Begriff des Kontextes zusammengefasst.

5.2.2. Kontextinformationen im sprachlichen Kontext

Da der einzige, zunächst zur Verfügung stehende Kontext, der sprachliche Kontext ist, kommt als Quelle für erste Kontextinformationen nur dieser infrage. Um eine Vorstellung der zum sprachlichen Kontext gehörenden Informationen und damit eine Idee eines möglichst vollständigen Abbildes des Kontextes zu erhalten muss analysiert werden, welche Kontextinformationen aus diesem ableitbar sind. Betrachten wir hierzu zunächst die zwei Sätze in Beispiel 17.

- (17) a. John, go to the small fridge next to the two cupboards and open it.
 b. Then close its door again.

Als Grundbausteine des Kontextes benötigt man zunächst Informationen über die beschriebenen Gegebenheiten, also Informationen darüber welche Geschehnisse mit welchen Objekten zuvor beschrieben wurden. Sätze und vor allem Sätze im Imperativ beinhalten zumeist Subjekte, Verben und Objekte, also durch die Verben beschriebene Ereignisse, die von den Subjekten ausgeführt werden und in Beziehung zu den Objekten stehen. Subjekte und Objekte beschreiben dabei konkret oder abstrakt existierende Dinge, die an den Ereignissen teilhaben können. Also konkret oder abstrakt existierende Entitäten, welche somit

die Grundlage für eine Betrachtung des sprachlichen Kontextes darstellen. Solche Entitäten sind z.B. Gegenstände, Personen oder Systeme, die im gegebenen Diskurs auftreten. In Beispiel 17 sind dies die Entitäten: *John*, *fridge*, *two cupboards* und *door*.

Über diese Entitäten lässt sich wiederum weiteres Wissen aus der Syntax gewinnen. Zum einen lässt sich das Geschlecht der Entitäten, also ob sie Maskulin, Feminin oder Neutrum sind, sowie Informationen über die Anzahl, der mit dem Ausdruck gemeinten Entitäten durch die Singular- oder Plural-Form sowie Quantifizierern wie bei „*two cupboards*“ ableiten. Zum anderen lassen sich die einzelnen Entitäten einer von zwei Gruppen zuordnen: Einmal die Entitäten, welche als handelnde Entitäten, also Subjekte infrage kommen, wie Personen oder Systeme, und diejenigen, welche Objekte bzw. behandelte Entitäten darstellen, also alle beschriebenen nicht handlungsfähigen Gegenstände. So ist in diesem Beispiel nur die Entität *John* Subjekt und damit handelnd. Die anderen drei Entitäten nehmen die Rolle von Objekten ein. Handelnde und behandelte Entitäten unterscheiden sich hierbei in den für sie ermittelbaren Informationen. Handelnde Entitäten beschreiben zumeist Personen oder Systeme, welche unter anderem zumeist ohne einen Artikel angesprochen werden.

Im Falle eines Objektes, wie z.B. *fridge*, lässt sich hingegen bestimmen, ob es sich um eine bestimmte oder unbestimmte Nennung des Objektes handelt, also welche Form eines Determinierers (*the*, *a*, ...) verwendet wurde. Dies kann Aufschluss über die bereits erfolgte Nennung einer solchen Entität im vorangegangenen Diskursteil geben. Außerdem können Entitäten mittels Adjektiven oder auch Possessivpronomen näher beschrieben werden, wie es bei „*small fridge*“ der Fall ist. Solche näher beschreibenden Eigenschaften können hilfreich bei der Unterscheidung von Identitäten gleich benannter Entitäten sein und dafür sorgen, dass sowohl die präzisere Beschreibung der Entität, als auch ihre generalisierte Form im Kontext auftreten. So handelt es sich bei einem *red cup* nicht um die gleiche Entität wie bei einem *green cup*, aber beide Entitäten können gemeint sein wenn im darauf folgenden Diskurs von einem *cup* die Rede ist.

Die Eigenschaft des Geschlechts der Entität ist zwar prinzipiell für alle Entitäten bestimmbar, spielt aber nur bei Subjektentitäten eine differenzierende Rolle, denn die als Objektentitäten betrachteten Entitäten haben in der englischen Sprache zumeist das Geschlecht Neutrum, da es sich um beschriebene Gegenstände handelt. Bei den Subjektentitäten und insbesondere bei beschriebenen Personen hingegen ist eine Information über das Geschlecht wichtig für eine Unterscheidung. So könnte von zwei Subjekten *John* und *Mary* gesprochen werden, bei der *John* maskulin und *Mary* feminin ist. Tritt daraufhin ein Pronomen wie *he* oder *she* auf, ist die Information über das Geschlecht essentiell um entscheiden zu können, welche der beiden Entitäten gemeint ist.

Das zweite wichtige Konzept, dessen Vertreter aus dem sprachlichen Kontext abgeleitet werden können, sind die durch die Verben beschriebenen Aktionen bzw. Ereignisse. So enthält das Beispiel die Aktionen `go(THEME:John, DEST:fridge)`, `open(THEME:it)` und `close(PATIENT:door)`. Mittels dieser Informationen lässt sich weiteres Wissen über die Entitäten im Kontext ableiten, denn die Aktionen beschreiben, welche Handlungen geschehen und welche Entitäten an diesen Handlungen teilnehmen. Interessant hierbei sind vor allem die Rollen, welche die Entitäten in den Handlungen einnehmen, denn hieraus lassen sich Aussagen über die Zustände und Zustandsübergänge der Entitäten ableiten und die Bedeutung der Ereignisse genauer verstehen. So lässt sich aus der Aktion `open` ableiten, dass die Entität auf die das Pronomen *it* referenziert einen Zustand „geöffnet“ haben muss. Außerdem besitzt sie vermutlich auch den gegensätzlichen Zustand „geschlossen“ und befindet sich zum Zeitpunkt nach der Aktion im Zustand „geöffnet“, denn die Aktion `open` führt den Zustandsübergang „geschlossen“ → „geöffnet“ durch.

Außerdem lassen sich über die Rollen der Entitäten in den Aktionen hinausgehende Beziehungen aus dem sprachlichen Kontext ableiten. Ein Beispiel hierfür ist die Ortsbeziehung

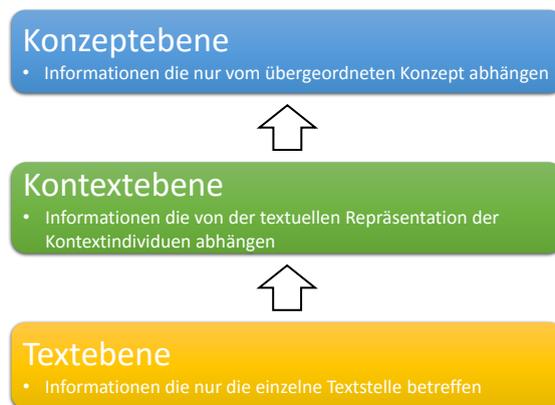


Abbildung 5.2.: Hierarchieebenen des Kontextwissens

zwischen der Entität *fridge* und den Entitäten beschrieben durch den Ausdruck *two cupboards* in Beispiel 17. Die adverbiale Bestimmung des Ortes, gebildet durch die Präposition *next to*, definiert die Position der Entität *fridge* als direkt neben den *two cupboards*.

- (18) Armar get the orange juice *and* the cup from the cupboard and then put them both on the table.

Ebenso lassen sich über eine Konjunktion in Beziehung stehende Entitäten identifizieren. So werden die beiden Entitäten *orange juice* und *cup* in Beispiel 18 im selben Kontext einer Aktion verwendet und mittels der Konjunktion *and* verbunden. Diese Information ermöglicht es dadurch erst das sich auf eine Mehrzahl beziehende Pronomen *them* aufzulösen. Solche Beziehungen bieten also ein tieferes Verständnis der Gegebenheiten und ermöglichen es Informationen über die Zusammenhänge von Aktionen und Entitäten präziser einzuordnen.

Zusätzlich zu diesen aus dem sprachlichen Kontext ableitbaren Informationen lassen sich die Entitäten und Aktionen noch mit Weltwissen und Domänenwissen anreichern. Eine Form von Wissen über Zusammenhänge in der Welt bieten die semantischen Relationen (siehe Abschnitt 2.1.8), wie z.B. Synonyme oder Hyperonyme einer Entität. Hieraus lassen sich Informationen über Beziehungen zwischen Entitäten ableiten, wie z.B. dass ein Kühlschrank im Normalfall eine Tür besitzt oder dass eine Nennung eines „TV“ gleichbedeutend mit einer Nennung „television“ sein kann. Außerdem lässt sich aus Informationen über Unter- und Oberbegriffe der Entitäten bestimmen, ob diese eine Gemeinsamkeit haben und somit eventuell in einer Äußerung tatsächlich zusammenhängen. Es lassen sich also Informationen über Synonyme, Hyperonyme, Hyponyme und Meronyme aus lexikalischen Datenbanken (vgl. Abschnitt 2.5) gewinnen und diese Informationen den Entitäten hinzufügen. Außerdem steht als weitere Wissensquelle die Domänenontologie zur Verfügung, welche Informationen über die Objekte und Systeme in der Domäne sowie deren Zustände und ausführbaren Aktionen enthält. Diese Informationen sind ebenso nützlich bei der Auflösung von Referenzen und können die aus dem Text und anderen Wissensquellen gewonnenen Erkenntnisse validieren und ergänzen, sodass z.B. klar ist, dass die Entität *fridge* eine Tür beinhaltet, welche die Zustände „geöffnet“ und „geschlossen“ besitzt.

Insgesamt lassen sich die soeben identifizierten Kontextinformationen noch hinsichtlich ihres Bezugs zur eigentlichen textuellen Repräsentation analysieren, denn nicht alle dieser Informationen sind an die eigentliche Textstelle gebunden von der sie ermittelt werden konnten. Es lässt sich nämlich feststellen, dass viele dieser Informationen unabhängig von

der eigentlichen textuellen Repräsentation und somit der Ausprägung der Entitäten und Aktionen sind, sondern vielmehr Informationen über die den Entitäten und Aktionen übergeordneten Konzepte darstellen. Informationen wie die semantischen Relationen oder auch die Geschlechtsinformation beziehen sich zwar auf den Namen der Entität oder Aktion aber nicht auf die konkrete Ausprägung, beispielsweise hinsichtlich dessen, ob es sich um eine bestimmte oder unbestimmte Nennung der Entität handelt. Die Entitäten und Aktionen, im Folgenden auch **Kontextindividuen** genannt, bilden wiederum Informationen ab die mehreren Textstellen gemeinsam sind. So wird die Entität *red cup* aus den konkreten Textstellen „red“ und „cup“ gebildet und stellt somit eine Information dar die diesen beiden Textstellen gemeinsam ist. Außerdem gibt es noch Informationen wie z.B. die aus der Vorverarbeitung stammenden Wortarten, welche für eine Textstelle gelten.

Es lassen sich also drei unterschiedliche Ebenen von Kontextwissen mit unterschiedlichen Fokussen identifizieren, welche zusammen die in Abbildung 5.2 dargestellte Hierarchie bilden, denn die Informationen der Konzeptebene beziehen sich nur auf die den Kontextindividuen übergeordneten Konzepte, die Informationen der Kontextebene hingegen auf die identifizierten Kontextindividuen, welche an die textuelle Repräsentation gebunden sind, und die Informationen der Textebene beziehen sich nur auf die Textstellen selbst.

5.2.3. Kontextgenerierung

Nachdem die zu betrachtenden Teile des Kontextes identifiziert wurden, ist zu klären, wie diese Informationen aus der zur Verfügung stehenden textuellen Repräsentation extrahiert werden können. Die in Abschnitt 5.2.2 identifizierten Informationen lassen sich in drei Hauptaspekte gliedern. Zunächst die als Grundbausteine zu betrachtenden Entitäten und alle Informationen, die diese Entitäten näher beschreiben, wie z.B. Ortsbeziehungen zwischen mehreren Entitäten. Des Weiteren die auftretenden Aktionen, die die Aktionen näher beschreibenden Informationen sowie Beziehungen zwischen den Aktionen und den erkannten Entitäten, wie z.B. die Rollen der Argumente der Aktionen. Außerdem die Informationen der Entitäten und Aktionen, welche über das Wissen aus der textuellen Repräsentation und den Vorverarbeitungsschritten hinausgehen, wie das Wissen aus der Domänen- und hinzugezogenen Weltwissenontologien und der darin eingeschlossenen Identifizierung von übergeordneten Konzepten, im Folgenden **Konzeptualisierung** genannt. Da die ersten beiden Aspekte in sich abgeschlossene Analysen der gegebenen textuellen Informationen darstellen bietet es sich an diese ebenso als eigenständige Schritte in der Kontextgenerierung zu betrachten. Dies ermöglicht die einzelnen Erkennungen unabhängig voneinander zu entwickeln und jederzeit unabhängig voneinander an neue Gegebenheiten, wie neue Informationen aus anderen Agenten des Projektes (siehe Abschnitt 3.3), anzupassen. Allerdings benötigt die Erkennung der Argument- und Rollenbeziehungen bereits ein Abbild der vorhandenen Entitäten, was dazu führt, dass eine Aktionserkennung erst nach der Erkennung der Entitäten sinnvoll durchgeführt werden kann. Die Anreicherung mit Wissen von außerhalb der textuellen Repräsentation und die dabei durchgeführte Konzeptualisierung setzt auf den gewonnenen Informationen der anderen beiden Teilaspekte auf, denn es handelt sich um eine Anreicherung basierend auf den bereits identifizierten Informationen. Dies führt dazu, dass dieser Aspekt wiederum einen eigenständigen Schritt in der Kontextgenerierung darstellt, welcher nach den beiden anderen Schritten durchgeführt wird, um Synergien der zuvor gewonnenen Informationen nutzen zu können. Auch hier ermöglicht dieses Vorgehen eine leichtere Anpassung auf veränderte Gegebenheiten oder auch einen flexibleren Austausch der als Informationsquellen herangezogenen Ontologien. Im Folgenden werden diese drei Schritte der Kontextgenerierung genauer analysiert und ein Entwurf einer möglichen Lösung der Aufgabe gegeben.

5.2.3.1. Entitätenerkennung

Um die für das aufzubauende Kontextwissen essentiellen Entitäten erkennen zu können, muss zunächst analysiert werden, in welchen Strukturen eines Satzes Entitäten beschrieben werden können. Entitäten sind in der Philosophie „seiende Dinge“, also Dinge, die sowohl abstrakt als auch real existieren können. Somit können Entitäten in Texten Personen, Gegenstände, Organisationen oder benannte Systeme, wie z.B. Roboter sein. Da allerdings Organisationen in Anweisungen nur sehr selten auftreten, da sie zumeist in beschreibenden Formulierungen verwendet werden, kann diese Kategorie in der Betrachtung vernachlässigt werden. Somit sind die zu betrachtenden Entitäten Personen, Gegenstände und agierende Systeme. Solche Entitäten werden im Regelfall durch ein Substantiv und verschiedene dieses Substantiv näher beschreibende Elemente dargestellt. Somit treten sie in der Syntaxanalyse als Bestandteil eines bestimmten Phrasentyps, der Nominalphrase (engl. *noun phrase*), auf. Hierin wird die Entität durch ein Substantiv dargestellt und durch die umgebenden Informationen, wie z.B. Adjektivattribute oder Präpositionalphrasen, weiter beschrieben. Eine Nominalphrase besteht immer aus mindestens einem Substantiv, dem sogenannten Kopf, und kann um die folgenden Elemente erweitert werden:

- **Determinierer** wie z.B. *the, a, this*
- **Adjektivattribute** wie z.B. *large, green*
- **Adjektivphrasen** und **Partizipphrasen** wie z.B. *extremely large, made of wood*
- **Substantivattribute** wie z.B. *orange* in *orange juice*
- **Präpositionalphrasen** wie z.B. *in the kitchen in the fridge in the kitchen*
- **Possessivpronomen** wie z.B. *its* in *the fridge and its door*
- **Adverbien** wie z.B. *(over) there* in *the fridge (over) there*
- **Relativsätze** wie z.B. *which is dirty*
- weitere Phrasen, die das Substantiv genauer bestimmen

Eine Ausnahme zu dieser Regel stellen Personalpronomen dar, welche ebenso eine Nominalphrase bilden. Diese Einordnung basiert auf der Auffassung einer Nominalphrase als in sich abgeschlossene syntaktische Einheit, welche durch ein Pronomen vollständig ersetzt werden kann, ohne den Satz grammatikalisch inkorrekt werden zu lassen [Loc05].

Eine Nominalphrase kann im Satz die Funktion eines Subjektes, wie bei „the fridge“ in Beispiel 19, eines Objektes, wie „the orange juice“ in Beispiel 20, oder eines Attributs zu einer anderen Phrase, wie „the dishwasher“ in Beispiel 21, erfüllen.

- (19) The fridge is green.
- (20) Armar bring me the orange juice.
- (21) Armar open the door of the dishwasher.
- (22) She has been studying all night.

Außerdem gibt es noch die Möglichkeit, dass die Nominalphrase eine adverbiale Funktion einnimmt, wie es in Beispiel 22 der Fall ist. Diese Form spielt allerdings bei der Erkennung von Entitäten keine Rolle, da sie keine neuen Entitäten einführt oder referenziert, sondern nur allgemeine Begrifflichkeiten darstellt.

Somit lassen sich Entitäten über eine Erkennung von Nominalphrasen im gegebenen Text ermitteln. Diese Information ist das Ergebnis der Phrasenerkennung (vgl. Abschnitt 2.3.2), welche Teil des Vorverarbeitungsfließbands des Projektes ist und somit in dieser Arbeit als

existent vorausgesetzt werden kann. Hierdurch lassen sich die einzelnen Nominalphrasen der Eingabe bestimmen und analysieren. Dabei können die in den Phrasen beschriebenen Entitäten über ihre Funktion im Satz und Informationen aus der Eigennamenerkennung (vgl. Abschnitt 2.3.3) in eine der drei Kategorien Subjekt (handelnde Entität), Objekt (behandelte Entität) und als Sonderform Pronomen eingeordnet werden. Pronomen sind hierbei Platzhalter für die eigentlichen Entitäten und ersetzen ihre Rolle im Satz, wobei die Ausprägung des Pronomens (z.B. 1. Person Singular) bereits Anhaltspunkt über die mögliche Kategorie der Entität liefert (z.B. *he* kann nur auf Subjektentitäten verweisen). Ordnet die Eigennamenerkennung eine der zur Nominalphrase gehörenden Textstellen als Person ein oder enthält die Nominalphrase den Eigennamen eines der in der Domänenontologie beschriebenen Systeme, so handelt es sich um eine Entität, welche in der Lage ist Anweisungen auszuführen. Ist dies nicht der Fall kann von einem beschriebenen Gegenstand ausgegangen werden oder überprüft werden ob die Nominalphrase nur aus einem Personalpronomen besteht, was dazu führen würde, dass diese Phrase eine Pronomenentität beschreibt. Da jede der Entitätsformen unterschiedliche Eigenschaften besitzt kann nach der Erkennung eine für den Typ spezielle Analyse durchgeführt werden.

Da Subjektentitäten nach der hier verwendeten Definition Personen oder Systeme darstellen besitzen die sie enthaltenen Nominalphrasen in den meisten Fällen keine der Erweiterungen, sondern werden nur aus den Eigennamen gebildet. Ein Beispiel hierfür ist die Entität *John*, welche nur aus der Nominalphrase bestehend aus der textuellen Repräsentation „John“ gebildet wird. Somit sind andere Informationen über sie deutlich wichtiger, wie z.B. das Geschlecht oder ob es sich um eine Person oder ein System handelt. Diese Informationen spezifizieren das Subjekt näher und bieten die Möglichkeit die Subjekte weiter zu klassifizieren.

Bei Objektentitäten hingegen spielen viele der Erweiterungen der Nominalphrase eine Rolle und gerade die beschreibenden Erweiterungen wie Adjektive, Possessivpronomen oder Präpositionalphrasen geben weitere Anhaltspunkte für eine spätere Differenzierung der einzelnen Objekte.

Die Pronomenentitäten-bildenden Nominalphrasen hingegen haben ebenso keine der Erweiterungen und bestehen nur aus dem Pronomen selbst, wie die Nominalphrase „it“.

Zusätzlich lassen sich einige für alle Entitäten allgemeine Informationen aus dem sprachlichen Kontext ableiten: So lässt sich aus dem Ergebnis der Wortart-Markierung (vgl. Abschnitt 2.3.1), welche ebenso Teil des Vorverarbeitungsfließbands ist, für jede der möglichen Entitäten ableiten, ob sie eine Mehrzahl von Dingen beschreibt, denn die Wortarten des Penn Tagsets beinhalten Informationen über Plural- oder Singularform der Köpfe der Nominalphrase. Einen weiteren Anhaltspunkt über Mehr- und Einzahl können die zur Nominalphrase gehörenden Wörter geben. Da die textuelle Repräsentation Zahlwörter beinhalten kann, welche mit der Wortart-Markierung *CD* versehen werden, ist es möglich zu überprüfen, ob es sich um eine Zahl handelt, die mehr als eine Entität beschreibt. Dies hilft dabei auch dann eine Aussage über Mehr- bzw. Einzahl treffen zu können, wenn die Wortarten des Nominalkopfes nicht eindeutig sind, wie es bei der Wortart-Markierung *NN*, welche ein Nomen beschreibt, welches Singular sein kann oder eine Menge darstellt, oder bei einem aus mehreren Nomen zusammengesetzten Nominalkopf wie bei „the two kitchen tables“ in Beispiel 23, der Fall ist. Diese Information, soweit vorhanden, bietet sogar Wissen über die genaue Anzahl der betrachteten Dinge.

(23) [DT the] [CD two] [NN kitchen] [NNS tables]

Bei Entitäten, die eine Subjektrolle im betrachteten Kontext einnehmen, sind die zusätzlich zu gewinnenden Informationen bei weitem nicht so vielschichtig wie bei Objektentitäten. Da Subjektentitäten durch ihre Definition nur Personen oder Systeme darstellen, die

zumeist durch Eigennamen identifiziert werden, besitzen sie keine Erweiterungen der Nominalphrase. Daher benötigt man zur genaueren Differenzierung, welche Art von Entitäten man betrachtet, weitere für Subjekte spezifische Informationen, wie z.B. eine Klassifizierung der Subjekte in Personen und handelnde Systeme. Diese Information kann durch eine Kombination der Eigennamenerkennung, welche Personen erkennt, und Informationen über die Benamung der Systeme in der Domänenontologie gewonnen werden.

(24) [NP Armar] [VP go] [PP to] [NP, PER John]

So lässt sich in Beispiel 24 die Nominalphrase „Armar“ mit dem Systemnamen *Armar* aus der Domänenontologie abgleichen und damit *Armar* als System und somit Subjektentität klassifizieren sowie die Nominalphrase „John“, welche über die Eigennamenerkennung als Person klassifiziert wurde ebenso als Subjektentität einordnen. Es wird also immer dann von einer Subjektentität ausgegangen, wenn entweder ein Teil der Nominalphrase mit dem Namen eines Systems aus der Domänenontologie übereinstimmt oder der Eigennamenerkennung einen Teil der Phrase als Person einordnet. Eine weitere Information, welche bestimmt werden kann, ist das Geschlecht der Entität, also Information darüber, ob die beschriebene Entität maskulin, feminin oder neutrum ist. Diese Information lässt sich mittels der von Bergsma und Lin [BL06] generierten Tabelle gewinnen, welche Informationen über das wahrscheinlichste Geschlecht von über 3 Millionen Nomen enthält, welche durch eine Kombination einer Korpusanalyse und Anfragen an Googles Suchmaschine entstanden ist. Das heißt ein Abgleich des als Namen der Entität aus dem Eigennamen abgeleiteten Textes mit den Einträgen aus der Tabelle gibt einen Aufschluss über das Geschlecht von häufig auftretenden Namen. Diese Informationsquelle ist vor allem bei im Englischen gebräuchlichen Personennamen relativ zuverlässig. Allerdings erhält man keine Geschlechtsinformationen für unbekannte Namen.

Die Informationen aus den Erweiterungen der Nominalphrasen, wie Adjektivattribute oder Determinierer, beziehen sich vor allem auf Objektentitäten und spezifizieren diese näher. Eine dieser Information ist die Information darüber, ob die Entität neu in den Diskurs eingeführt wurde oder bereits im Diskurs existent bzw. dem Sprecher bekannt ist. Eine Heuristik, um diese Information zu gewinnen nutzt die Determinierer, d.h. bestimmte und unbestimmte Artikel oder Ausdrücke wie *some* oder *any*, welche mit der Wortart-Markierung *DT* versehen werden. Handelt es sich um unbestimmte Artikel, beschreibt die Nominalphrase eine Entität, die nicht explizit angesprochen wird, was darauf hindeutet, dass die Entität noch nicht existiert oder eine Entität aus einer Menge von bereits spezifizierten Entitäten gemeint ist. Andersherum folgt aus einem bestimmten Artikel, dass der Sprecher die beschriebene Entität bereits kennt, also ein bestimmtes Objekt meint, und somit die Wahrscheinlichkeit steigt, dass es bereits eine vorherige Referenz auf das gleiche Konzept gibt.

(25) Go to the yellow cupboard.

Des Weiteren lassen sich zusätzliche beschreibende Informationen aus den Nominalphrasen und ihrem Zusammenspiel im Satz extrahieren. Zunächst lassen sich beschreibende Informationen, welche sich nur auf das Objekt selbst beziehen, wie Adjektivattribute/-phrasen, Adverbien oder Substantivattribute, wie *yellow* in Beispiel 25, ableiten. Diese Informationen lassen sich über ihre Wortart-Markierung und Zugehörigkeit zur Nominalphrase bestimmen und der Objektentität als spezialisierende Information zuordnen. Außerdem lassen sich näher beschreibende Zusammenhänge zu anderen Entitäten, wie sie durch Possessivpronomen ausgedrückt werden, bestimmen. Diese sind anhand der Wortart-Markierung *PRP\$* zu erkennen.

- (26) Go to the fridge and take the milk out of the refridgerator.
 (27) Take the apple from the table and put the fruit in the fridge.
 (28) This is my car. The wheels are brand new.

Des Weiteren können Informationen extrahiert werden, welche sich mittels des Namens der Entität aus anderen Wissensquellen ableiten lassen. So sind vor allem Informationen über die semantischen Relationen (siehe Abschnitt 2.1.8) des Objektes interessant, denn diese geben Aufschluss über eventuelle Zusammenhänge gleichartiger oder hierarchisch verwandter Entitäten. Diese Informationen lassen sich mittels Abfragen an eine Lexikalische Datenbank, wie z.B. WordNet (siehe Abschnitt 2.5.1), ermitteln. Gerade ein Wissen über Synonyme ermöglicht es z.B. Schlüsse über mögliche Kandidaten einer Identitätsbeziehung zu erkennen, wie dass die textuellen Repräsentationen *refridgerator* und *fridge* in Beispiel 26 dasselbe Objekt darstellen können.

Ebenfalls bietet ein Wissen über Hyper- und Hyponyme die Möglichkeit Identitäten zu erkennen, denn eine Information darüber, dass z.B. *apple* ein Hyponym von *fruit* darstellt, erlaubt es zu erkennen, dass in Beispiel 27 die Objektentität *fruit* dasselbe Objekt bezeichnet wie die Objektentität *apple*.

Die dritte interessante semantische Relation stellt das Wissen über Mero- und Holonyme dar, denn diese bieten ebenso wie Hyper- und Hyponyme Informationen über hierarchische Beziehungen, aber in diesem Fall von Teil-Ganzes-Beziehungen. So lässt sich über dieses Wissen erkennen, dass die Entität *wheels* in Beispiel 28 einen Teil der Entität *car* darstellt.

Zusätzlich existieren Informationen, welche durch Beziehungen zwischen Objekten oder Objekten und Subjekten beschrieben werden. Solche Informationen sind zumeist in Präpositionalphrasen oder Relativsätze eingebettet und enthalten Informationen, die für beide an der Beziehung teilnehmenden Entitäten relevant sind. So beschreibt die Beziehung in Beispiel 29 sowohl die Entität *cup* als auf der Entität *table*, als auch die Entität *table* als unter der Entität *cup*.

- (29) Take the cup on the table.
 (30) Take the cup [CC and] the dishes
 (31) Take the cup and wash the dishes
 instrNr=0 instrNr=1

Im Anwendungsfall von *PARSE*, Anweisungen in gesprochener Sprache, beschränken sich diese Beziehungen zumeist auf Ortsbestimmung oder stellen eine Kombination der Entitäten über eine Konjunktion dar. Die Konjunktionsbeziehungen lassen sich dabei über die Wortart-Markierung *CC* in direkter Kombination zweier identifizierter Entitäten erkennen, wie es in Beispiel 30 der Fall ist. Bei dieser Heuristik werden nur Konjunktionen mit *and* betrachtet, da alle anderen Konjunktoren, wie z.B. *but* oder *otherwise*, keine Information über eine zusammenhängende Beziehung der beiden Entitäten darstellen, welche aber die Information ist die man für ein Abbild der Beziehungen zwischen Entitäten extrahieren möchte. Bei der Erkennung dieser Beziehung ist zu beachten, dass nicht jede Kombination von zwei angrenzenden Entitäten mit einer Konjunktion zwischen ihnen auch eine Konjunktionsbeziehung bildet. Denn die Konjunktion *and* in Beispiel 31 bezieht sich auf eine Konjunktion der beiden Aktionen bzw. Satzteile und nicht auf eine Konjunktion der Entitäten. Daher können die Instruktionsnummern, welche in den die Textstellen repräsentierenden Token enthalten sind, einen Aufschluss darüber geben, ob beide Entitäten zur gleichen Instruktion gehören und damit eine Konjunktionsbeziehung existieren kann.

Ortsbestimmung	gebildet durch
Neben (<i>next_to</i>)	next to, beside
Vor (<i>front_of</i>)	in front of
Hinter (<i>behind</i>)	behind
Zwischen (<i>between</i>)	between
Gegenüber (<i>opposite</i>)	opposite to
Auf (<i>on</i>)	on, above, on top of, onto
Unter (<i>under</i>)	under, below, underneath, beneath
In (<i>in</i>)	in, inside, within, into
Links von (<i>left</i>)	left, left of, on the left side of
Rechts von (<i>right</i>)	right, right of, on the right side of
Von (<i>from</i>)	from, out of

Tabelle 5.1.: Präpositionalphrasen mit Ortsbestimmung

Die Varianten der Ortsbestimmungen sind vielseitiger als die auf die Konjunktion *and* reduzierten Konjunktionsmöglichkeiten. So existieren die in Tabelle 5.1 dargestellten Arten von Präpositionalphrasen mit Ortsbezug und die Wörter mit denen sie häufig gebildet werden. Der in der Forschung klassische Weg diese zu bestimmen wäre auf eine korrekte Syntaxanalyse aufzubauen und daraus die Präpositionalphrasen abzulesen. Da aber wie bereits zu Anfang beschrieben, bei gesprochener Sprache nicht von einer korrekten Syntaxanalyse ausgegangen werden kann, ist der einfachste Ansatz diese zu erkennen über eine Suche nach den Schlüsselwörtern zwischen zwei aufeinanderfolgenden Entitäten und einem Instruktionsnummer-Abgleich wie bei den Konjunktionsbeziehungen.

(32) $\underbrace{\text{Armar get the cup from the table}}_{\text{instrNr}=0} \underbrace{\text{then go out of the room}}_{\text{instrNr}=1}$

(33) Bring me the cup on the table next to the popcorn.

So werden in Beispiel 32 die Entitäten *Armar*, *cup*, *table* und *room* erkannt und zwischen ihnen nach Schlüsselwörtern gesucht. Die Wörter zwischen *Armar* und *cup* enthalten kein passendes Schlüsselwort, allerdings wird zwischen *cup* und *table* das Schlüsselwort *from* verwendet. Da sowohl *cup* als auch *table* Teil der Instruktion mit der Nummer Null sind wird hier die Ortsbestimmung *cup* --from--> *table* erkannt. Auch zwischen *table* und *room* wird eine Schlüsselphrase, nämlich *out of*, erkannt. Da hier allerdings die Instruktionsnummern nicht übereinstimmen wird keine Ortsbestimmung angenommen, was auch korrekt ist, da sich die Schlüsselphrase auf das Verb bezieht.

Außerdem können Ortsbestimmungen auch in Kombination auftreten, wie es in Beispiel 33 der Fall ist. Solche Kombinationen sind deutlich komplexer aufzulösen und gerade bei einer nicht vorhandenen korrekten Syntaxanalyse schwierig zu erkennen, da eine einfache Suche keine klaren Ergebnisse der Beziehungen liefert. Da eine Auflösung dieser Formulierungen für diese Arbeit zu aufwendig gewesen wäre, wird dieser Fall für spätere Erweiterungen offengelassen. Dieses Vorgehen ist zunächst valide, da im betrachteten Korpus kein Fall von Kombinationen dieser Beziehungen auftritt und somit anscheinend beim Geben von Anweisungen an Systeme solche Formulierungen vermieden werden.

Somit ergeben sich die in Abbildung 5.3 dargestellten Schritte für die Entitätenerkennung. Es wird also zunächst nach Nominalphrasen gesucht und diese gegebenenfalls an der Konjunktion *and* gespalten um beide zusammen beschriebenen Entitäten gleich in der Analyse behandeln zu können. Daraufhin lässt sich die für alle Entitätstypen allgemeine Information des Numerus bestimmen und anschließend anhand der Wortarten und den Informationen aus der Eigennamenerkennung die in der Phrase erkannte Entität in eine der

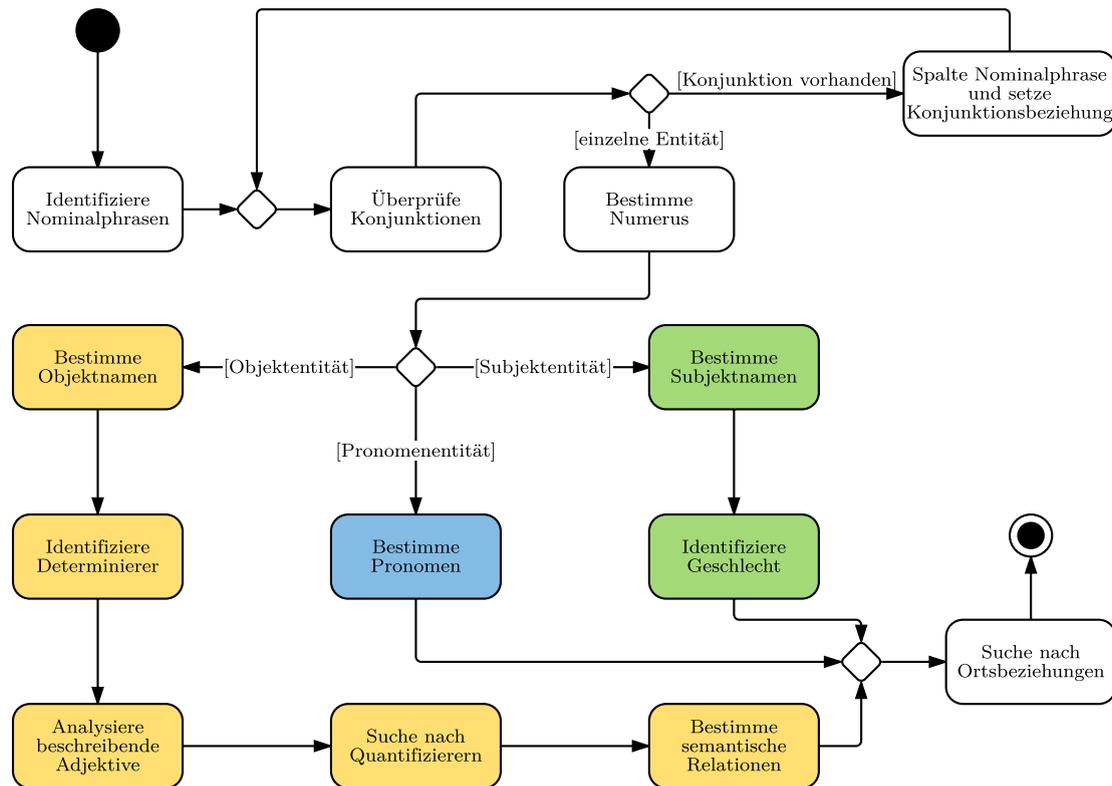


Abbildung 5.3.: Schritte der Entitätenerkennung

Kategorien Subjekt-, Objekt- oder Pronomenentität einordnen um für jeden Typ die spezifischen Informationen ermitteln zu können. Bei den Subjekten sind dies der Eigenname bzw. Systemname sowie das Geschlecht im Fall einer Person, bei den Pronomenentitäten das verwendete Pronomen und bei Objekten der Objektname aus den Substantiven, die Form des verwandten Determinierers, die verwendeten Adjektive und Quantifizierer sowie als Vorbereitung für die spätere Konzeptbildung die semantischen Relationen. Daraufhin kann zwischen den Entitäten nach Ortsbestimmungen gesucht werden.

5.2.3.2. Aktionserkennung

Um den Aufbau des Kontextwissens um Wissen über die auftretenden Aktionen und Ereignisse ergänzen zu können ist zu analysieren, wie diese in den betrachteten Anweisungen auftreten und beschrieben werden. Außerdem möchte man bestimmen wie weitere Informationen, wie z.B. die Argumente und Rollen der Argumente der Aktionen, ermittelt werden können. Als Aktionen werden in den Anweisungen auftretende Beschreibungen von konkret oder abstrakt ausführbaren Handlungen verstanden, da dies die Informationen sind, die man für ein Verständnis der beschriebenen Abläufe benötigt. Handlungen werden in der Regel durch Prädikate bestehend aus Verben und Hilfsverben beschrieben, welche analog zu den Entitäten Teil eines bestimmten Phrasentyps sind, der Verbalphrase (engl. *verb phrase*). Eine solche Verbalphrase besteht immer aus mindestens einem Verb, welches den Kopf der Phrase bildet und die Handlung beschreibt, und weiteren der Formulierung dienenden Hilfsverben. Interessant für den Kontext sind hierbei zum einen die Grundform der Aktion selbst, also das Lemma bzw. die Lemmata der die Aktion abbildenden Verben, da mittels dieser ein Abgleich zwischen mehreren die gleiche Handlung beschreibenden Aktionen durchgeführt werden kann oder ein Abgleich mit den in der Do-

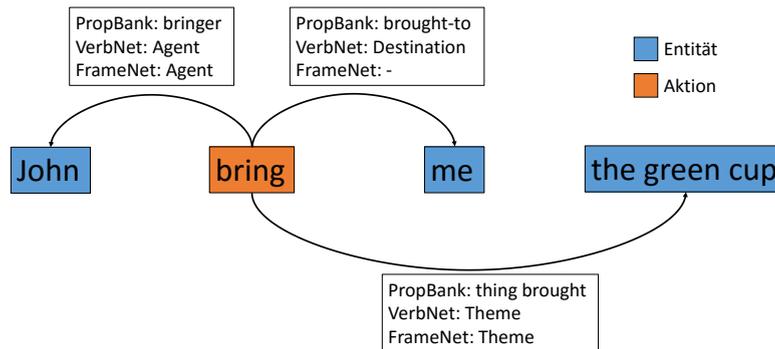


Abbildung 5.4.: Ergebnisse des SRL in Verbindung mit Aktionen und Entitäten

mänenontologie beschriebenen Methoden möglich ist. Zum anderen sind dies die an der Aktion teilnehmenden Entitäten und ihre Rollen im Bezug zu dieser Aktion, welche es ermöglichen Zusammenhänge zwischen Aktionen und Entitäten abzubilden und somit z.B. Schlüsse über Zustände und Zustandsübergänge der Entitäten ziehen zu können.

Ersteres lässt sich aus der Verbalphrase selbst ableiten, indem die Lemmata der Vollverben in der Verbalphrase herangezogen werden. Letzteres hingegen stellt eine komplexere Aufgabe dar. Die Erkennung der semantischen Rollen der Argumente einer Aktion entspricht der in Abschnitt 2.3.4 beschriebenen Aufgabe des Markierens semantischer Rollen für die es in der Forschung bereits einige nutzbare Werkzeuge gibt. Das Problem bei der Verwendung dieser Werkzeuge stellen wiederum die nicht vorhandenen Satzzeichen in aus gesprochener Sprache generiertem Text dar, was dazu führt, dass die Ergebnisse der Werkzeuge schlechter werden. Dieses Problem lässt sich im Fall von Anweisungen zumindest teilweise dadurch umgehen, dass Argumente einer Aktion in Anweisungen meist innerhalb des gleichen von Verbalphrasen und Konjunktionen begrenzten Teiles eines Satzes stehen. Durch diese Annahme lassen sich die in der Vorverarbeitung bereits identifizierten Instruktionen, welche genau diese Satzteile abbilden, einzeln als Eingabe für die Verarbeitung der Werkzeuge benutzen und somit in vielen Fällen annehmbare Ergebnisse erzeugen. Es lässt sich also ein Werkzeug zu Markierung der semantischen Rollen (im Folgenden kurz SRL-Werkzeug) verwenden, welches Informationen über die semantischen Rollen liefert.

Damit lassen sich für jede aus den Verbalphrasen bestimmte Aktion die Rollen der Argumente dadurch bestimmen, dass nach einer vom verwendeten SRL-Werkzeug erkannten Aktion in der Verbalphrase gesucht wird und die vom SRL-Werkzeug zu dieser gefundenen Aktion bestimmten Argumente samt ihrer Rollen der Entität zugeordnet werden, auf die das Argument zeigt. Dabei lassen sich die Rollen verschiedener Datenbanken wie PropBank, VerbNet oder FrameNet je nach verwendetem Werkzeug und vorhandenen Übereinstimmungen der Datenbanken ermitteln und dem Kontextwissen hinzufügen. In Abbildung 5.4 ist die Aktion *bring* mit den jeweiligen Zuordnungen der Entitäten *John*, *me* und *green cup* und ihren Rollen dargestellt.

Zusätzlich zu diesen Informationen lässt sich der Aktionsname, welcher anhand des erkannten Hauptverbs wird, dazu nutzen weitere Informationen über die Aktion zu ermitteln. So wäre es von Vorteil zu wissen, welche anderen Bezeichnungen ebenfalls die gleiche Aktion beschreiben oder welche Begrifflichkeiten eine Aktion beschreiben die eine Verallgemeinerung der beschriebenen Aktion darstellen. Solche Informationen lassen sich aus den semantischen Relationen (siehe Abschnitt 2.1.8) der die Aktion beschreibenden Verben ableiten und über eine Anbindung an eine Lexikalische Datenbank wie z.B. WordNet (siehe Abschnitt 2.5.1) generieren. Interessant sind hierbei vor allem Synonyme und Antonyme, da diese Aufschluss über weitere Aktionen enthalten können, die ähnliche oder

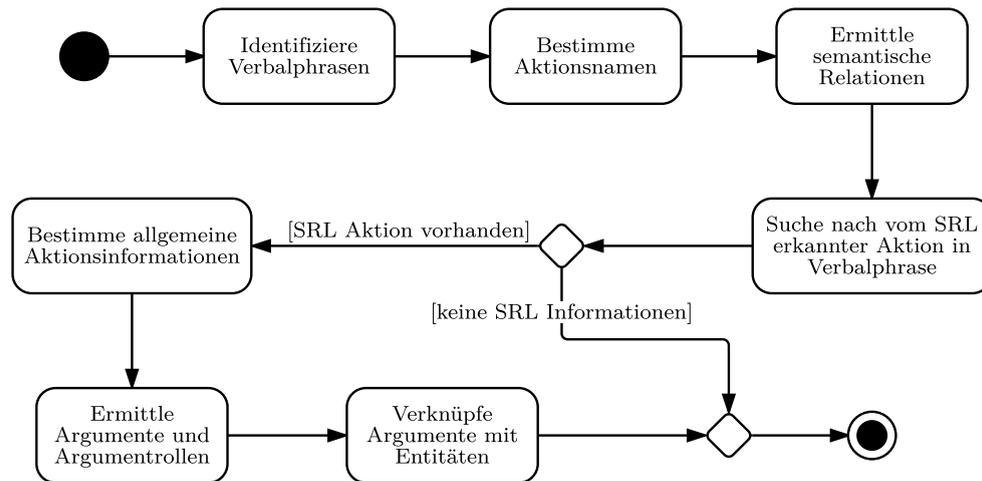


Abbildung 5.5.: Schritte der Aktionserkennung

entgegengesetzte Aktionen beschreiben und damit dazu genutzt werden können eventuelle Zusammenhänge wie Zustände von Entitäten aus den Aktionen abzuleiten. Aber auch Hyper- und Hyponyme sind interessant, da diese hierarchische Zusammenhänge zwischen Aktionen abbilden können. Diese Informationen lassen sich wie auch schon bei den Entitäten aus WordNet (siehe Abschnitt 2.5.1) bestimmen und werden in der Aktionserkennung vorbereitend für die Konzeptualisierung bestimmt.

Somit ergeben sich die in Abbildung 5.5 dargestellten Schritte für die Aktionserkennung. Zunächst werden die Verbalphrasen bestimmt und aus ihnen der Aktionsname, bzw. das vom SRL-Werkzeug erkannte Verb bestimmt. Sofern vorhanden lassen sich dann die Aktionsinformationen des SRL-Werkzeugs bestimmen und die Argumente und semantischen Rollen mit den Entitäten verknüpfen.

5.2.3.3. Konzeptualisierung

Der dritte Schritt der Kontextgenerierung befasst sich mit der Abstraktion von Informationen aus den erkannten Kontextindividuen, welche unabhängig von der konkreten textuellen Repräsentation der Kontextindividuen sind, und der darin enthaltenen Bildung von übergeordneten Konzepten, der sogenannten **Konzeptualisierung**. Zusätzlich beinhaltet dieser Prozess das Miteinbeziehen des Wissens aus der Domänenontologie und Weltwissen, denn in der Domänenontologie sind nur von textuellen Repräsentationen unabhängige Informationen enthalten, welche somit nur in die Konzeptebene einfließen können. Man möchte also identifizieren können, dass zwei an die textuellen Repräsentationen *fridge* und *refridgerator* gebundenen Entitäten das gleiche Konzept darstellen.

Zunächst lässt sich analysieren, welche der in Abschnitt 5.2.2 identifizierten Informationen und letztlich in Abschnitt 5.2.3 für die Kontextindividuen als generierbar ermittelten Informationen sich auf ein übergeordnetes Konzept übertragen lassen und damit unabhängig von der konkreten textuellen Ausprägung sind. Generell lässt sich feststellen, dass alle sich auf den Namen des Kontextindividuums beziehenden Informationen und somit aus den Lemmata ermittelten Informationen sich auf das übergeordnete Konzept übertragen lassen, denn sie sind durch die Lemmatisierung unabhängig von den konkreten Ausprägungen. Dies sind bei den Entitäten zunächst der Name selbst sowie das Geschlecht, die Synonyme, Hyper-/Hyponyme und Mero-/Holonyme sofern es sich um den entsprechenden Typ von Entität handelt. Bei den Aktionen sind dies ebenso der Name sowie die

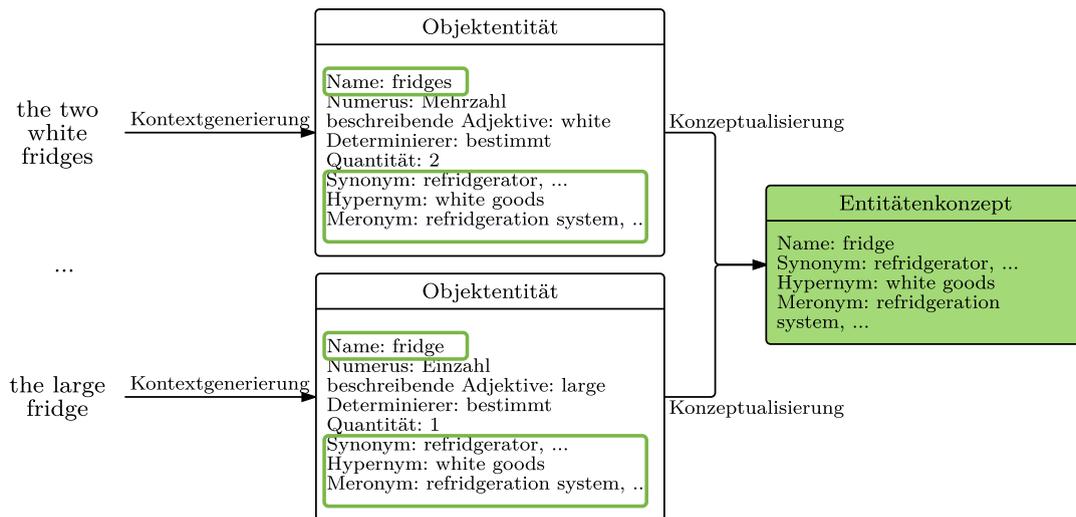


Abbildung 5.6.: Abstraktion eines Konzeptes aus Objektentitäten

Beschreibung der Aktion aus einer der Lexikalischen Datenbanken, die Syno-/Antonyme und Hyper-/Hyponyme. Das heißt aus einer Objektentität mit der textuellen Repräsentation „the two white fridges“ und einer Objektentität mit der textuellen Repräsentation „the large fridge“ lässt sich das gemeinsame übergeordnete Konzept *fridge* ableiten. Dieser Vorgang ist in Abbildung 5.6 dargestellt.

Neben diesen aus den Kontextindividuen ableitbaren Informationen lassen sich auf der Konzeptebene weitere Informationen durch eine Anbindung an die Domänenontologie ermitteln. Die Domänenontologie enthält Informationen über die in der Domäne auftretenden Systeme, Objekte und Methoden (siehe Abschnitt 3.4), welche sich mit den gefundenen Kontextindividuen bzw. den aus ihnen abgeleiteten Konzepten abgleichen lassen. Für diesen Abgleich kann die Information über den Typ des Kontextindividuums, also ob es sich um eine Subjekt- oder Objektentität oder eine Aktion handelt, genutzt werden, um den betrachteten Suchraum einzuschränken. So können Subjektentitäten wegen ihrer Definition (vgl. Abschnitt 5.2.2) nur mit einem System aus der Domänenontologie zusammenhängen, Objektentitäten hingegen Objekte der Domänenontologie widerspiegeln und Aktionen sich auf Methoden beziehen. Um den eigentlichen Abgleich durchzuführen müssen vergleichbare Benamungen der möglichen Konzept-Ontologieindividuum-Paarungen geschaffen werden. Die Ontologieindividuen können Namen in der Form *OrangeJuice* besitzen, welche bei einem einfachen String-Vergleich nicht mit einem *orange juice*-Konzept übereinstimmen würden. Daher liegt es nahe eine Ähnlichkeitsmetrik für Strings als Abgleich zu verwenden. Gerade für die Fälle von einzelnen abweichenden Zeichen bietet sich hierbei die Jaro-Winkler-Distanz an, welche einen prozentualen Wert für die Ähnlichkeit zweier Strings angibt [CRF03]. Hiermit lässt sich dann ein Schwellwert für einen Treffer eines Abgleichs definieren der zwischen 90% und 100% liegen sollte, denn alle darunter liegenden Werte würden zu viele falsch positive Einordnungen ergeben.

Zusätzlich zu diesem Abgleich-Problem existieren in der Domänenontologie auch Objektindividuen der Form *Fridge.Door*, welche ebenso mit einem Konzept *door* übereinstimmen sollten. Hierbei liegt die Schwierigkeit darin, dass es mehrere Objektindividuen mit dem Teilobjekt *Door* geben kann und deshalb auszuwählen ist, welches Objektindividuum den eigentlichen Treffer darstellt oder ob es sich um ein eigenständiges *door*-Konzept handelt. Um dies zu bestimmen lässt sich zunächst überprüfen, ob es sich bei der Nennung des Kontextindividuums um eine unbestimmte Nennung handelt und damit eine eigenständige

Entität *door* beschrieben wird oder ob die Nennung mittels des Determinierers bestimmt ist und somit die *door*-Entität bereits in den Kontext eingeführt wurde. Existieren in letzterem Fall keine zuvor genannten Entitäten, die mit einem der *Door* übergeordneten Objektindividuen aus der Domänenontologie übereinstimmen, handelt es sich vermutlich um eine eigenständige *door*-Entität. Ansonsten lässt sich der Bezug auf das Objektindividuum der am nächsten stehenden Entität mit zugeordnetem passendem Objektindividuum auflösen.

(34) Go to the fridge and then to the cupboard afterwards close the door

Das Konzept *door* in Beispiel 34 lässt sich mit dem Objektindividuum *Cupboard.Door* verknüpfen, obwohl ebenfalls ein Objektindividuum *Fridge.Door* in der Domänenontologie vorhanden ist, da die *cupboard*-Entität syntaktisch näher zur Entität *door* steht und eindeutig mit dem Objektindividuum *Cupboard* der Domänenontologie verbunden werden konnte. Die Heuristik findet hierbei zunächst für die Entität *door* die beiden passenden Objektindividuen *Fridge.Door* und *Cupboard.Door* woraufhin nach Entitäten gesucht wird, die vor dieser Entität genannt wurden und ein diese Konzepte beinhaltendes Konzept repräsentieren. In diesem Fall sind dies die Entitäten *fridge* und *cupboard*, wobei *cupboard* die syntaktisch näher zur *door*-Entität stehende Entität darstellt und damit als das wahrscheinlichere Konzept angenommen wird. Hieraus folgt dann die Zuordnung der *door*-Entität zum *Cupboard.Door*-Konzept.

Existiert also ein für das gefundene Konzept passendes Domänenontologieindividuum lassen sich je nach Typ weitere Informationen aus der Domänenontologie in die Konzeptebene einbauen. Für Systeme, Objekte und Methoden enthält die Domänenontologie Informationen über gleichwertige Benamungen desselben Konzeptes, wodurch sich die Synonyme des Konzeptes erweitern lassen. Zusätzlich enthält die Domänenontologie Beziehungen zwischen Objekten wie Teil-Ganzes-Beziehungen, in Form von Sub- und Superobjekten, sowie Beziehungen zu Zuständen, die das Objekt annehmen kann. Ersteres entspricht den Meronym-Beziehungen, die in Abschnitt 5.2.3.1 bereits als interessant für den Kontext identifiziert wurden. Letzteres hingegen bietet eine neue Information, die sehr interessant für die Betrachtung der auftretenden Aktionen ist. Da ein Wissen über Zustände der Objekte und Wissen über Zustandsübergänge, ausgelöst durch die Aktionen, dabei helfen kann Konstrukte wie in Beispiel 35 aufzulösen.

(35) Open the fridge and go to the closed cupboard then close it again

Ist bekannt, dass das Konzept *fridge* die Zustände *opened* und *closed* hat und die Aktion *open* einen Zustandsübergang in den Zustand *opened* erzeugt sowie dass die Aktion *close* einen Zustandsübergang nach *closed* durchführt, kann für den referierenden Ausdruck *it* nur die Entität *fridge* als Referent infrage kommen. Außerdem bieten Zustände ein genaueres Verständnis der aktuellen Situation zu einem bestimmten Zeitpunkt einer Äußerung, was für ein umfassendes Verständnis der Zusammenhänge nötig ist.

Deshalb ist es sinnvoll die Zustände der verknüpften Objekte ebenfalls in die Konzeptebene zu übertragen und als eigenständige Konzepte zu modellieren, da diese von mehreren Konzepten geteilt werden können und in der Domänenontologie weitere die Zustände näher beschreibende Informationen enthalten sind. Für Zustände enthält die Domänenontologie nämlich ebenfalls Informationen über mit einem Zustand in Verbindung stehende andere Zustände. Dies erlaubt es mittels eines gegebenen Zustands und eines Zustandsübergangs zu überprüfen, ob der gegebene Zustand in einem Zusammenhang mit dem Zustand zu dem gewechselt wird steht und damit dieser Übergang den gegebenen Zustand verändern kann.

Für Methoden enthält die Domänenontologie genau die für eine Überprüfung der eben

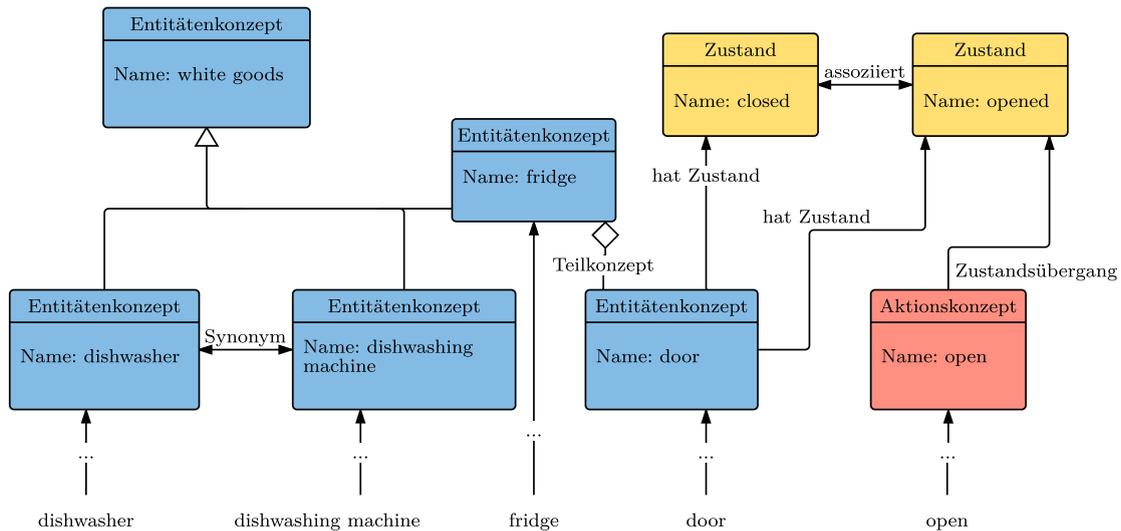


Abbildung 5.7.: Mit Domänenontologie erweiterte Konzeptualisierungshierarchie

beschriebenen Konstruktion erforderlichen Zustandsübergänge, welche von einer Methode ausgelöst werden können. Ein solcher Zustandsübergang gibt dabei den Zustand an in dem sich das behandelte Objekt nach der Ausführung der Aktion befindet. Dies wäre im Fall der Methode `open` ein Zustandsübergang in den Zustand `opened`.

Somit lassen sich die Zustände sowie neue Teil-Ganzes- und Synonym- bzw. Antonym-Beziehungen in die Konzeptebene übertragen und einige der durch das Fehlen der Perception und eines erlernten Weltwissens dem System im Gegensatz zum Menschen nicht zur Verfügung stehenden Informationen ausgleichen. Bei den Informationen aus der Domänenontologie kann, im Gegensatz zu den aus den Kontextindividuen extrahierten Informationen, eine gewisse Korrektheit angenommen werden, da die Informationen in der Ontologie nicht auf einer Heuristik beruhen sondern von Hand erzeugt wurden. Dies bietet die Möglichkeit die aus einer Lexikalischen Datenbank extrahierten Hierarchie- und Beziehungsinformationen, wie Synonyme und Meronyme, zu validieren sowie noch einen Schritt weiter in der Konzeptualisierung zu gehen. Die aus der Domänenontologie übertragenen gleichartigen Konzepte sowie Teil- bzw. Überkonzepte lassen sich als eigenständige Konzepte in die Konzeptebene einbauen und damit eine Konzepthierarchie erstellen. Ein Beispiel für eine solche resultierende Hierarchie für die gefundenen Kontextindividuen *dishwasher*, *dishwashing machine*, *fridge*, *door* und *open* ist in Abbildung 5.7 dargestellt. Hierbei entstammen die Synonym-, Teilkonzept- und Zustandsinformationen aus der Domänenontologie.

Das Entitätenkonzept *white goods* und die diesem zugeordneten Vererbungsbeziehungen hingegen stellen einen weiteren Schritt der Konzeptualisierung dar, denn die Domänenontologie deckt zwar Teil-Ganzes-Beziehungen zwischen Objekten ab, enthält aber keine Informationen über Hyperonym- bzw. Hyponym-Beziehungen zwischen Objekten. Diese Information lässt sich aber aus einer Anbindung an WordNet generieren, welches eine Hyperonym-Hierarchie bietet, die für Nomen den Wurzelknoten *entity* besitzt. Hiermit lassen sich über alle gefundenen Objektkonzepte hinweg gemeinsame Überbegriffe suchen und diese der Konzeptebene hinzufügen. Dies bietet die Möglichkeit das generierte Kontextwissen auch um ein Zusammenhangswissen zu erweitern und damit einordnen zu können, welche Konzepte aus einem ähnlichen Bereich stammen bzw. ähnliche Überbegriffe teilen. Da allerdings für viele Kombinationen der zentrale Wurzelknoten oder andere wenig spezifische Überbegriffe als gemeinsame Überbegriffe gefunden werden ist es sinnvoll ein

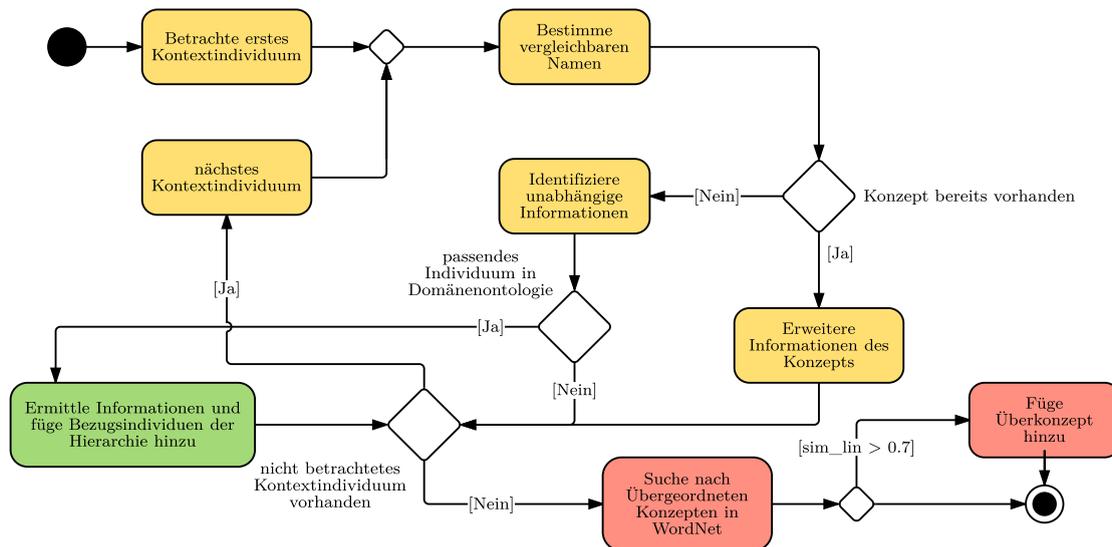


Abbildung 5.8.: Übersicht über den Ablauf der Konzeptualisierung

gemeinsames Überkonzept nur dann in die Konzeptebene einzuführen, wenn es sich um ein hinreichend spezifisches Überkonzept handelt. Um dies zu bewerten existieren verschiedene Ähnlichkeitsmetriken, die auf dem niedrigsten gemeinsamen Überbegriff (engl. *Lowest Common Subsumer* kurz *LCS*) aufsetzen. Da eine reine Distanzmessung bzw. Pfadlängenmessung in WordNet keine Aussagekraft über die Spezifität des Überbegriffes gibt, bietet sich für den Vergleich eine auf dem Informationsgehalt basierende Ähnlichkeitsmetrik, wie die Methode von Lin [Lin98], an.

$$sim_{Lin}(c_1, c_2) = \frac{2 \log(P(LCS(c_1, c_2)))}{\log(P(c_1)) + \log(P(c_2))} \quad (5.1)$$

Hierbei sind sich zwei Konzepte dann ähnlicher, wenn der Informationsgehalt den beide Konzepte teilen (im WordNet-Fall die doppelte Auftrittswahrscheinlichkeit des *LCS*) eine geringere Differenz zum summierten gesamten Informationsgehalt der Konzepte hat (vgl. Gleichung 5.1). Somit haben zwei Konzepte die einen *LCS* teilen, der einen ähnlich hohen Informationsgehalt besitzt und somit ähnlich spezifisch ist wie die beiden betrachteten Konzepte, eine hohen Ähnlichkeitswert nach Lin. Damit lässt sich ein Schwellwert für in die Konzeptebene zu übertragende Überkonzepte wählen, der dafür sorgt, dass nur spezifische Überbegriffe in die Hierarchie übertragen werden. Eine Evaluation der im Korpus auftretenden Konzepte ergab einen Schwellwert von 0,75 von möglichen 1,0 als validen Ähnlichkeitswert. Außerdem lassen sich hierdurch weitere Hyperonym- und Hyponym-Beziehungen aus WordNet extrahieren, denn ist der *LCS* zweier Konzepte eines dieser Konzepte selbst, lässt sich dieses als Überkonzept des Anderen identifizieren.

Somit ergibt sich für die Konzeptualisierung der in Abbildung 5.8 dargestellte Ablauf mit den drei farblich unterschiedlich dargestellten Teilschritten. Gelb beschreibt die Identifizierung der Grundkonzepte aus den von der textuellen Repräsentation unabhängigen Informationen. Mit grün sind die Schritte der Domänenontologieverknüpfung dargestellt und rot stellt die Suche nach übergeordneten Konzepten in WordNet dar.

5.2.3.4. Zustandsanalyse

In der Konzeptualisierung wurden bisher nur die aus der Domänenontologie extrahierten Zustände betrachtet. Diese stellen alle in der Domänenontologie abgebildeten möglichen

Zustände der Konzepte dar, geben aber keine Auskunft darüber, in welchen Zuständen sich die Kontextindividuen zu einem bestimmten Zeitpunkt befinden. Diese Information ist wiederum eine Information auf der Kontext- und nicht auf der Konzeptebene, da die Zustände der Entitäten von den konkreten Ausprägungen und ausgeführten Aktionen in den Anweisungen abhängen. Es handelt sich also hierbei um ein Beispiel in dem das Wissen aus der Konzeptebene in Kombination mit dem Wissen aus der Kontextebene zu neuen Informationen über die Individuen der Kontextebene führt. Es wäre ebenso möglich auch Zustände zu betrachten, die nicht in der Domänenontologie vorhanden sind, was aber die Ermittlung potenzieller Kandidaten für Zustände um einiges komplexer macht, weshalb sich nur auf die Zustände konzentriert wurde, welche in der Domänenontologie abgebildet sind.

Da die Perzeption der Umgebung und auch eine manuelle Beschreibung der initialen Gegebenheiten im *PARSE*-Projekt zwecks einer Austauschbarkeit der Domäne und des Zielsystems nicht betrachtet wird, benötigt man zunächst eine Möglichkeit den initialen Zustand der Entitäten zu ermitteln um darauf aufbauend die durch die Aktionen beschriebenen Aktionsübergänge zu analysieren. Es existieren also initial nur Informationen über die möglichen Zustände der Entitäten, die einem Konzept angehören, das einen Gegenpart in der Domänenontologie besitzt, welcher Zustände modelliert hat. Dies können in der bisherigen Ausprägung der Domänenontologie nur Objektentitäten sein, denn diese können mit Objekten in der Domänenontologie verknüpft werden, welche als einzige Ontologieindividuen mögliche Zustände besitzen. Diese möglichen Zustände geben also die Menge der infrage kommenden initialen Zustände vor. Um nun aber herauszufinden, in welchem Zustand sich eine auftretende Entität gerade befinden kann, lassen sich die Informationen aus den Aktionen sowie den beschreibenden Teilen der Nominalphrasen nutzen.

(36) Take the full cup from the table

(37) Go to the television which is turned on and turn it off

Der einfachste auftretende Fall ist eine Beschreibung des gegenwärtigen Zustands einer Entität in den Adjektivattributen der Nominalphrase. Hierbei lässt sich ein Abgleich zwischen den Adjektiven und den möglichen Zuständen des übergeordneten Konzeptes durchführen und darüber erkennen, in welchem Zustand sich die Entität befindet. In Beispiel 36 lässt sich somit der Zustand *full* aus der die *cup*-Entität beschreibenden Nominalphrase *the full cup* ableiten, sofern das der *cup*-Entität übergeordnete *cup*-Konzept, den Zustand *full* als möglichen Zustand besitzt. Um diesen Abgleich durchzuführen muss die Adjektivbeschreibung mit der textuellen Repräsentation des Zustands übereinstimmen oder zumindest ein Synonym dieser sein, denn ohne eine solche Übereinstimmung lässt sich der Zustand nicht zuordnen. Hierbei gibt es einige Sonderfälle in denen diese simple Heuristik versagt oder in der die Beschreibungen komplexer sind, wie z.B. der Einsatz eines Relativsatzes zur Beschreibung des Zustands, wie es in Beispiel 37 der Fall ist. Hier wird der Zustand *on* der *television*-Entität durch den Relativsatz „which is turned on“ beschrieben und würde eine Erkennung von Relativsätzen voraussetzen, welche ohne eine korrekte Syntaxanalyse nicht ohne Weiteres durchgeführt werden kann. Da außerdem solche Beschreibungen im jetzigen Korpus nicht auftreten werden diese Fälle in dieser Arbeit nicht betrachtet und für eine mögliche spätere Arbeit offen gelassen. Daher stellt die in dieser Arbeit durchgeführte Zustandsanalyse eine erste simple Heuristik für die Bestimmung des Zustands der Entitäten dar, welche zeigen soll, dass einige Fälle von Zuständen mit einfachen Kontextinformationen ermittelt werden können, sowie dass diese Information nützlich für die spätere Korreferenzanalyse sein kann.

Eine Möglichkeit, die allerdings mit den als Kontext generierten Informationen abgedeckt werden kann, ist das Bestimmen des Zustands einer Entität über den Zustandsübergang

der Aktion an der sie teilnimmt. Voraussetzung sind hierfür, zusätzlich zu einer Information über die möglichen Zustände der Entität, auch Informationen über die durch die Aktion ausgelösten Zustandsübergänge, welche ebenso von der Anbindung an die Domänenontologie abhängt. Sind beide Informationen vorhanden lässt sich der momentane Zustand einer Entität dadurch bestimmen, dass überprüft wird, ob der Endzustand des Zustandsübergangs, den die Aktion auslöst, einen der möglichen Zustände der Objektentität darstellt und somit die Entität sich nach der Durchführung dieser Aktion vermutlich in diesem Zustand befindet.

(38) Go to the fridge and open the fridge

So kann in Beispiel 38 für die zweite *fridge*-Entität der Zustand *opened* gefolgert werden, da sie mit dem Konzept *Fridge* verbunden werden konnte, welches die möglichen Zustände *opened* und *closed* besitzt und die Aktion *open* mit dem Konzept *open* verbunden wurde, welches einen Zustandsübergang in den Zustand *opened* definiert. Dadurch dass die *fridge*-Entität das behandelte Objekt der *open*-Aktion darstellt, lässt sich daher der Zustand folgern.

Interessant an dieser Möglichkeit ist aber vor allem, dass sie es ebenso ermöglicht für Pronomen Zustände zu ermitteln, was für eine spätere Korreferenzanalyse von Vorteil sein könnte. Dies liegt daran, dass auch die Pronomen als behandelte Objekte einer Aktion auftreten können. Dies ermöglicht es, die von den Zustandsübergängen der Aktionen beschriebenen Endzustände als mögliche Zustände in denen sich das Pronomen befinden können muss zu identifizieren und damit einen Referenten zu finden der dies erfüllt.

Die bisher beschriebenen Zustandsinformationen schließen nur die direkt mit der Textstelle zusammenhängenden Zustände mit ein. Aber eine Entität, die z.B. eine Wiederholung einer anderen bereits beschriebenen Entität darstellt, also Teil einer Objektidentität ist oder aber ein Pronomen, welches als Referenten eine andere Entität besitzt, befindet sich nicht nur in den Zuständen, die direkt an dieser Textstelle beschrieben werden, sondern erbt die Zustände der zuvor beschriebenen Referenten. Dies stellt also eine der Möglichkeiten dar, in denen sich Kontext- und Korreferenzanalyse durch ein Zusammenspiel gegenseitig unterstützen können, und untermauert damit, dass der im *PARSE*-Projekt gewählte agentenbasierte Ansatz sinnvoll ist.

(39) Take the dirty cup fill it with water and bring it to me

(40) If the fridge is opened get the cup out of it

In einer initialen Zustandsanalyse würden die Entitäten in Beispiel 39 die folgenden Zustände besitzen: *cup[dirty]*, *it[full]* und *it[]*. Nachdem allerdings Informationen über die Korreferenzbeziehungen zwischen den Entitäten existieren und damit klar ist, dass sich das letzte *it* auf das vorherige *it* und dieses wiederum auf die *cup*-Entität bezieht, können die folgenden Zustände gefolgert werden: *cup[dirty]*, *it[dirty, full]* und *it[dirty, full]*.

Bei diesem folgernden Ansatz müssen die Informationen über Bedingungen miteinbezogen werden. Würde man diese Informationen nicht mit einbeziehen, könnte aus Formulierungen von Aktionen oder Adjektiven im Bedingungsteil einer Anweisung, die Zustände beschreiben, fälschlicherweise ein angenommener Zustand gefolgert werden. So darf für die Entität *fridge* und das sich darauf beziehende Pronomen *it* in Beispiel 40 nicht der Zustand *opened* gefolgert werden, da es sich bei der *open*-Aktion um eine Bedingung handelt und nicht um eine Anweisung.

5.2.4. Repräsentation des Kontextwissens

Die in der Kontextanalyse als ermittelbarer Kontext identifizierte Informationen lassen sich wie im Folgenden beschrieben zusammenfassen. Zunächst sind dies die aus den Nominal- und Verbalphrasen ermittelten Kontextindividuen. Dies sind die Entitäten und Aktionen, welche in der betrachteten Eingabe auftreten. Die Entitäten lassen sich hierbei in Subjekt- und Objektentitäten sowie als Platzhalter für die eigentliche Entität eingesetzte Pronomenentitäten unterteilen. Die Subjektentitäten, welche handelnde Subjekte repräsentieren, beinhalten hierbei Informationen über den Namen der Entität, das Geschlecht sowie den Numerus. Objektentitäten, die behandelte Gegenstände darstellen, enthalten ebenfalls Informationen über den Namen und den Numerus der Entität aber zusätzlich Informationen über den verwendeten Determinierer, die die Entität beschreibenden Adjektive, Possessivpronomen und Quantifizierer sowie Informationen über semantische Relationen zu anderen Begriffen wie Synonyme, Hyper- / Hyponyme und Mero- / Holonyme. Zusätzlich dazu können Objektentitäten Zustände besitzen. Die Pronomenentitäten, welche eigentlich nur einen Bezug zu einer anderen Entität darstellen und somit im Kontext nur eine Platzhalterfunktion für eine spätere Korreferenzanalyse darstellen, enthalten ebenso Informationen über mögliche Zustände, sowie über das Pronomen mit dem sie gebildet wurden und den Numerus den dieses Pronomen innehat.

Des Weiteren wurden zwei Formen von Beziehungen zwischen Entitäten identifiziert: Zunächst die Konjunktionsbeziehung, welche angibt, dass zwei Entitäten im gleichen Kontext einer Aktion mit einer Konjunktion verknüpft auftreten sowie die Ortsbeziehung, welche Informationen über adverbiale Bestimmungen des Ortes zwischen Entitäten abbildet.

Für die Aktionen wurden als Kontextinformationen über das bildende Prädikat, die Beschreibung der ausgeführten Handlung sowie ebenfalls Informationen über semantische Relationen wie synonyme oder antonyme Aktionen identifiziert. Außerdem ist das Zusammenspiel der Aktionen mit den Entitäten, vor allem die semantischen Rollen der an einer Aktion teilhabenden Entitäten, besonders wichtig für ein möglichst umfassendes Verständnis der geäußerten Gegebenheiten.

Diese Kontextinformationen sind die aus der textuellen Repräsentation ermittelten initialen Informationen. Über diese Kontextindividuen hinaus wurden diesen übergeordnete Konzepte identifiziert, welche die von der textuellen Repräsentation und somit konkreten Ausprägung unabhängigen und mit Domänen- und Weltwissen erweiterten Informationen enthalten. Diese lassen sich wiederum je nach repräsentierbarer Art der Kontextindividuen in die Kategorien Subjekt-, Objekt- und Aktionskonzept unterteilen. Zusätzlich lassen sich aus der Domänenontologie noch annehmbare Zustände der Objektkonzepte ableiten. Diese enthalten wiederum unterschiedliche Informationen je nachdem, was aus den untergeordneten Kontextindividuen ableitbar und mit der Domänenontologie und Weltwissensdatenbanken anreicherbaren Informationen zu gewinnen ist.

Allgemein lassen sich für alle Konzepte gleichwertige sowie über- und untergeordnete Konzepte identifizieren. Für Subjektkonzepte ist außerdem die Geschlechtsinformation unabhängig von der konkreten Ausprägung. Bei den Objektkonzepten lassen sich außerdem Teil-Ganzes Beziehungen zwischen Konzepten identifizieren sowie mögliche Zustände aus der Domänenontologie ableiten. Für die Aktionskonzepte sind außerdem antonyme Konzepte interessant sowie die durch sie ausgelösten Zustandsübergänge. Die Zustände stellen ein eigenständiges Konzept dar, welches den Namen des Zustands sowie seine assoziierten Wechselzustände enthält.

Betrachtet man nun diese identifizierten Informationen im Hinblick einer möglichen Repräsentation und berücksichtigt, dass die verschiedenen Informationen, wie in Abschnitt 5.2.2 beschrieben, unterschiedliche Fokusse besitzen und somit zusammen mit den aus den Vorverarbeitungsschritten (siehe Abschnitt 3.3) stammenden Informationen über die einzelnen Textstellen eine Hierarchie bilden, bietet sich ebenso für die Repräsentation eine hierar-

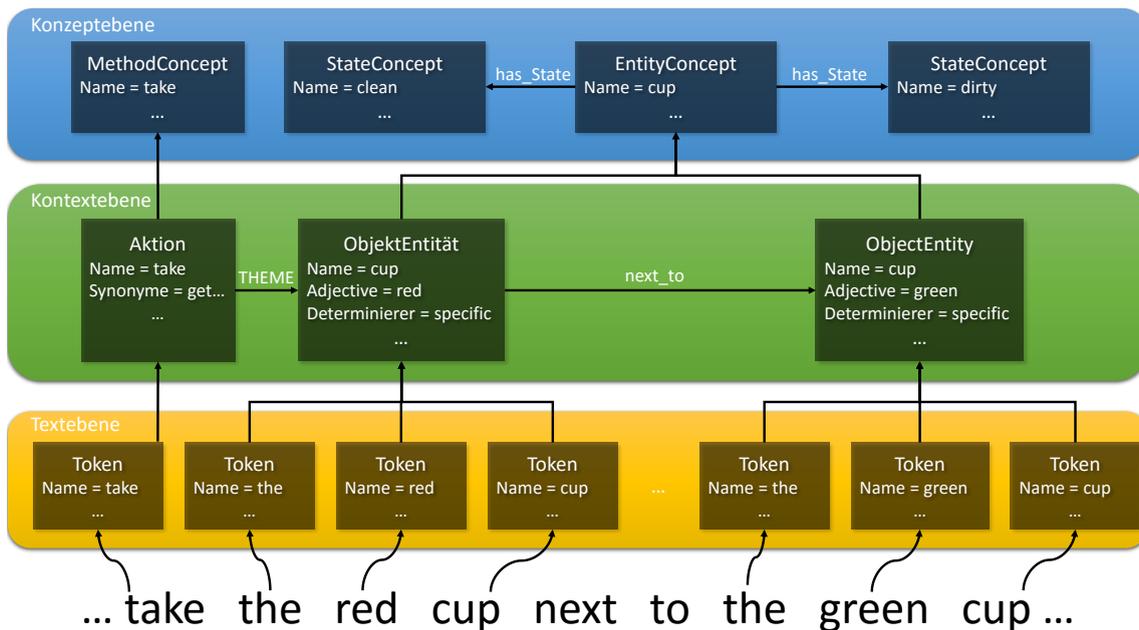


Abbildung 5.9.: Repräsentationsebenen des Kontextes

chischer Aufbau an. So lassen sich die Informationen über die Kontextindividuen, welche eine erste Abstraktion der Textstellen darstellen, oberhalb der einzelnen Textstellen als Knoten in den Graph einbauen. Dies bietet die Möglichkeit die meisten der Informationen durch Attribute der Knoten zu repräsentieren und die bereits im Graph als Token-Knoten dargestellten Textstellen über Kanten als Referenz für die Kontextindividuen zu nutzen. Dies ermöglicht es jederzeit die eigentlichen Textstellen zu ermitteln, die ein Kontextindividuum gebildet haben. Außerdem bietet der Gebrauch von Kanten die Möglichkeit, ohne eine Veränderung der Knoten, die Referenzen anzupassen, sollten fehlerhafte Zuordnungen entdeckt werden.

Die Informationen über die Konzepte stellen wiederum eine weitere Abstraktion der Kontextindividuen dar und bilden somit eine weitere Hierarchieebene von Graph-Knoten. Die konkreten Kontextindividuen zu den Konzepten lassen sich dabei ebenfalls mittels Kanten zuordnen, was auch hier den Vorteil einer leichten Anpassung der Bezüge bietet. Diese Form der Repräsentation erlaubt es nämlich, die einem Konzept zugeordneten Kontextindividuen jederzeit mit einer weiteren Kante zu erweitern oder durch löschen einer Kante zu verändern ohne das das eigentliche Konzept repräsentiert durch einen Konzeptknoten angefasst werden müsste. Dasselbe gilt für die Beziehungen zwischen den Kontextindividuen, wie semantische Rollen oder Ortsbeziehungen aber auch für den Aufbau der Beziehungen zwischen Konzepten, wie z.B. über- und untergeordnete Konzepte. Dieser Entwurf führt zu dem in Abbildung 5.9 dargestellten Aufbau der Kontextinformationen.

5.3. Korreferenzanalyse

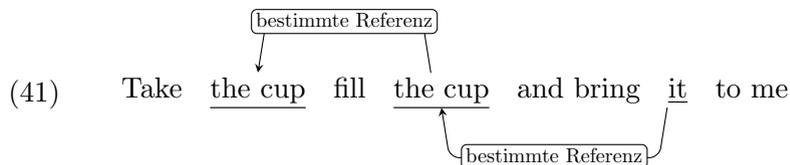
Die Korreferenzanalyse beschäftigt sich mit der Auflösung von Referenzen in natürlich-sprachlichen Texten. Mit Referenzen bezeichnet man in der Sprachwissenschaft Ausdrücke, die auf einen anderen Ausdruck oder eine Entität verweisen. Hierbei wird unterschieden zwischen dem Ausdruck, der die Referenz darstellt, dem sogenannten **referierenden Ausdruck** und der Entität auf die die Referenz verweist, dem **Referenten** (siehe Abschnitt 2.4.1).

Die Aufgabe der Korreferenzanalyse ist es also natürlichsprachliche Referenzen zu identifizieren und den Referenten eines jeden referierenden Ausdrucks zu ermitteln. Diese Aufgabe

ist essentiell für die Umsetzung von natürlicher Sprache in Quelltext, da ohne ein Wissen über diese Paarungen nicht bestimmt werden kann, welche Nennungen auf die gleiche Entität zeigen. Ebenso benötigt man diese Paarungen, um überhaupt eine dargestellte Entität für Formulierungen wie Pronomen bestimmen zu können. Hätte man diese Informationen bei der Umsetzung nicht zur Verfügung, wären die Bezugsobjekte für ausgeführte Aktionen nicht eindeutig bestimmbar, was zu fehlerhaftem Quellcode führen würde.

Referenzen können dabei in unterschiedlichen Formen auftreten. Zunächst lassen sich hierbei bestimmte und unbestimmte Referenzen unterscheiden. Unbestimmte Referenzen bilden eine Referenz auf eine Entität ab, die bisher nicht im Diskurs erwähnt wurde und werden zumeist mit referierenden Ausdrücken, die einen unbestimmten Artikel beinhalten gebildet. Bestimmte Referenzen hingegen verweisen auf bereits im Diskurs erwähnte Entitäten oder aber dem an der Konversation Teilhabenden, durch z.B. seine Wahrnehmung der Umgebung, bekannte Entitäten. Unbestimmte Referenzen können dabei nur mittels unbestimmter Nominalphrasen gebildet werden, bestimmte Referenzen hingegen sowohl durch bestimmte Nominalphrasen, Pronomen als auch Namen. Den Referent einer unbestimmten Referenz stellt somit die Entität dar, welche die unbestimmte Nominalphrase abbildet.

Um den Referenten einer bestimmten Referenz zu identifizieren benötigt man hingegen eine Abbildung vom betrachteten referierenden Ausdruck zu dem referierenden Ausdruck, welcher die letzte vorherige Nennung der Entität darstellt. Dieser referierende Ausdruck kann dabei wiederum eine Abbildung zu einer weiteren vorherigen Nennung besitzen oder aber es handelt sich bei den Ausdruck bereits um die erste Nennung der Entität im Diskurs. Ist dies der Fall so besitzt der referierende Ausdruck keine weiteren Abbildungen mehr sondern bildet den Referenten direkt ab. Durch diese Abbildungen ergeben sich Korreferenzketten zwischen den zusammengehörigen referierenden Ausdrücken, die damit verdeutlichen, welche Ausdrücke zusammen den gleichen Referenten abbilden.



So bilden die beiden Ausdrücke „the cup“ sowie der Ausdruck „it“ in Beispiel 41 eine Korreferenzkette, die als gemeinsamen Referenten die Entität *cup* besitzt. Diese ist bedingt durch die bestimmte Nennung des ersten Ausdrucks den Konversationspartnern bereits bekannt. Die Aufgabe solche Korreferenzketten zu finden ist Ziel der Korreferenzanalyse, denn ohne ein Wissen über zusammengehörige Ausdrücke und somit Nennungen einer gleichen Entität ist kein vollständiges Verständnis komplexer Aussagen möglich.

5.3.1. Stand der Forschung

Die Ansätze zur Lösung der Korreferenzanalyse, welche in der aktuellen Forschung betrachtet werden, beschäftigen sich zum großen Teil mit der Auflösung von Referenzen in geschriebener Sprache. Zumeist werden Merkmalsüberprüfungen eingesetzt, die mittels maschinellem Lernen auf großen Textkorpora erzeugt wurden. So setzt z.B. das *Stanford Deterministic Coreference Resolution System* (siehe Abschnitt 2.4.2) auf mehrere nach ihrer Präzision nacheinander ausgeführte Siebe, welche in jedem Schritt ein anderes Merkmal mittels eines gelernten Modells überprüfen.

Allerdings erzielen viele dieser Ansätze durch ihre Ausrichtung auf geschriebene Sprache nur mäßige Ergebnisse auf gesprochener Sprache, da sie auf Merkmale setzen, die in

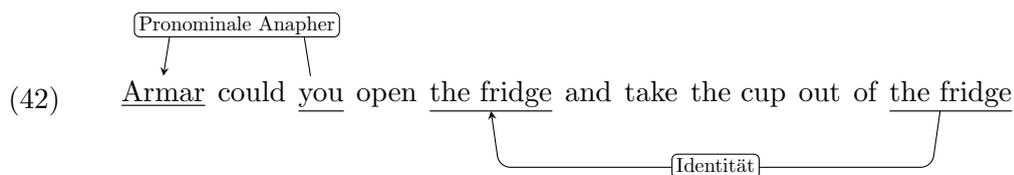
gesprochener Sprache nicht immer korrekt vorhanden sind. So setzt der Stanford Korreferenzen auf Syntaxbäume, welche ohne Satzzeichen und mit Wortwiederholungen nur sehr schwierig korrekt bestimmbar sind.

Die Ansätze die bisher für die Korreferenzanalyse von gesprochener Sprache durchgeführt wurden, beschäftigen sich zumeist nur mit der Auflösung von Referenzen gebildet durch Pronomen, wie bei Strube und Müller [SM03] sowie Tetreault und Allen [TA06] für Dialoge. Da aber für die Umsetzung von gesprochener Sprache in Quelltext auch Referenzen aufgelöst werden müssen, die nicht durch Pronomen gebildet werden und sich Dialoge in der Formulierungen von Anweisungen unterscheiden wird für die Umsetzung der Korreferenzanalyse dieser Arbeit ein anderer Ansatz benötigt.

Hierfür bietet sich das in der Kontextanalyse erstellte Kontextwissen an, welches bereits viele der Merkmale der genannten Arbeiten enthält und darüber hinaus das Potenzial bietet durch die Anreicherung von Wissen über die einzelnen Kontextindividuen eine detailliertere Betrachtung der möglichen Referenzen zu ermöglichen. Daher ist der Ansatz dieser Arbeit eine kontextbasierte Korreferenzanalyse von in gesprochener Sprache gegebenen Anweisungen durchzuführen. Da keine großen Korpora für Anweisungen in gesprochener Sprache existieren und der bisherige Korpus des Projekts nicht ausreicht um auf ihm maschinelle Lernverfahren anzuwenden, wird eine regelbasierte Umsetzung der kontextbasierten Korreferenzanalyse durchgeführt.

5.3.2. Kontextbasierte Korreferenzanalyse

Für die Umsetzung der kontextbasierten Korreferenzanalyse müssen also zunächst die Eigenschaften der Referenzen auf die Informationen aus der Kontextanalyse übertragen werden. Es lässt sich dabei feststellen, dass unbestimmte Referenzen nur von Objektentitäten mit unbestimmtem Artikel gebildet werden können, bestimmte Referenzen hingegen von allen anderen Formen von identifizierten Entitäten. Insbesondere stellen die identifizierten Entitätsobjekte der Kontextanalyse bereits Referenten der referierenden Ausdrücke im Text dar, denn sie bilden die beschriebene abstrakte Entität ab. Unterscheiden lässt sich hierbei also noch, ob die Entitätsobjekte eine Abbildung einer noch nicht im Diskurs aufgetretenen Entität darstellen und somit bereits die eigentliche Referenz auf die Entität darstellen oder ob sie nur eine wiederholte Nennung oder wie im Fall eines Pronomens ein Platzhalter für die eigentliche Entität sind. Somit besteht die in dieser Arbeit durchgeführte kontextbasierte Korreferenzanalyse aus der Auflösung der Referenzen der Entitätsobjekte, die dem letzteren Fall angehören, auf die eigentlich erste Nennung dieser Entität und damit die Bestimmung der Korreferenzketten zu einer abstrakten Entität.



Des Weiteren lassen sich diese betrachteten Referenzen in zwei Kategorien einteilen. Einerseits Referenzen gebildet durch Pronomen, wie *you* in Beispiel 42 und andererseits Referenzen gebildet durch erneute Nennungen einer bereits erwähnten Entität, im Folgenden Identitätsabbildung genannt, wie die Referenz der letzteren *fridge* Nennung zur vorherigen. Erstere stellen einen Platzhalter einer Entität dar, der nicht so spezifisch ist wie die wiederholte Nennung und besitzen in jedem Fall¹ einen referierenden Ausdruck im vorangegangenen Diskurs, welcher die ersetzte Entität näher beschreibt. Letztere hingegen

¹Außer in Sonderfällen wie z.B. *it* in „it is raining“ wie in Abschnitt 2.4.1 beschrieben

bilden nur dann eine Korreferenzabbildung zu einem anderen Ausdruck, falls eine vorherige Nennung der gleichen Entität existiert. Falls diese nicht existiert stellen sie selbst bzw. die Entität, die sie beschreiben, den Referenten dar. Die erste Nennung „the fridge“ in Beispiel 42 ist also der referierende Ausdruck, der auf die Entität *fridge* verweist. Diese Form wird im Folgenden zur Vereinfachung selbst Referent genannt, da sie als erstes im Diskurs auf den Referent verweist. Außerdem können Pronomen neben Anaphern auch Kataphern bilden. Da diese allerdings generell im englischen Sprachgebrauch sehr selten sind und zumeist in Beschreibungstexten und nicht in Anweisungen auftreten werden diese in dieser Arbeit vernachlässigt und sich auf die Auflösung von mit Pronomen gebildeten Anaphern konzentriert.

Somit lassen sich für das in Kapitel 5 definierte Ziel der kontextbasierten Korreferenzanalyse zwei Schritte identifizieren: Die Auflösung pronominaler Anaphern sowie die Identitätsauflösung von Entitätsidentitäten.

5.3.3. Pronominale Anapherauflösung

Für die Auflösung pronominaler Anaphern muss zunächst analysiert werden, welche Arten von Pronomen in englischsprachigen Anweisungen auftreten und für die Umsetzung in Quelltext relevant sind. Die wohl am häufigsten auftretende Form von Pronomen sind Personalpronomen, denn sie können anstatt einer wiederholten Nennung einer bereits in den Kontext eingeführten Entität im Textverlauf verwendet werden. Sie treten gerade in Anweisungen vermehrt auf, da in solchen Formulierungen häufig mehrere aufeinanderfolgende Aktionen mit derselben Entität durchgeführt werden und durch die Nutzung von Personalpronomen die Formulierungen kürzer werden. Damit stellen sie auch für die Korreferenzanalyse die wichtigste Form von Referenzen dar, denn ohne eine Möglichkeit diese aufzulösen lässt sich keine korrekte Umsetzung in Quelltext durchführen.

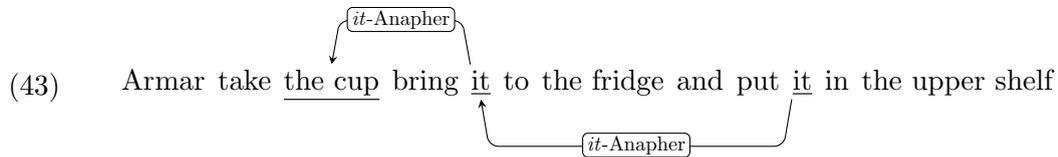
Zusätzlich zu den Personalpronomen können sowohl Reflexiv-, Demonstrativ- als auch Possessivpronomen Referenzen erzeugen. Da allerdings Demonstrativpronomen zumeist in Verbindung mit einer zeigenden Handlung, wie z.B. „this cup and not that one“ mit gleichzeitigem Zeigen auf den Gegenstand, auftreten und eine Wahrnehmung der Umwelt im *PARSE*-Projekt nicht betrachtet wird, können diese nicht vollständig untersucht werden, da keine Informationen über die Objekte auf die gezeigt wird existiert. Daher werden die Demonstrativpronomen in der Betrachtung dieser Arbeit ausgelassen. Reflexiv- und Possessivpronomen hingegen lassen sich mit einigen wenigen Veränderungen ähnlich zu den Personalpronomen auflösen. Daher betrachten wir zunächst die Auflösung von Personalpronomen.

Um eine möglichst vollständige Betrachtung der Auflösung, von durch Personalpronomen erzeugten Anaphern, durchführen zu können muss zunächst geklärt werden, welche Eigenschaften Referenten von Personalpronomen generell erfüllen müssen und welche Eigenschaften spezifisch für das jeweilige Personalpronomen bzw. den Personalpronomentyp sind.

5.3.3.1. Vorangehende Entitäten

Zunächst lässt sich festhalten, dass durch die Betrachtung von Anaphern nur diejenigen Entitäten als Referenten infrage kommen, die bereits vor dem betrachteten referierenden Ausdruck im Diskurs genannt wurden. Auch der Mensch hat bei der Auflösung von Referenzen, zum Zeitpunkt des Hörens des referierenden Ausdrucks, nur Informationen über die zuvor genannten Entitäten. Ebenso wird bei der Textproduktion darauf geachtet auf bereits erwähnte Entitäten zu verweisen um klare und möglichst eindeutige Sätze zu formulieren. Personalpronomen können hierbei sowohl auf Entitäten verweisen, die selbst keine

Pronomen darstellen, als auch auf Pronomen der gleichen Art. In dieser Arbeit sind zwei Pronomen, immer dann gleichartig, wenn sie mit dem gleichen Wort gebildet werden.



Dabei stellen Erstere den eigentlichen Referenten dar, Letztere hingegen bilden eine Korreferenzkette bis zur ersten nicht Pronomenentität, welche daraufhin für die gesamte Kette als Referent dient. So dient die Entität *cup* in Beispiel 43 als Referent der Korreferenzkette bestehend aus den beiden *it*-Pronomen. Hierbei gilt, dass keine Entität als Kandidat infrage kommt, die im Text vor einem Pronomen derselben Art steht. Würde dies nämlich der Fall sein könnten die beiden Pronomen auf unterschiedliche Entitäten zeigen, was aber durch dieselbe Art der Pronomen ausgeschlossen ist.

5.3.3.2. Numerus

Der Numerus stellt eine weitere generelle Eigenschaft für Referenten-Kandidaten dar: Der Numerus, den das Personalpronomen widerspiegelt, muss mit dem Numerus des möglichen Kandidaten für den Referenten übereinstimmen oder es muss im Falle einer mehrzahligen Referenz eine Konjunktionsbeziehung zwischen mehreren Entitäten existieren, die dann zusammen als Referent infrage kommen.

(44) Armar go to the kitchen and get the two oranges, they are in the fridge

(45) Armar open the fridge and the dishwasher then close them both again

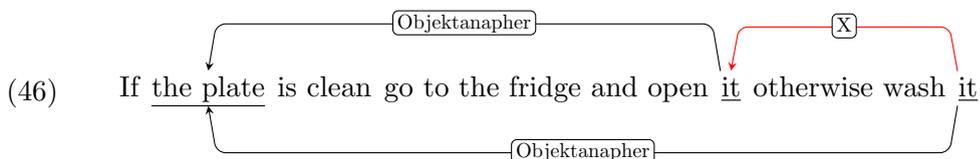
So kann eine mehrzahlige Referenz entweder auf eine Entität im Plural verweisen, wie es bei *they* und *oranges* in Beispiel 44 der Fall ist, aber nicht auf die Entität *kitchen* die im Singular steht, oder aber wie *them* in Beispiel 45 auf zwei Entitäten, die mit einer Konjunktionsbeziehung verbunden sind wie *fridge* und *dishwasher*.

5.3.3.3. Bezugsrahmen

Eine weitere Eigenschaft, die mögliche Kandidaten für Referenten von Personalpronomen erfüllen müssen, ist, dass sie in einer gewissen Nähe zum referierenden Ausdruck beschrieben sein müssen. Wie bereits in Abschnitt 2.4 beschrieben beziehen sich Pronomen in der Regel auf Entitäten, die nicht weiter als ein bis zwei Sätze entfernt stehen. Diese Eigenschaft resultiert daraus, dass bei einer größeren Distanz zumeist, selbst für das menschliche Gehirn, zu viele mögliche Kandidaten in Betracht gezogen werden müssen und damit die Auflösung schwer eindeutig umsetzbar ist. Deshalb beschreiben Menschen zumeist automatisch weiter entfernte Entitäten mit einer erneuten direkten Entitätsreferenz und nicht mit Pronomen. Da durch die betrachtete gesprochene Sprache, keine Informationen über die Satzgrenzen existieren, lässt sich für diese Eigenschaft nur die Instruktionsnummer-Heuristik heranziehen und auf die Satzzahl abbilden. Da die meisten Sätze in der englischen Sprache und vor allem in Anweisungen nicht mehr als drei Verben bzw. Instruktionen enthalten, bietet sich eine Spanne von sechs Instruktionsnummern als valide Grenze für pronominale Referenzen an. Diese Spanne repräsentiert einen Betrachtungshorizont von circa zwei Sätzen.

5.3.3.4. Bedingungen

Im Fall von im Text auftretenden Bedingungen lässt sich eine weitere allgemeine Eigenschaft identifizieren: Beschreibungen, die Bedingungen enthalten, verändern ebenfalls den Bezugsrahmen für Anaphern.



Betrachtet man das Beispiel 46 so existieren dort die beiden Pronomen *it* in jeweils unterschiedlichen Bedingungs Zweigen. Die beiden Bedingungs Zweige sind dabei logisch und semantisch getrennte unabhängige Ausführungszweige. Dadurch beziehen sich alle in ihnen enthaltenen Entitäten jeweils auf die Entitäten in ihrem Zweig oder auf die Bedingung selbst. Hierdurch existieren also keine Korreferenzketten zwischen den Zweigen. Daher kommen referierende Ausdrücke, aus der jeweils anderen Alternative, nicht für den Bezugsrahmen der Auflösung der betrachteten Alternative infrage. Somit existiert die rot eingezeichnete Referenz im Beispiel nicht, sondern nur die Referenz des zweiten *it* auf die Entität *plate* selbst.

5.3.3.5. Gleiche Aktion

Eine weitere für Referenten von Personalpronomen allgemeine Eigenschaft lässt sich aus der Tatsache ableiten, dass Personalpronomen als Platzhalter für die eigentlich gemeinte Entität benutzt werden und diese vollständig im Satzbau ersetzen. Hierdurch lässt sich ableiten, dass keine Entität als Referent infrage kommt, die an derselben Aktion teilnimmt, wie das aufzulösende Pronomen.

(47) Get the cup and put the water in it

(48) John washes himself

(49) Close the water bottle with its cap

Wäre dies der Fall, könnte ein und dieselbe Entität zwei unterschiedliche semantische Rollen innerhalb einer Aktion annehmen, was wiederum der Logik der semantischen Rollen widerspräche. So kann die Entität *water* nicht als Referent für das Pronomen *it* in Beispiel 47 infrage kommen. Anders verhält sich dies bei Reflexivpronomen, welche explizit für einen solchen Fall gedacht sind (siehe Beispiel 48) oder auch bei den Possessivpronomen, welche in Verbindung mit einer anderen Entität stehen (siehe Beispiel 49). Bei diesen beiden Formen gilt die Eigenschaft nicht.

5.3.3.6. Typübereinstimmung

Generell lassen sich die Personalpronomen in vier Gruppen einteilen, die unterschiedliche Arten von Entitäten referenzieren können. Zunächst sind dies die Pronomen der ersten Person Singular und Plural, welche sich immer auf den Sprecher oder eine Gruppe von Personen beziehen, denen der Sprecher angehört und im Folgenden als *Sprecherpronomen* bezeichnet werden. Da eine sprechende Person in der Regel auf sich selbst nur mit einem Pronomen verweist, lässt sich für diese Anaphern keine im Text existierende Entität bestimmen, die nicht ein Pronomen darstellt. Daher ist es sinnvoll eine virtuelle Sprecherentität einzuführen, die als Referent des ersten im Text auftretenden Sprecherpronomen dient und es damit ermöglicht ebenfalls eine vollständige Korreferenzkette für Sprecherpronomen aufzubauen.

Pronomen- typ	gebildet durch	mögliche Referenz auf
Sprecher	1. Person Singular: <i>i, me, myself, mine, my</i> 1. Person Plural: <i>we, us, ourself, ourselves, ours, our</i>	Sprecherpronomen & virtuelle Sprecherentität
Subjekt	2. Person: <i>you, yourself, yourselves, your, yours</i> 3. Person Singular: <i>he, him, himself, his</i> (Maskulin) 3. Person Singular: <i>she, her, herself, hers</i> (Feminin)	Subjektpronomen (gleicher Art) & Subjektentitäten
Objekt	3. Person Singular: <i>it, itself, its</i> (Neutrum)	Objektpronomen & Objektentitäten
Gruppen	3. Person Plural: <i>they, them, themselves, themselves, their, theirs</i>	Gruppenpronomen & mehrzählige Subjekt- oder Objektentitäten

Tabelle 5.2.: Typen von Personalpronomen und ihre möglichen Bezugsobjekte

Pronomen der zweiten Person Singular und Plural, sowie der dritten Person Singular, in männlicher und weiblicher Form, beziehen sich auf Subjektentitäten im Kontext und werden im Folgenden als *Subjektpronomen* bezeichnet werden. Referenten der Pronomen der zweiten Person werden direkt vom Sprecher angesprochen. Daher können sie sich nur auf handelnde Entitäten beziehen, welche im Kontext als Subjektentitäten abgebildet werden. Die dritte Gruppe besteht aus den Pronomen der dritten Person Singular im Neutrum, welche sich auf eine Objektentität beziehen. Diese werden im weiteren Verlauf *Objektpronomen* genannt.

Die vierte und letzte Pronomengruppe stellen die Pronomen der dritten Person Plural dar, welche sowohl auf Subjekt- als auch Objektentitäten in der Mehrzahl oder über Konjunktionen verbundene Entitäten referenzieren können. Pronomen dieser Gruppe werden nachfolgend *Gruppenpronomen* genannt.

Eine Übersicht über die verschiedenen Pronomentypen, die Pronomen, die sie bilden, sowie die Entitäten auf die sie referenzieren können wird in Tabelle 5.2 gegeben. Bei den bildenden Pronomen sind hierbei bereits die Reflexiv- und Possesivpronomen miteinbezogen, auf die in Abschnitt 5.3.3.12 genauer eingegangen wird.

Zusätzlich zu den allgemeinen Eigenschaften, welche die Referentenkandidaten erfüllen müssen, lassen sich ebenfalls für den Pronomentyp spezifische Eigenschaften identifizieren, welche im Folgenden genauer erläutert werden.

5.3.3.7. Geschlecht

Für die Auflösung von Subjektpronomen lässt sich die Geschlechtsinformation, welche sowohl die Pronomen, durch ihre Ausprägung, als auch die Subjektentitäten enthalten, als Eigenschaft identifizieren. So kann sich z.B. ein feminines Pronomen der dritten Person Singular nur auf eine Entität beziehen, die ebenfalls feminin ist.

(50) Mary and John are in the kitchen can you bring her an empty cup

So bezieht sich das Pronomen *her* in Beispiel 50 eindeutig auf die feminine Subjektentität *Mary* und nicht auf die Subjektentität *John*, da diese maskulin ist. Im Fall eines Sub-

jektpronomen, gebildet durch die 2. Person, bleibt die Geschlechtsinformationen allerdings ohne Aussage, da die Referenz hierbei unabhängig vom Geschlecht ist.

Diese Eigenschaft lässt sich nicht auf die anderen Pronomentypen übertragen, da diese keine Information über das Geschlecht enthalten oder wie im Fall der Objektpronomen, durch die Definition der Objektentitäten (siehe Abschnitt 5.2.2), bereits implizit abgedeckt wird.

5.3.3.8. Ausführbarkeit von Aktionen

Für die Auflösung von Referenzen, die sich auf Objektentitäten beziehen können (Objekt- und Gruppenpronomen), bietet der aufgebaute Kontext eine weitere Eigenschaft zur Validierung möglicher Kandidaten. Durch die Anbindung an die Domänenontologie (siehe Abschnitt 5.2.3.3) existieren Informationen über die möglichen Zustände, die eine Objektentität annehmen kann, sowie Informationen über die Zustandsübergänge, die eine Aktion auslöst. Außerdem enthält der Kontext die in Abschnitt 5.2.3.4 beschriebenen Informationen über initiale Zustände der Objektentitäten und Pronomen. Hiermit lässt sich überprüfen, ob ein Pronomen, welches eine Proto-Patient-Rolle (siehe Abschnitt 2.3.4) in einer Aktion einnimmt, also von der Aktion „behandelt“ wird, einem Zustandsübergang ausgesetzt ist, der im Zustand des Kandidaten durchführbar ist. Dies ist immer dann der Fall, wenn die Objektentität oder das gleichwertige Pronomen, welches als Kandidat infrage kommt, sich nicht im gleichen Zustand befindet, wie der Zustand in den übergegangen wird.

(51) Armar bring me the full cup next to the empty glass and pour me from it

Betrachtet man also das Pronomen *it* in Beispiel 51 unter der Annahme, dass bekannt ist, dass die Entität *cup* sich im Zustand *full* und die Entität *glass* sich im Zustand *empty* befindet, lässt sich mit der Information, dass die Methode *pour* einen Zustandsübergang in den Zustand *empty* erzeugt die Entität *glass* als Referent für das Pronomen ausschließen. Diese Eigenschaft basiert auf dem einfachen binären Zustandskonzept, welches in der Domänenontologie abgebildet ist. Eine Überprüfung dieser Eigenschaft hängt stark von der Güte, der in Abschnitt 5.2.3.4 beschriebenen initialen Zustandsbestimmung, und der Verfügbarkeit der Zustands- und Zustandsübergangsinformationen in der Domänenontologie ab.

5.3.3.9. Ortsbeziehungen

Eine weitere Möglichkeit die Kandidaten für Referenzen einzugrenzen, bieten die in der Kontextanalyse für Entitäten gesammelten Ortsbeziehungen. Diese stellen eine Form von Entitäten in natürlicher Sprache dar, die eine Art Nebenordnung zweier Entitäten beschreibt. So stellt das Ziel der Ortsbeziehung im Kontext zumeist nur eine genauere Beschreibung der vorherigen Entität dar. Dies lässt sich in der Korreferenzanalyse dafür nutzen, dass wenn immer zwei gleichartige Entitäten, die eine Ortsbeziehung zwischen sich bilden, als Kandidaten für eine Referenz infrage kommen, die mittels der Ortsbeziehung näher beschriebene Entität mit einer höheren Wahrscheinlichkeit den Referenten darstellt. Dies lässt sich daraus schließen, dass die zweite Entität nur als Beschreibung genutzt wird und daher nicht im Ausführungsfokus der Anweisung steht. Eine Ausnahme hierzu bilden Pronomen selbst, die sich im Fall von zwei Pronomen als Kandidaten immer auf das im Text näher stehende gleichartige Pronomen beziehen.

(52) Take the cup next to the popcorn and bring it to me

(53) Put the cup on the table and take the popcorn next to it then bring it to me

(54) Take the cup next to the fridge and put the cup in it

Somit lässt sich über diese Eigenschaft der Kandidat *cup* in Beispiel 52 als wahrscheinlicherer Kandidat als der Kandidat *popcorn* identifizieren. Ein Fall der beschriebenen Ausnahme ist in Beispiel 53 dargestellt, in welchem sich das letzte *it* Pronomen auf das vorherige *it* bezieht, obwohl dieses in einer Ortsbeziehung zur ebenfalls infrage kommenden *popcorn*-Entität steht.

Es gibt allerdings auch Formulierungen in denen diese Eigenschaft nicht korrekt ist, wie in Beispiel 54, in dem trotzdem die beschreibende Entität *fridge* als Referent gemeint ist. Außerdem lässt sich diese Eigenschaft nicht auf Sprecherpronomen anwenden, da Bezüge dieser nur in Form von Pronomen auftreten.

5.3.3.10. Übereinstimmung semantischer Rollen

Eine weitere Eigenschaft, die man für die Auflösung von Objektpronomen benutzen kann, stellt die semantische Rolle, die das Pronomen einnimmt, dar. Diese Eigenschaft resultiert aus der Beobachtung, dass häufig Referenten und referierende Ausdrücke in der gleichen semantischen Rolle in unterschiedlichen Aktionen auftreten. Gerade bei Objektpronomen lässt sich dieses Phänomen feststellen.

(55) Armar get the cup from the table and put it into the dishwasher.

Die Entität *cup* nimmt in Beispiel 55 die semantische Rolle **Theme** in der Aktion *get* ein. Die Entität *table* hingegen die Rolle **Source**. Da das Pronomen *it* ebenfalls die Rolle **Theme** in der Aktion *put* einnimmt, lässt sich wegen Rollengleichheit die Entität *cup* als passenderer Kandidat für eine Referenz bestimmen. Diese Beobachtung ist, aber ebenso wie die Ortsbeziehung zuvor, nur in vielen aber nicht allen Fällen zutreffend und bietet daher wiederum nur einen Hinweis auf einen wahrscheinlicheren Kandidaten. Aus dieser Eigenschaft resultiert auch die Einordnung dieser Beobachtung, als nur für Objektpronomen zutreffend, da sie zwar auch für andere Pronomen zutreffend sein kann, aber diese zumeist bereits über die anderen Eigenschaften sicher erkannt werden können. Somit würde sie bei den anderen Pronomen keinen großen Mehrwert und einige falsch-positive Ergebnisse erzeugen.

5.3.3.11. Abstand zum referierenden Ausdruck

Eine weitere Eigenschaft die nicht nur für alle Pronomen, sondern allgemein für alle Korreferenzen besteht ist, dass eine Korrelation zwischen dem Abstand eines Kandidaten zum referierenden Ausdruck und der Wahrscheinlichkeit, dass dieser als Referent infrage kommt, existiert. Diese Eigenschaft leitet sich daraus ab, dass beobachtet werden kann, dass nähere Ausdrücke im Allgemeinen den wahrscheinlicheren Kandidaten darstellen. Diese Eigenschaft kommt immer dann zum Tragen, wenn es mehr als einen möglichen und gleich passenden Kandidaten für eine Referenz gibt. Da in diesem Fall alle Kandidaten gleich wahrscheinlich sind, ist der Ausdruck, welcher zum referierenden Ausdruck am nächsten im Text auftaucht, als Referent zu betrachten. Gerade bei der Auflösung von pronominalen Anaphern spielt diese Eigenschaft eine große Rolle, denn hier existieren häufiger mehrere gleich wahrscheinliche Kandidaten.

(56) Please bring me the popcorn and take the cup it is on the table

So lässt sich mit den bisher identifizierte Eigenschaften weder die Entität *popcorn*, noch die Entität *cup*, als eindeutiger Referent für den referierenden Ausdruck *it* in Beispiel 56 identifizieren. Da nun aber die Entität *cup* näher zum referierenden Ausdruck steht, stellt diese auch den wahrscheinlicheren Referenten dar.

5.3.3.12. Reflexiv- und Possessivpronomen

Wie bereits erwähnt lässt sich das referierende Verhalten von Reflexiv- und Possessivpronomen an dem von Personalpronomen ausrichten. So erfüllen Referenzen von Reflexivpronomen die gleichen Eigenschaften, wie die Referenzen von Personalpronomen. Zusätzlich besitzen sie allerdings einen kleineren Bezugsrahmen als Personalpronomen, denn sie bilden Bezüge von Entitäten auf sich selbst ab. Also Formulierungen in denen sowohl das Subjekt, als auch das Objekt einer Aktion, dieselbe Entität darstellen. Somit lässt sich das Vorgehen für Personalpronomen auf Reflexivpronomen übertragen, indem eine Überprüfung, ob der referierende Ausdruck und der Referentenkandidat in der gleichen Instruktion auftreten, hinzugefügt wird.

Bei Possessivpronomen handelt es sich hingegen um beschreibende Zusatzinformationen zu einer Entität, weshalb diese auch einer Entität zugeordnet sind. Auch hier gelten in vielen Fällen die gleichen Eigenschaften wie bei Personalpronomen. Allerdings ist die Auflösung enger mit der näher beschriebenen Entität verknüpft und erlaubt daher weitere Eigenschaften zu ermitteln, wie z.B. ein genaueres Abgleich der Teil-Ganzes-Beziehungen des Kandidaten, und der näher beschriebenen Entität zuzuordnen. Dies wäre mit einem alleinstehenden Personalpronomen nicht möglich.

(57) Please go to the fridge open its door and take the cup then close its door again

So lässt sich in Beispiel 57 die Teil-Ganzes Beziehung zwischen *fridge* und *door* dazu nutzen, die Referenzen der beiden *its* korrekt aufzulösen. Da Possessivpronomen Entitäten zugeordnet sind und diese Entitäten wiederum untereinander Identitätsabbildungen aufbauen können, ist es sinnvoll Referenzen von Possessivpronomen neben Identitätsreferenzen zu betrachten, da nicht in allen Fällen beide Referenzen übereinstimmen müssen. So beziehen sich die beiden *its* im Beispiel auf die Entität *fridge*, aber die Identitätsabbildung der hinteren *door*-Entität bezieht sich auf die vorherige *door*-Entität. Diese Eigenschaft lässt sich allerdings nur für Possessivpronomen der Typen Gruppe und Objekt umsetzen. Dies liegt daran, dass nur diese Objektentitäten als Referenten besitzen können, welche Informationen über Teil-Ganzes-Beziehungen enthalten. Außerdem lassen sich die Eigenschaften gleiche Aktion (siehe Abschnitt 5.3.3.5) und Übereinstimmung der semantischen Rollen nicht auf Possessivpronomen übertragen, da sie durch die Bindung an eine Entität nicht für sie erfüllt sein können.

5.3.4. Entwurf der pronominalen Anapherauflösung

Für die Umsetzung der Auflösung von pronominalen Anaphern lässt sich also festhalten, dass es sowohl Eigenschaften gibt, die jeder Referentenkandidat eines Pronomens erfüllen muss und somit ein hartes Ausschlusskriterium für die sie nicht erfüllende Kandidaten darstellen. Die weiteren Eigenschaften geben hingegen vielmehr einen Hinweis über eine wahrscheinlichere Paarung. Zur ersten Gruppe zählen die Eigenschaften vorangehende Entitäten, Numerus, Bezugsrahmen, Bedingungen, gleiche Aktion und Geschlecht. Letztere Gruppe beinhaltet Eigenschaften, die bedingt durch ihre Konzeption, nicht in allen Fällen, sicher Kandidaten ausschließen können, wie die Ausführbarkeit von Aktionen, die Ortsbeziehungen und die Übereinstimmung der semantischen Rollen. Die Ausführbarkeit von Aktionen (siehe Abschnitt 5.3.3.8) ist deshalb unsicher, da die Erkennung der Zustände fehlerhaft oder unvollständig sein kann, was wiederum zu fehlerhaften Einschätzungen führen würde. Die Ortsbeziehungen kommen deshalb nicht als hartes Auswahlkriterium infrage, weil es Referenten gibt bei denen die Eigenschaft nicht erfüllt ist (siehe Abschnitt 5.3.3.9). Gleiches gilt für die Übereinstimmung der semantischen Rollen (siehe Abschnitt 5.3.3.10). Da diese Eigenschaften aber trotzdem in vielen Fällen die Kandidatenmenge gut einschränken möchte man sie als unsichere Eigenschaften trotzdem in der

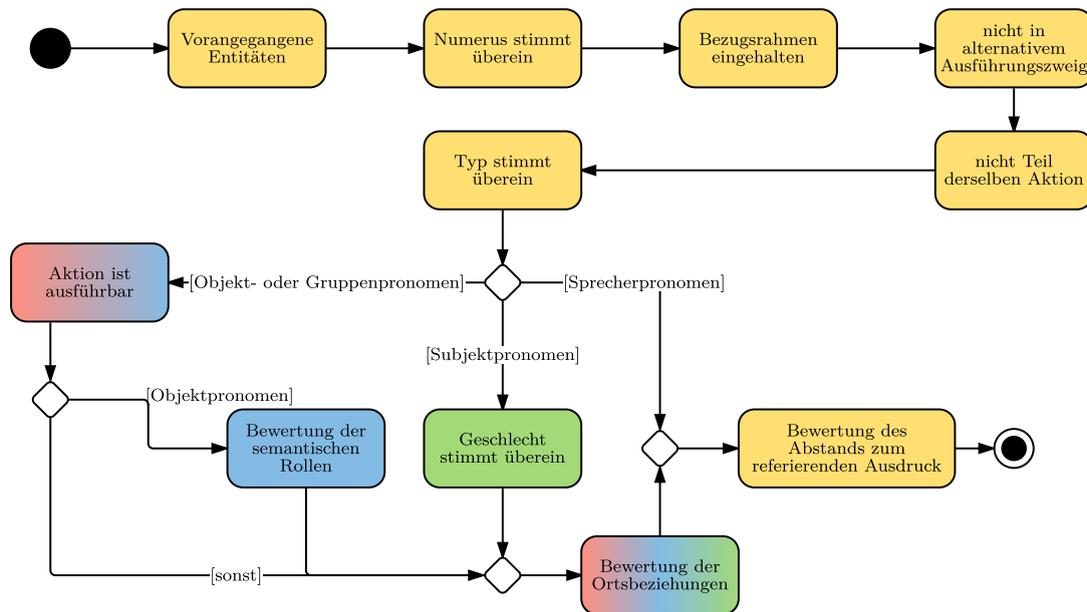


Abbildung 5.10.: Übersicht über den Ablauf der pronominalen Anapherauflösung

Auflösung betrachten. Außerdem existieren Eigenschaften, die erst dann sinnvoll angewendet werden können, wenn andere Eigenschaften überprüft wurden, wie es z.B. beim Abstand (siehe Abschnitt 5.3.3.11) der Fall ist.

Insgesamt lassen sich die Eigenschaften, unter dem Aspekt der Fehleranfälligkeit in der Überprüfung, in einer Reihenfolge anordnen, in der möglichst früh viele sicher nicht korrekte Kandidaten ausgefiltert werden und damit die Fehlerrate der unsicheren Eigenschaften verringert wird.

Betrachtet man zunächst das Vorgehen im Fall von Personalpronomen, so ergeben sich die in Abbildung 5.10 dargestellten Schritte. Es lassen sich also zunächst die sicheren, für alle Pronomentypen allgemeinen, Eigenschaften überprüfen und damit zuerst die Kandidaten ausschließen, die keine vorangegangene Entität darstellen (siehe Abschnitt 5.3.3.1), deren Numerus nicht übereinstimmt (siehe Abschnitt 5.3.3.2), die mehr als 6 Instruktionen entfernt auftreten (siehe Abschnitt 5.3.3.3), die sich in einem zum aktuellen Bezugspunkt alternativen Aussagenzweig befinden (siehe Abschnitt 5.3.3.4) und Entitäten, die Teil der gleichen Aktion wie der referierende Ausdruck sind (siehe Abschnitt 5.3.3.5). Da diese Eigenschaften alle harte Auswahlkriterien darstellen kann die Kandidatenmenge hierdurch deutlich verkleinert werden. Außerdem handelt es sich bei diesen Eigenschaften, um für alle Pronomentypen gleich durchzuführende Überprüfungen, was es erlaubt diese allgemein zu definieren und noch vor einer Unterteilung in Pronomentypen mit jedem Pronomen durchzuführen. Daraufhin schränkt die Überprüfung des Typs die Kandidatenliste weiter ein (siehe Abschnitt 5.3.3.6) und sorgt dafür, dass klar ist, welche weiteren Schritte durchgeführt werden können.

Bei den Sprecherpronomen wird keine weitere Überprüfung durchgeführt, da diese in Anweisungen generell nur als Pronomen auftreten und somit eindeutig identifizierbar sind.

Für Subjektpronomen wird zunächst das Geschlecht überprüft (siehe Abschnitt 5.3.3.7), welche eine weitere harte Ausschlusseigenschaft darstellt, sowie die Ortsbeziehungen ausgewertet (siehe Abschnitt 5.3.3.9), welche nur Wahrscheinlichkeiten bieten.

Im Fall von Objektpronomen lässt sich die Ausführbarkeit der gefundenen Aktion überprüfen (siehe Abschnitt 5.3.3.8), welches entweder eine Aussage über wahrscheinlich nicht

infrage kommende Aktionen zulässt oder zwecks nicht vorhandener Informationen keinen Effekt hat. Daraufhin können die Kandidaten anhand ihrer semantischen Rollen bewertet werden (siehe Abschnitt 5.3.3.10) sowie letztlich die Ortsbeziehungen ausgewertet werden (siehe Abschnitt 5.3.3.9).

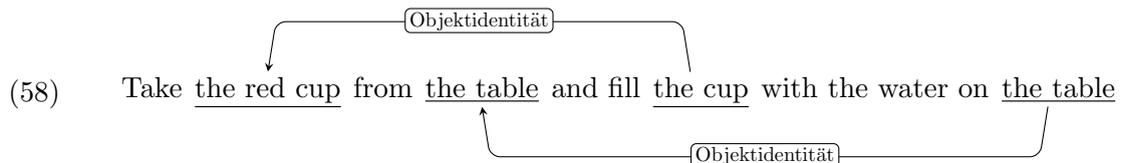
Für Gruppenpronomen kommen die gleichen Schritte wie für Objektpronomen infrage, mit der Ausnahme, dass eine Übereinstimmung der semantischen Rollen für Kandidaten von Gruppenpronomen schwierig und fehleranfällig ist. Daher wird diese ausgelassen.

Abschließend lassen sich dann die, nach allen Überprüfungen übrig gebliebenen, Kandidaten anhand ihrer bisherigen Wahrscheinlichkeit und dem Abstand zum referierenden Ausdruck (siehe Abschnitt 5.3.3.11) bewerten. Daraus ergibt sich dann eine Wahrscheinlichkeitsverteilung der Referenten.

Für die Reflexivpronomen muss hierbei nur der Bezugsrahmen dahingehend angepasst werden, dass nur in der gleichen Instruktion gesucht wird, sowie die Eigenschaft Teil der gleichen Aktion (siehe Abschnitt 5.3.3.5) nicht eingesetzt wird. Die Anpassungen für die Possessivpronomen bestehen darin, dass für die Objekt- und Gruppenpronomen, die nicht mehr geltenden Eigenschaften Aktion ausführbar und Übereinstimmung semantischer Rollen durch die Teil-Ganzes Eigenschaft ersetzt wird (siehe Abschnitt 5.3.3.12).

5.3.5. Identitätsauflösung

Unter Identitätsauflösung wird in dieser Arbeit die Auflösung von Referenzen, gebildet durch eine wiederholte Nennung oder Umschreibung derselben Entität, verstanden. Solche Identitäten können auftreten, müssen aber nicht zwangsläufig vorhanden sein. So kann die Nennung einer Entität zwar potenzielle Referenten besitzen, aber keine dieser Kandidaten bilden die exakt gleiche Entität ab und bilden somit keinen Referenten. Wenn kein potenzieller Referent die Entität abbildet wurde diese zum ersten Mal im Diskurs genannt und stellt damit selbst die Abbildung des Referenten dar. Identitäten können sowohl durch eine Wiederholung des Namens der Entität oder aber über Unter- bzw. Überbegriffe oder Synonyme auftreten.



So bildet sowohl die Wiederholung der *table*-Entität in Beispiel 58 eine Referenz auf die vorher genannte *table*-Entität, als auch die allgemeinere *cup*-Entität auf die spezifischere *red cup*-Entität. Die *red cup*-Entität allerdings hat keinen Referenten, da sie die erste Nennung der Entität im Diskurs darstellt.

Da Identitätsabbildungen immer Wiederholungen einer bereits eingeführten Entität darstellen, können diese keine Kataphern sondern nur Anaphern bilden, was dazu führt, dass insgesamt in dieser Arbeit keine Kataphern betrachtet werden und einige, der für die Auflösung pronominaler Anaphern identifizierten, Eigenschaften sich ebenso auf die Identitätsabbildungen übertragen lassen.

Zunächst lässt sich festhalten, dass Entitäten nur auf Entitäten des gleichen Typs verweisen können. Dies sorgt dafür, dass sich die Identitätsauflösung von Imperativsätzen in zwei Kategorien unterscheiden lässt: Die *Subjektidentitätsauflösung*, welche Referenzen von Subjektentitäten auf Subjektentitäten auflöst und die *Objektidentitätsauflösung*, welche Referenzen von Objektentitäten auf Objektentitäten auflöst. Somit lassen sich die Kandidaten einer möglichen Referenz, je nach betrachteter Entität, direkt auf den jeweiligen Typ einschränken.

Zusätzlich hierzu gibt es noch Eigenschaften, die für alle Formen von Identitätsabbildungen gelten. Die Erste dieser Eigenschaften resultiert aus dem nicht Vorhandensein von Kataphern und ist analog zur in Abschnitt 5.3.3.1 beschriebenen Eigenschaft für Pronomen. Für eine Anapherreferenz kommen nur der betrachteten Entität vorangegangene Entitäten infrage. Zwei weitere analog zu den Eigenschaften der Pronomen identifizierbare Eigenschaften stellen der Numerus (siehe Abschnitt 5.3.3.2) und die Überprüfung alternativer Ausführungszweige (siehe Abschnitt 5.3.3.4) dar. Der Numerus des Kandidaten muss mit dem Numerus der betrachteten Entität übereinstimmen. Außerdem darf der Kandidat nicht in einem alternativen Ausführungszweig zum aktuellen Zweig auftreten.

5.3.5.1. Konzeptübereinstimmung

Eine weitere Eigenschaft lässt sich aus den in der Konzeptualisierung extrahierten Überkonzepten ableiten (siehe Abschnitt 5.2.3.3). Stimmt das extrahierte Überkonzept des möglichen Kandidaten nicht mit dem Konzept der betrachteten Entität überein, existiert mindestens eine von der textuellen Repräsentation unabhängige Eigenschaft, die nicht übereinstimmt. Ist dies der Fall kann es sich nicht um die gleiche Entität handeln, es sei denn es wurde ein Synonym oder Hyperonym benutzt.

- (59) Can you please turn off the television because the sound of the TV annoys me
 (60) Take the lemon juice out of the fridge and fill the cup with the juice

Stellt das Konzept der betrachteten Entität ein Synonym des Konzepts des Kandidaten dar, handelt es sich zwar nicht um dasselbe Konzept, aber um ein Konzept mit gleicher Bedeutung und kann daher ebenso eine Identitätsabbildung erzeugen. Dieser Fall tritt in Beispiel 59 bei den Entitäten *television* und *TV* ein, welche die synonymen Konzepte *Television* und *TV* besitzen.

In Beispiel 60 stellt das Konzept *Juice* der *juice*-Entität das Überkonzept des *LemonJuice*-Konzepts der *lemon juice*-Entität dar. Es handelt sich somit um eine Beschreibung der Entität und stellt ebenfalls eine Identitätsabbildung dar.

Zusätzlich zu diesen allgemeinen Eigenschaften von Identitätsabbildungen besitzen die zwei identifizierten Kategorien weitere für die Kategorie spezifische Eigenschaften. Der Abstand zum referierenden Ausdruck (siehe Abschnitt 5.3.3.11) kann auch bei der Identitätsauflösung als Maß für gegebenenfalls gleich wahrscheinliche Kandidaten eingesetzt werden. Auch hier gilt, kommen mehrere Entitäten als Referenten infrage, kann der Bezug zum jeweils nächstgelegenen Kandidaten als eine stärkere Bindung betrachtet werden.

5.3.5.2. Subjektidentitätsauflösung

Bei der Auflösung von Identitätsabbildungen von Subjektentitäten lässt sich eine weitere bereits für die Auflösung von Subjektpronomen als relevant identifizierte Eigenschaft der Subjektentitäten nutzen. Subjektentitäten besitzen als weiteres Merkmal Informationen über das Geschlecht (siehe Abschnitt 5.3.3.7), sofern diese ermittelbar sind. Ist dies der Fall, lassen sich die betrachtete Entität und die möglichen Kandidaten dahingehend überprüfen, ob ihr Geschlecht übereinstimmt.

Eine weitere Eigenschaft, die für eine Identität zwischen zwei Subjektentitäten spricht, ist eine Übereinstimmung ihrer textuellen Repräsentation. Stimmen diese nicht überein, kann es sich im Fall eines Subjekts nur dann um dieselbe Entität handeln, wenn diese ein Synonym besitzt, wie im Fall einer Person, die mittels einer anderen Form der Anrede z.B. nur mit Vor- bzw. Nachnamen angeredet wird oder im Fall eines angesprochenen Systems durch Synonyme für dieses, wie z.B. *robo* für *Roboter*. Da die Fälle der Synonyme und einer direkten Übereinstimmung der textuellen Repräsentation durch die Konstruktion der

Konzepte und der beschriebenen Überprüfung der Konzeptübereinstimmung bereits abgedeckt werden, müssen nur die verschiedenen Formen der Anrede einer Person überprüft werden. Hierbei existiert für eine betrachtete Subjektentität mit Vor- und Nachnamen die Möglichkeit eines Referenten mit nur dem Vornamen, nur dem Nachnamen oder einer Anredeform wie *Mr.* und dem Nachnamen. Trifft keine solche Form zu, kann es für eine Subjektentität auch keine Referenz geben. Damit wird diese Entität als erste Nennung der Entität klassifiziert.

5.3.5.3. Objektidentitätsauflösung

Die Auflösung von Objektidentitäten bietet durch deutlich detailliertere Informationen über Objektentitäten aus dem Kontextwissen weitere Eigenschaften, die für eine Übereinstimmung erfüllt sein müssen.

- (61) On the table are five cups put three cups on the cupboard and bring two cups to me
- (62) Put the green cup on the table and take the red cup then fill the cup with water and bring the red cup to me

Zunächst einmal lässt sich feststellen, dass falls die betrachtete Entität einen unbestimmten Artikel oder einen anderen unspezifischen Determinierer, wie z.B. *any*, *another* oder *some* besitzt, sie keine vorangegangene Entität als Referenten besitzt, da diese Formulierungen immer dann benutzt werden, wenn eine Entität das erste Mal in den Diskurs eingeführt wird (siehe Abschnitt 2.4) oder keine spezifische Entität als Bezug gemeint ist.

Außerdem lassen sich die Informationen über die Quantität der Entitäten vergleichen, denn stimmt die Anzahl zweier im Plural stehenden Entitäten nicht überein, können sie keine Identitätsabbildung darstellen. In Beispiel 61 ist dieser Fall dargestellt. Hier stellt keine der *cups*-Entitäten eine vollständige Identität der Anderen dar, da alle Quantitäten nicht übereinstimmen.

Des Weiteren lassen sich die beschreibenden Adjektive als Eigenschaft heranziehen: Hat die betrachtete Entität beschreibende Adjektive, so muss auch ein möglicher Kandidat, für eine Referenz dieser Entität, dieselben Adjektive besitzen. Wäre dies nicht der Fall, ließe sich nicht überprüfen, ob die Entitäten gleichartig sind. Hierdurch lässt sich in Beispiel 62 bestimmen, dass die *red cup*-Entität keine Identität der *green cup*-Entität darstellt, da das Adjektiv *red* für die *green cup*-Entität nicht vorhanden ist. Man könnte bei diesem Vergleich ebenso Synonyme der Adjektive zulassen. Da aber der Mensch zumeist Objekte, die er zuvor mit einem Adjektiv belegt hat, wieder mit demselben Adjektiv beschreibt, kann dies vernachlässigt werden. Hat die betrachtete Entität keine beschreibenden Adjektive, können allerdings alle Entitäten, egal ob mit oder ohne beschreibende Adjektive, infrage kommen, da hiermit eine allgemeinere Aussage über eine Entität getroffen wurde. Dieser Fall ist in Beispiel 62 durch die *cup*-Entität ohne Adjektive verdeutlicht, bei der beide vorherigen *cup*-Entitäten infrage kommen. Stellt der betrachtete Kandidat hingegen selbst eine allgemeinere Aussage der betrachteten Entität dar und es ist bekannt, dass er auf einen anderen Kandidaten verweist, der die gleichen Adjektive besitzt, so kommt er ebenso infrage. Diesen Fall stellt die *cup*-Entität in Beispiel 62 dar, welche eine allgemeinere Form der *red cup*-Entität darstellt und somit der Referent der letzten *red cup*-Entität ist.

Auch hierbei muss, wie bei der Subjektidentitätsauflösung, ein möglicher Kandidat eine Übereinstimmung der textuellen Repräsentation bzw. ein Synonym oder Hyperonym darstellen. Ohne eine Übereinstimmung einer dieser Fälle, kann ein Kandidat keine gleichartige Entität darstellen und damit keine Identität.

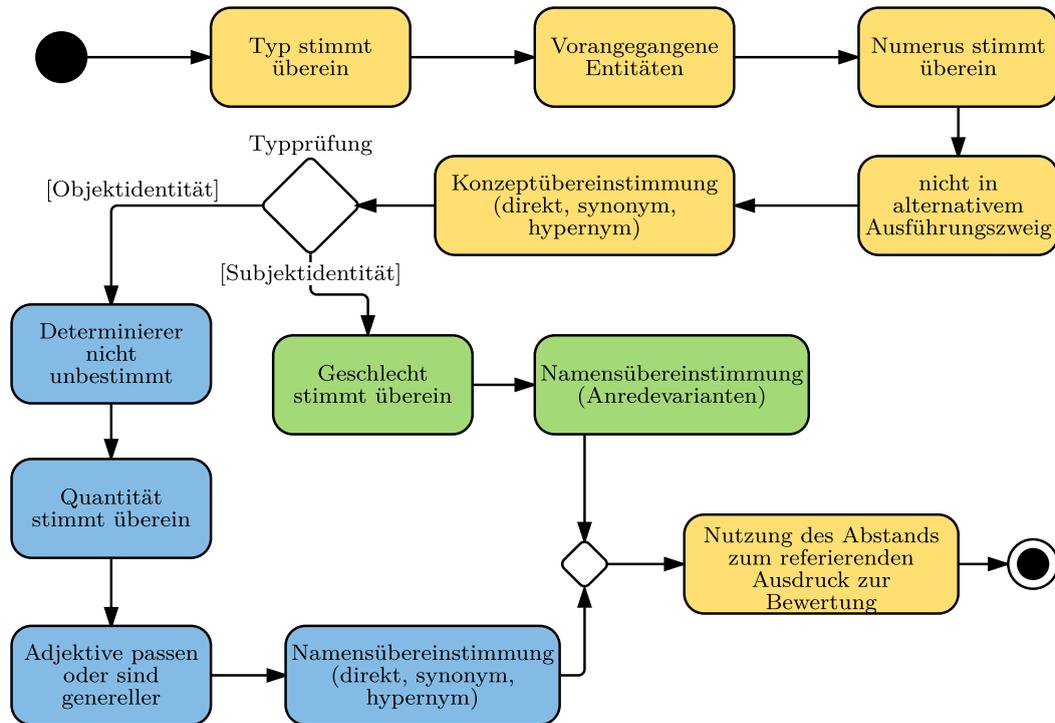


Abbildung 5.11.: Schritte der Identitätsauflösung

- (63) The cup is on the table can you bring the cup to me
- (64) Open the fridge and then put the cup next to the refridgerator
- (65) The lemon juice is in the fridge can you bring the juice to me

Der einfachste Fall ist hierbei eine Übereinstimmung der textuellen Repräsentation der betrachteten Entität mit der textuellen Repräsentation eines möglichen Kandidaten, wie es bei den beiden *cup*-Entitäten in Beispiel 63 der Fall ist. Aber auch der Fall, bei dem es sich bei der textuellen Repräsentation der betrachteten Entität, um ein Synonym der textuellen Repräsentation des möglichen Kandidaten handelt, kann eine Identität darstellen, wie bei *refridgerator* und *fridge* in Beispiel 64. Der dritte mögliche Fall ist der, dass die betrachtete Entität ein Hyperonym des Kandidaten und somit eine generalisierte Form darstellt, wie im Fall der *juice*-Entität, welche ein Hyperonym des *lemon juice* in Beispiel 65 darstellt. Trifft keine dieser drei Möglichkeiten zu, existiert für die betrachtete Entität vermutlich keine Identitätsabbildung oder sie lässt sich mit den zur Verfügung stehenden Eigenschaften nicht identifizieren.

Somit lassen sich die für eine Überprüfung auf Identitätsabbildungen einer Entität durchzuführenden Schritte wie in Abbildung 5.11 dargestellt beschreiben. Es wird also zunächst überprüft ob der Typ eines Kandidaten mit dem Entitätstyp des referierenden Ausdrucks übereinstimmt, sowie ob der Kandidat vor dem referierenden Ausdruck genannt wurde. Daraufhin lässt sich überprüfen ob der Numerus des Kandidaten mit dem Numerus des referierenden Ausdrucks übereinstimmt und ob sich der Kandidat in einem alternativen Ausführungszweig zum aktuellen Bezugspunkt befindet. Anschließend wird überprüft ob das dargestellte Konzept des referierenden Ausdruck das gleiche, ein Synonymes oder Hyperonymes Konzept des Kandidaten darstellt (siehe Abschnitt 5.3.5.1). Nun lassen sich

noch die für den Typ der Identität spezifischen Informationen überprüfen.

So wird für eine Objektidentität überprüft ob der Determinierer des referierenden Ausdrucks spezifisch ist, denn sonst würde es sich um eine unbestimmte Referenz handeln, ob die Quantitäten übereinstimmen und ob die Adjektive des referierenden Ausdrucks gleich oder genereller als die des Kandidaten sind. Außerdem wird überprüft ob die Namen der Entitäten übereinstimmen, synonym oder der Name des referierenden Ausdrucks hyperonym zum Namen des Kandidaten ist.

Für eine Subjektidentität hingegen wird überprüft ob das Geschlecht übereinstimmt und im Fall einer Person, ob die textuelle Repräsentation des Kandidaten eine Anredevariante der textuellen Repräsentation des referierenden Ausdrucks darstellt.

Die Eigenschaften der Identitätsauflösung sind im Gegensatz zu denen der Pronomen alle harte Auswahlkriterien und sorgen somit dafür, dass die Menge der möglichen Kandidaten nach und nach verkleinert wird, sodass im letzten Schritt eine Bewertung anhand des Abstands zum referierenden Ausdruck (siehe Abschnitt 5.3.3.11) durchgeführt werden kann.

6. Implementierung

Für die Umsetzung der in Kapitel 5 beschriebenen Analysen und ihrer identifizierten Teilschritte wurde das in Abschnitt 3.2 beschriebene Projekt-Rahmenwerk um zwei weitere Agenten erweitert. Diese Wahl wurde getroffen, da hiermit die Implementierung der als Anwendungsbeispiel für den Nutzen der Kontextanalyse gedachten Korreferenzanalyse die Umsetzung der Kontextanalyse nicht betrifft. Zusätzlich bietet sich so die Möglichkeit, die mehrfache unabhängige Ausführung der Agenten zu verwenden, um Synergien zwischen den beiden Analysen zu nutzen, wie es im Fall der Zustandsanalyse geschieht (siehe Abschnitt 5.2.3.4).

Doch vor der Umsetzung der beiden Agenten wurden mehrere Schritte, die nötig für eine sinnvolle Umsetzung dieser sind, als Verarbeitungen identifiziert, die nicht die Analysen betreffen, sondern vielmehr Schritte der Vorverarbeitung darstellen und zur Bereitstellung von allgemeinen Informationen dienen.

6.1. Anpassungen an der Vorverarbeitung

Für die Kontextanalyse wurden in Abschnitt 5.2.3.1 die Erkennung von Eigennamen (siehe Abschnitt 2.3.3) sowie in Abschnitt 5.2.3.2 das Markieren von semantischen Rollen (siehe Abschnitt 2.3.4) als Informationsquellen für die Analyse identifiziert. Da beide Schritte allerdings zuvor nicht im Vorverarbeitungsfließband des Projektes behandelt wurden, wurde dieses zunächst erweitert. Als Werkzeug zur Generierung dieser Informationen fiel die Wahl auf *Senna* [CWB⁺11], welches ein neuronales Netzwerk nutzt um die Eigennamen sowie semantische Rollen in einem gegebenen Text zu erkennen. Die Wahl von *Senna* als Werkzeug resultiert daraus, dass *Senna* bereits als Teil des Vorverarbeitungsfließbands für die Generierung der Informationen der Wortart-Markierungen und der Phrasenerkennung genutzt wurde (siehe [Koc15]). Somit musste kein weiteres Werkzeug für die Gewinnung der Informationen eingebunden werden, sondern nur die gewonnenen Informationen ebenso geeignet in den als geteilten Datenspeicher genutzten Graphen überführt werden.

6.1.1. Eigennamenerkenner

Für die Erkennung der Eigennamen (vgl. Abschnitt 2.3.3) wurde eine weitere Pipelinestufe hinzugefügt, welche die pro Wort bestimmten Eigennamen-Markierungen (siehe Tabelle 2.3) aus *Senna* als weiteres Attribut der bereits für die Wortart-Markierungen und Phraseninformationen im Graph vorhandenen Token-Knoten hinzufügt. Hierzu wird der

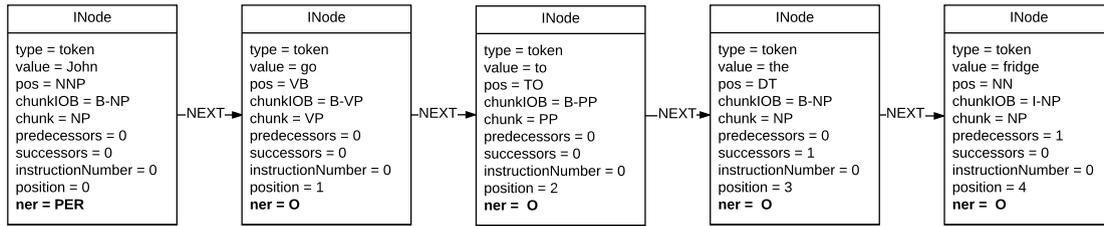


Abbildung 6.1.: Vorverarbeitungsfließband-Ausgabe nach Ausführung des Eigennamenerkenners

Eingabetext der *Senna*-Implementierung mit der Option *-ner* übergeben. *Senna* gibt als Ausgabe daraufhin pro Wort entweder, bei erkanntem Eigennamens-Typ, die Eigennamen-Markierung oder die Markierung *O* für den Fall, dass kein Eigenname gefunden wurde, zurück. Diese Information wird daraufhin mit dem entsprechenden Token-Knoten verknüpft. Somit sieht das Ergebnis des Vorverarbeitungsfließbands nach der Ausführung des Eigennamenerkenners für den Satz „John go to the fridge“ wie in Abbildung 6.1 aus. Hierbei wurde „John“ als Eigenname der Kategorie Person klassifiziert und die restlichen Wörter des Satzes mit der Markierung *O* versehen, welche angibt, dass es sich bei ihnen um keine Eigennamen handelt.

6.1.2. Semantische-Rollen-Markierer

Auch das Markieren der semantischen Rollen, der Argumente von in der Eingabe auftretenden Prädikaten, ist mit *Senna* möglich. *Senna* gibt pro Token jeweils eine Spalte pro erkanntem Prädikat aus, in welchem im IOBES-Format (siehe Abschnitt 2.3.2) die jeweilige semantische Rolle in der PropBank-Argument-Notation (siehe Abschnitt 2.5.2) enthalten ist.

(66) [John, -, S-A1, -] [go, go, S-V, -] [to, -, B-A4, -] [the, -, I-A4, -] [fridge, -, E-A4, -]
 [and, -, -, -] [take, take, -, S-V] [the, -, -, B-A1] [cup, -, -, E-A1]

So ist die Ausgabe für den Satz „John go to the fridge and take the cup“ wie in Beispiel 66 dargestellt. Hierbei stellen die eckigen Klammern die einzelnen Ausgaben dar bei denen der erste Eintrag das Token selbst angibt, der zweite um welches Prädikat aus PropBank es sich handelt, der dritte die Argumentnummer für das erste auftretende Prädikat und der vierte die Argumentnummer für das zweite auftretende Prädikat. Da die Argumentnummern allerdings so noch keine eindeutige Aussage über die eigentliche semantische Rolle des Arguments bieten, wurde noch eine Anbindung an PropBank durchgeführt, um die eigentlichen Rollen zu den Argumenten zu bestimmen. Da viele der PropBank-Rahmen ebenso ein Pendant in VerbNet und FrameNet besitzen, wurde die Abbildungsdatei des Wissensextraktion-Werkzeugs *pikes* [CRAAr] dazu verwendet, die Argumente auf die Rollen in PropBank abzubilden, sowie falls vorhanden ebenso die Rollen aus VerbNet und FrameNet zu erhalten. Die Abbildungsdatei enthält pro PropBank-Bedeutung eines Prädikats die folgenden Einträge: Bedeutungsname, Lemma des Verbs, Aktionsbeschreibung, VerbNet-Rahmen, FrameNet-Rahmen und Eventtypen. Außerdem enthält sie pro Argument der PropBank-Bedeutung: Argumentnummer, PropBank Rolle, mögliche VerbNet Rollen und mögliche FrameNet Rollen. Um nun das Ergebnis von *Senna* auf die Informationen in der Abbildungsdatei projizieren zu können muss die PropBank-Bedeutung des von *Senna* identifizierten Prädikats bestimmt werden. Da von *Senna* nur Informationen über die tatsächlich in der Formulierung gegebenen Argumente geliefert werden, kann die Bedeutung nur über einen Abgleich der gegebenen Argumente mit den für eine

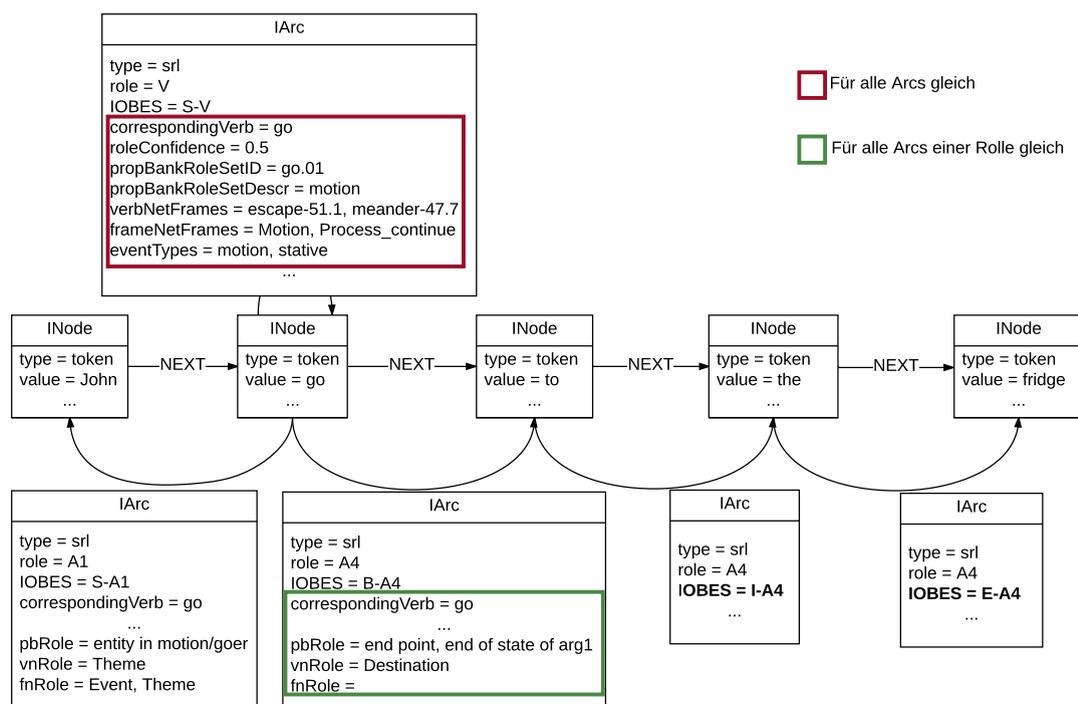


Abbildung 6.2.: Vorverarbeitungsfließband-Ausgabe nach Ausführung des Semantischen Rollen-Markierers

Bedeutung möglichen Argumenten erfolgen und damit nicht immer eindeutig bestimmt werden. Da aber die Bedeutungen in PropBank nach Auftretenswahrscheinlichkeit sortiert aufgelistet sind, kann in diesem Fall die zuerst aufgelistete infrage kommende Bedeutung angenommen werden. Da man eventuell im Projektverlauf Informationen über die Güte dieser Zuordnung haben möchte, wird ebenfalls eine Wahrscheinlichkeit für die Korrektheit der Abbildung berechnet, welche sich aus der Anzahl der ebenfalls infrage kommenden Bedeutungen ergibt.

Sind die Informationen zugeordnet, wird das Ergebnis als neue Kanten in den Graph eingefügt. So werden ausgehend von dem Token-Knoten, welcher das erkannte Prädikat repräsentiert, Kanten zu den die Argumente repräsentierenden Token-Knoten gespannt. Diese Kanten enthalten als Attribute die Informationen aus der Abbildung. Damit sieht das Ergebnis des Vorverarbeitungsfließbands nach dem Ausführen des Semantischen Rollen-Markierers für das Beispiel „John go to the fridge“ wie in Abbildung 6.2 dargestellt aus.

Die Erkennung semantischer Rollen ist ebenso wie viele andere Schritte der Semantikanalyse anfällig für die Gegebenheiten der gesprochenen natürlichen Sprache: Auch für diese Aufgabe sind die Satzzeichen ein wichtiges Indiz für die infrage kommenden Argumente eines Prädikats. Da die Satzzeichen aber bedingt durch die gesprochene Sprache nicht zur Verfügung stehen, enthält das Ergebnis von *Senna* viele fehlerhafte Argumente. Um trotzdem ein für diese und weitere Arbeiten verwertbares Ergebnis des Semantischen Rollen-Markierers zu erzielen, wurde auf die Instruktionsnummer-Heuristik (siehe [Koc15]) zurückgegriffen. Indem jede erkannte Instruktion einzeln als Eingabe an *Senna* übergeben wird, kann es zu deutlich weniger zusätzlich oder falsch erkannten Argumenten kommen. Dafür ist es allerdings möglich, dass durch die Heuristik falsch getrennte Instruktionen Argumente abschneiden. Da dies aber in den meisten Fällen des momentanen Korpus nicht der Fall ist, bietet diese Herangehensweise eine valide Option, um an die semantischen Rollen der Argumente zu gelangen. Außerdem wurde die Instruktionsnummer-Heuristik noch um weitere Trennwörter, wie Begriffe mit zeitlicher Einordnung (*then*, *afterwards*

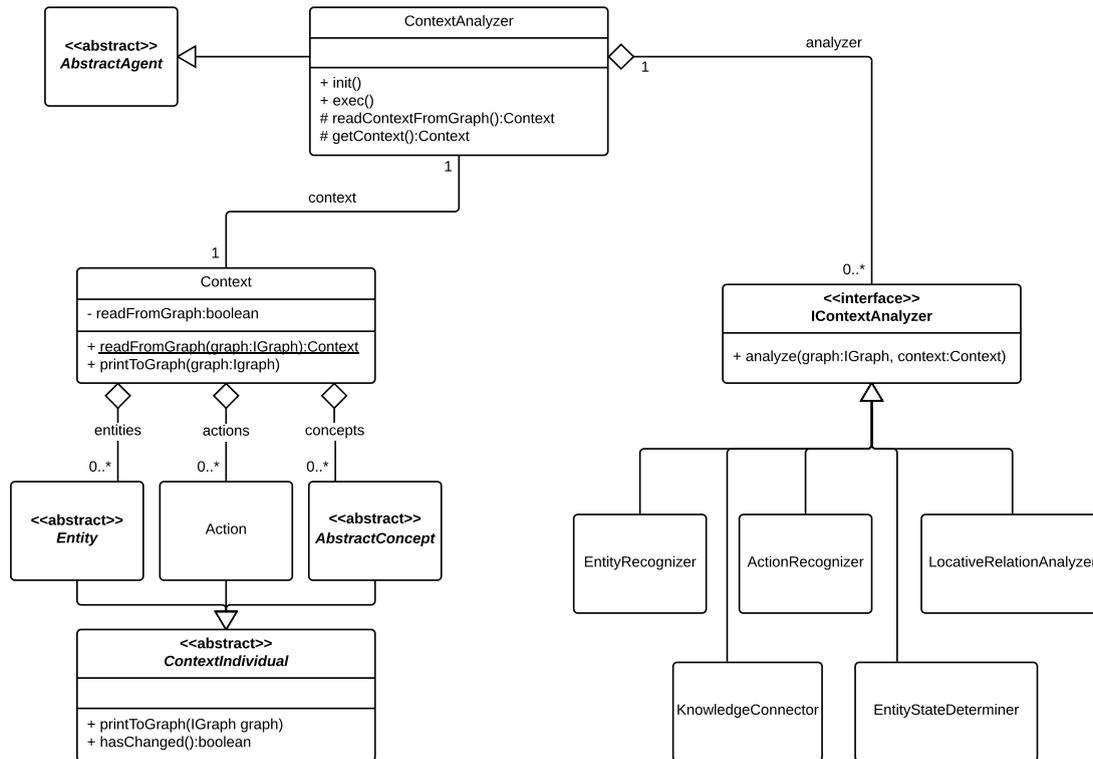


Abbildung 6.3.: Aufbau des Kontextanalyserers mit angedeutetem Kontext

usw.) und Begriffe, welche neue Bedingungszeige einführen (wie z.B. *if*, *otherwise* usw.) erweitert, die ebenso in den meisten Fällen eine neue Instruktion erzeugen.

6.2. Kontextanalyse

Für die Umsetzung der in Abschnitt 5.2 beschriebenen Kontextanalyse mit ihren identifizierten Teilschritten wurde der Agent `ContextAnalyzer` implementiert, welcher das Ergebnis des Vorverarbeitungsfließbands nach Ausführung des Semantische-Rollen-Markierers (vgl. Abbildung 6.2) als Eingabe entgegen nimmt und aus den gegebenen Informationen eine Repräsentation des Kontextwissens erstellt. Hierzu wurden die identifizierten Einzelschritte als aufeinanderfolgende Analyseschritte umgesetzt sowie die Erkennung der Ortsbeziehungen aus der Entitätenerkennung extrahiert, da diese durch Informationen aus der Aktionserkennung verbessert werden konnte. Hierdurch ergibt sich die folgende Reihenfolge für die Analyseschritte: Entitätenerkennung (Abschnitt 6.2.1), Aktionserkennung (Abschnitt 6.2.2), Ortsbeziehungserkennung (Abschnitt 6.2.3), Welt- und Domänenwissensanreicherung (Abschnitt 6.2.4) sowie Zustandsbestimmung (Abschnitt 6.2.5). Um jederzeit weitere Analyseschritte hinzufügen zu können wurde die Schnittstelle `IContextAnalyzer` mit der Methode `analyze(IGraph graph, Context context)` eingeführt, welche die einzelnen Analyseschritte implementieren. Übergeben wird hierbei der aktuell betrachtete Graph sowie die momentane Repräsentation des Kontextes, welche in den Analyseschritten verändert wird. Hierdurch entsteht der in Abbildung 6.3 dargestellte Aufbau mit zusammengefasstem Aufbau des Kontextes. Der vollständige Aufbau der Kontextrepräsentation ist in Abbildung 6.4 dargestellt.

Außerdem wurde die in Abschnitt 5.2.4 für die Kontextinformationen identifizierte Repräsentation in eine aus dem Graph standardisiert les- und schreibbare Struktur überführt, die sich flexibel erweitern lässt. Hierzu wurden die Kontextindividuen sowie die

Konzepte unter der abstrakten Klasse `ContextIndividual` zusammengefasst, welche es ermöglicht sie in den Graph zu schreiben (siehe Abbildung 6.4), sofern sie sich von der aus dem Graph gelesenen Version unterscheiden. Zusätzlich lassen sich durch die abstrakte Klasse alle Kontextwissenelemente mit Relationen, welche die Beziehungen zwischen den Kontextwissenelementen abbilden und beliebig erweitert werden können, versehen. Die Hierarchieebenen der Kontextrepräsentation (siehe Abschnitt 5.2.1) wurden ebenso in die Implementierung übertragen. Der Übergang von der Hierarchieebene der textuellen Repräsentation zur Kontextebene wird durch die in den Entitäten (`Entity`) und Aktionen (`Action`) enthaltenen Referenzlisten, welche die sie bildenden Token-Knoten des Graphen enthalten, dargestellt. Der Übergang von der Kontextebene auf die Konzeptebene wird über Relationen zwischen `Entity` bzw. `Action` und ihrem jeweiligen übergeordneten `AbstractConcept` abgebildet. Außerdem verfügen die Konzepte über eine konzepthierarchiebildende Struktur. Jedes das `AbstractConcept` erweiternde Konzept hat zunächst die Möglichkeit, gleichwertige Konzepte über die `equalConcept`-Menge zu definieren. Des Weiteren können Über- und Unterkonzepte über die `superConcept`- und `subConcept`-Mengen sowie Teil-Ganzes-Beziehungen zwischen Konzepten über die `partConcept`- und `partOfConcept`-Mengen definiert werden (siehe Abbildung 6.4).

Dieser Aufbau bietet also die Möglichkeit die Kontextwissenselemente beliebig zu erweitern und jederzeit neue Beziehungen zwischen ihnen zu definieren sowie alles in einer festen Struktur in den Graph zu überführen und wieder auslesbar zu machen. Dies erlaubt es anderen Agenten einfach mittels der `Context`-Klasse die im Graph enthaltenen Kontextinformationen auszulesen (`Context.readFromGraph`) und für sich zu nutzen, sowie falls neue Erkenntnisse in das Kontextwissen überführt wurden, die Informationen im Graph zu aktualisieren (`Context.printToGraph`).

6.2.1. Entitätenerkennung

Der Entitätenerkennung (`EntityRecognizer`) deckt den ersten in der Kontextanalyse durchgeführten Teilschritt, die Entitätenerkennung (siehe Abschnitt 5.2.3.1), ab und erhält dabei als Eingabe für seine Analyse den momentanen Graph sowie die sich momentan im Graph befindende Version des Kontext (`Context`). Falls es sich um die erste Ausführung des Kontextanalysierers handelt, ist der Kontext leer.

Zunächst wird die für eine Erkennung der Subjektentitäten nötige Information über die möglichen Namen der in der Domäne auftretenden Systeme aus der Domänenontologie ausgelesen. Daraufhin können die als Bearbeitungsgrundlage benötigten Nominalphrasen aus den Token-Knoten des Graphs ausgelesen werden, indem die `chunk`-Attribute auf *NP* untersucht werden. Da allerdings nicht jedes Auftreten eines Systems korrekt als Nominalphrase erkannt wird, wird außerdem in den anderen Phrasen nach Vorkommen der Systemnamen gesucht und diese ebenfalls der Liste der zu betrachtenden Phrasen hinzugefügt. Außerdem können die Nominalphrasen ebenso mehrere über eine Konjunktion verbundene Entitäten enthalten, was dazu führt, dass diese Nominalphrasen noch einmal unterteilt werden. Daraufhin wird für jede der zu betrachtenden Phrasen(-teile) überprüft, welche Art von Entität sie darstellt. Diese Überprüfung geschieht anhand der in Abschnitt 5.2.3.1 identifizierten Eigenschaften. So werden Subjektentitäten dadurch erkannt, dass die Phrase entweder ein Token enthält, welches eine vom Eigennamenerkennung identifizierte Person darstellt, oder aber dadurch, dass sie einen der Namen für die Systeme der Domäne darstellt. Pronomenentitäten hingegen sind in Nominalphrasen enthalten, die kein Nomen aber ein Personalpronomen enthalten und Objektentitäten sind alle übrigbleibenden Nominalphrasen, die zumindest ein Nomen als Kopf enthalten.

Nach dieser Einteilung, lassen sich die weiteren in Abschnitt 5.2.3.1 für die jeweiligen Pronomentypen identifizierten Informationen überprüfen und daraus die jeweilige Entität erstellen und dem Kontext hinzufügen. Der Name der Entität ergibt sich dabei aus den

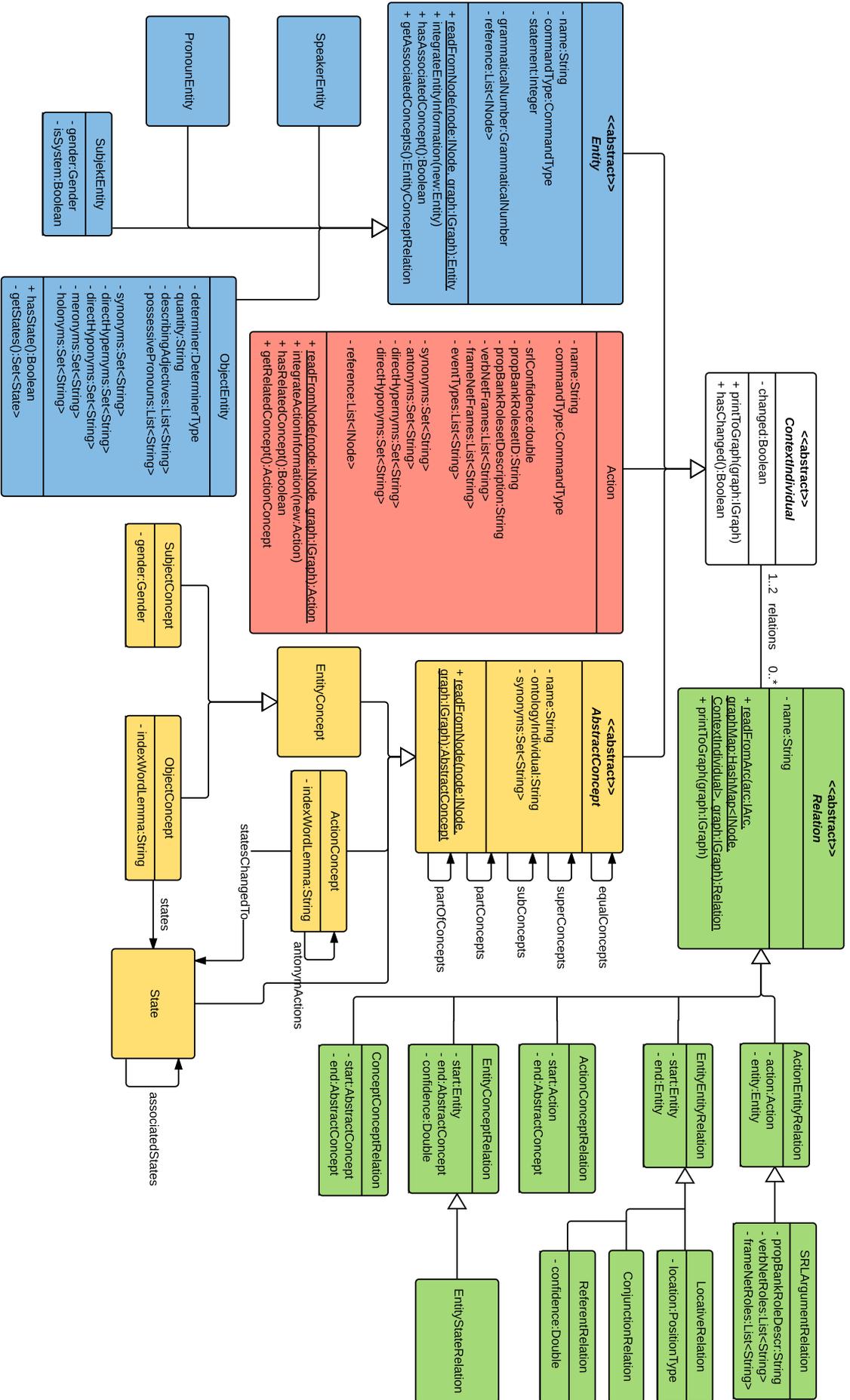


Abbildung 6.4.: Repräsentation des Kontextwissens

Substantiven in der Nominalphrase oder im Fall einer Pronomenentität aus dem erkannten Personalpronomen. Der Numerus lässt sich aus den Wortart-Markierungen der Substantive¹ ableiten oder im Fall einer Objektentität über eine in der Phrase enthaltene, erkannte Kardinalzahl (Wortart-Markierung *CD*) bestimmen. Wird hingegen keine Kardinalzahl gefunden, erhält das Attribut `quantity` den Wert `UNKNOWN`. Die beschreibenden Adjektive (Wortart-Markierung *JJ*) sowie Possessivpronomen (Wortart-Markierung *PRP\$*), die die Objektentitäten näher beschreiben, werden ebenso über ihre jeweilige Wortart-Markierung identifiziert.

Die Informationen über semantische Relationen der Objektentitäten werden mittels des Werkzeugs *extJWNL* [Aut] aus den Informationen der WordNet-Datenbank (siehe Abschnitt 2.5.1) in Version 3.0 gewonnen. Hierbei werden alle für den Namen der Entität infrage kommenden Synsets betrachtet, was dazu führen kann, dass die Informationen nicht mit der Bedeutung im aktuellen Kontext übereinstimmt. Da sich aber die aktuelle Bedeutung der Entität zu diesem Zeitpunkt nur schwer korrekt ermitteln lässt, wurde sich dafür entschieden, mehr, wenn auch eventuell falsche Informationen, zu generieren, die später durch eine vollständige Wortbedeutungsanalyse wieder eingeschränkt werden können.

Die Information über das Geschlecht der Subjektentitäten wird mittels der Geschlechtererkennung des *Stanford Deterministic Coreference Resolution Systems* (siehe Abschnitt 2.4.2) aus der Werkzeugsammlung *Stanford CoreNLP* [sta] umgesetzt, welche auf der Geschlechtstabelle von Bergsma und Lin [BL06] basiert.

Außerdem lässt sich, falls eine Teilung der Phrase durch eine Konjunktion erfolgt ist, eine Konjunktionsbeziehung (`ConjunctionRelation`) zwischen den beiden in den Teilen identifizierten Entitäten definieren. Außerdem wird für jede Entität überprüft, ob die Token-Knoten, die sie bilden, Informationen aus dem Bedingungserkennungs-Agenten ([Ste16]) enthalten. Ist dies der Fall kann so eine Angabe darüber gespeichert werden, ob die Entität in einer Bedingung auftritt.

Sollte es sich bei der gefundenen Entität um eine bereits im vorherigen Abbild des Kontextes enthaltene Entität handeln, kann die bereits enthaltene Entität mittels der Methode `integrateEntityInformation(new:Entity)` aktualisiert werden.

Das Ergebnis dieses ersten Schrittes der Kontextanalyse ist also der Kontext angereichert mit den verschiedenen in der Äußerung enthaltenen `Entity`-Ausprägungen. So ergeben sich für das Beispiel „John go to the one white fridge“ die in Abbildung 6.5 dargestellten Entitäten. Hier bildet die Nominalphrase „John“ die Subjektentität (`SubjectEntity`) mit Namen *John* ab, welche kein System darstellt, sondern eine Person im Singular. Die Nominalphrase „the two white fridges“ hingegen bildet die Objektentität (`ObjectEntity`) mit Namen *fridges*, welche einen spezifischen Determinierer besitzt und im Plural mit der Quantität zwei beschrieben wurde. Außerdem wird sie vom Adjektiv „white“ näher beschrieben und enthält bereits die für die spätere Konzeptualisierung verwendeten Informationen zu den semantischen Relationen des sich aus dem Namen *fridges* ergebenden Lemma *fridge*.

6.2.2. Aktionserkennung

Der Aktionserkennung (`ActionRecognizer`) befasst sich mit dem zweiten Teilschritt der Kontextanalyse, der Erkennung von Aktionen (siehe Abschnitt 5.2.3.2). Hierfür erhält er den Graph sowie den bisherigen Kontext (`Context`), nach Ausführung des Entitätenerkennung, als Eingabe. Dies ist nötig, damit die für die Aktionen erkannten semantischen Rollen direkt mit den erkannten Entitäten verbunden werden können. Als Mittel zur Erkennung der Aktionen werden die Verbalphrasen herangezogen, also alle Token-Knoten aus dem

¹ *NNP = Singular, NNPS & NNS = Plural* sowie *NN = MassOrSingular*

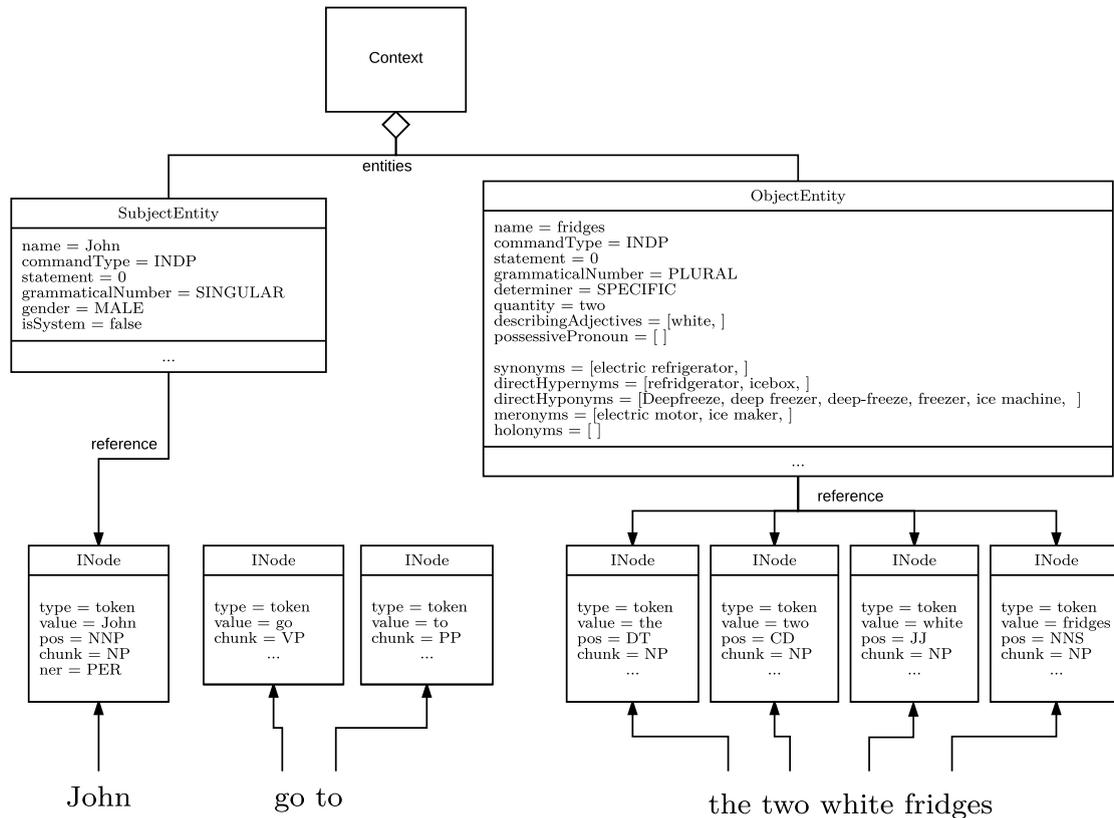


Abbildung 6.5.: Ergebnis des Entitätenerkenners

Graph, welche zusammen eine Verbalphrase² bilden. Sind diese aus dem Graph bestimmt, lässt sich pro Verbalphrase das vom Semantische-Rollen-Markierer identifizierte Prädikat bestimmen und für dieses eine neue Aktion (**Action**) erzeugen. Diese erhält als Namen das erkannte Prädikat und wird, wie bereits im Entitätenerkenners, mit den semantischen Relationen aus WordNet angereichert. Außerdem lassen sich die für das Prädikat vom Semantische-Rollen-Markierer bestimmten Informationen, wie z.B. die PropBank-Roleset-Beschreibung, die VerbNet- und die FrameNet-Rahmen, auslesen. Im nächsten Schritt werden die semantischen Rollen des Prädikats aus den Kanten des bildenden Token-Knoten ausgelesen und die Entität (**Entity**) aus dem Kontext (**Context**) gesucht, die als Referenz die Token-Knoten besitzt, welche die semantische Rolle darstellen, also das Ziel der Kante sind. Ist die Entität bestimmt, lässt sich eine semantische Rollen-Beziehung (**SRLArgumentRelation**) zwischen Beiden beschreiben, welche die Information über die PropBank-, VerbNet- sowie FrameNet-Rolle der Entität in dieser Aktion beschreibt.

Wie auch schon bei den Entitäten lässt sich überprüfen, ob die erkannte Aktion (**Action**) bereits im Kontext enthalten ist und falls ja, diese über die Methode `integrateAction-Information(new:Action)` aktualisieren. Somit ist das Ergebnis nach diesem Teilschritt der Kontextanalyse eine Kontextrepräsentation, die die erkannten Entitäten und Aktionen enthält sowie die semantischen Rollenbeziehungen zwischen ihnen. So ergibt sich für das vorherige Beispiel der in Abbildung 6.6 dargestellte Kontext. Hier wurde nun für die Verbalphrase „go“ die Aktion *go* erkannt und mit den Informationen des Semantische-Rollen-Markierers versehen, sowie um die WordNet-Informationen zu semantischen Relationen erweitert. Außerdem sind die zwei semantischen Rollen der Aktion über Beziehungen zu den jeweiligen Entitäten repräsentiert, welche die Informationen über die eingenommene

²chunk = VP

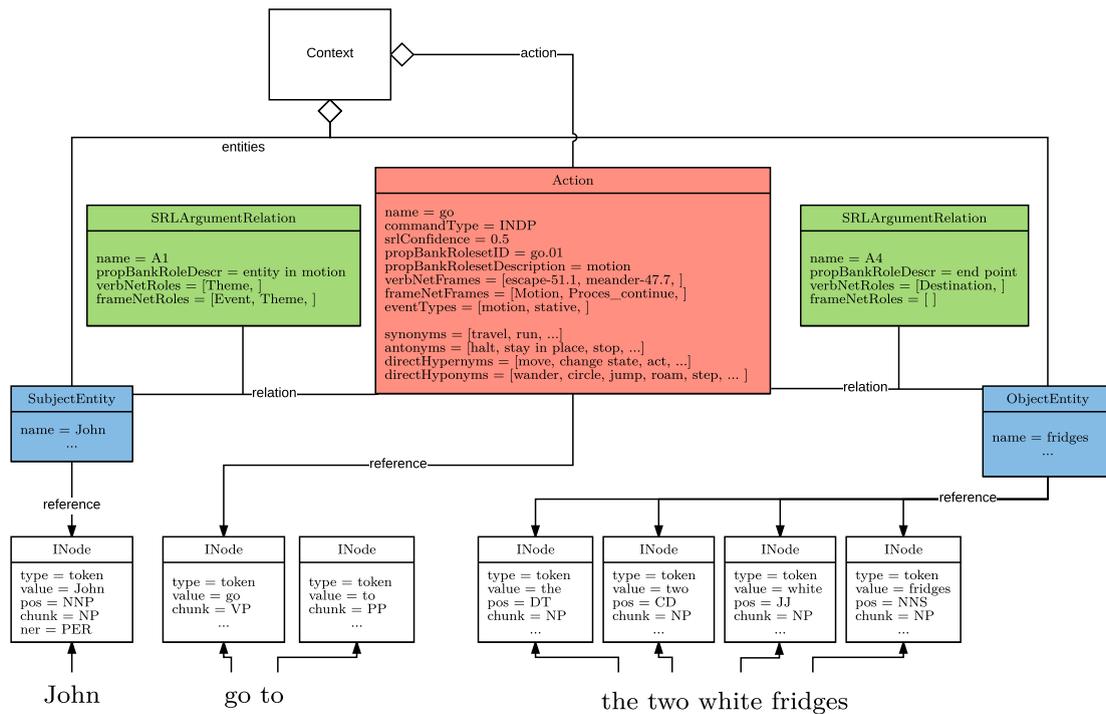


Abbildung 6.6.: Ergebnis des Kontext nach dem Aktionserkennung

Rolle aus den drei Datenbanken enthält. *John* stellt also die sich bewegende Entität der Aktion *go* dar und *fridges* das Ziel der Bewegung.

6.2.3. Ortsbeziehungserkennung

Der Ortsbeziehungserkennung (*LocativeRelationAnalyzer*) untersucht die Äußerung auf bestehende Ortsbeschreibungen zwischen den erkannten Entitäten. Um ohne Syntaxbäume überhaupt eine Erkennung von Beziehungen zwischen Entitäten zu ermöglichen, wird in dieser Arbeit eine simple Heuristik zur Bestimmung der Ortsbeziehungen eingesetzt, die zumindest in der Lage ist, viele der in Anweisungen auftretenden Ortsbeschreibungen zu erkennen. Hierzu werden nur Formulierungen zwischen zwei aufeinanderfolgenden Entitäten betrachtet und in diesem „Zwischenraum“ nach den in Abschnitt 5.2.3.1 identifizierten Signalwörtern gesucht. Dieser Ansatz wurde gewählt, da Präpositionalphrasen in der englischen Sprache zumeist nach der beschriebenen Entität auftreten. In den Formulierungen zwischen zwei Entitäten wird dann nach den Schlüsselwörtern (siehe Tabelle 5.1) gesucht indem die textuellen Repräsentationen verglichen werden.

(67) Armar look inside the fridge

Außerdem wird überprüft, ob die beiden aufeinanderfolgenden Entitäten an ein und derselben Aktion teilnehmen. Ist dies der Fall, wird nur dann nach Ortsbeziehungen zwischen ihnen gesucht, wenn die später genannte Entität eine ortsbestimmende Rolle (Argumentnummer *AM-LOC*) in der Aktion einnimmt sowie die vorherige Entität der Proto-Patient der Aktion ist. Wäre dies nicht der Fall, würden Beide eine nicht ortsbeschreibende semantische Rolle einnehmen und es würde somit keine Beziehung existieren, wie es in Beispiel 67 der Fall ist. Hier wurde zwar das Schlüsselwort *inside* zwischen der Entität *Armar* und der Entität *fridge* gefunden, aber *Armar* stellt den Handelnden der Aktion *look* dar und wird damit durch die Aktion nicht als in der Entität *fridge* beschrieben. Nehmen die beiden Entitäten nicht an derselben Aktion teil, wird nur dann nach einer Ortsbeziehung

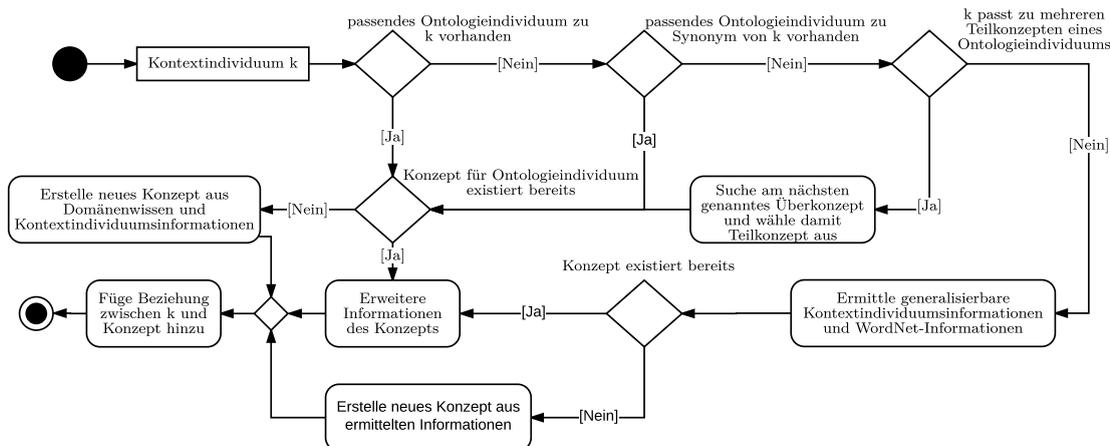


Abbildung 6.7.: Schematischer Ablauf der Konzeptbildung

gesucht, wenn keine Aktion zwischen den beiden Entitäten auftritt. Würde nämlich eine Aktion zwischen Beiden auftreten, wäre die entdeckte Formulierung nicht Teil einer Präpositionalphrase und hätte somit eine andere Bedeutung als eine Ortsbestimmung. Dieses Vorgehen wurde gewählt, um Fehler des Semantische-Rollen-Markierers abzufangen, falls dieser eine der Entitäten nicht als Teil der ausgeführten Aktion erkennt.

Wird daraufhin eine Ortsbeziehung zwischen zwei aufeinanderfolgenden Entitäten erkannt, wird diese mittels einer Ortsbeziehung (*LocativeRelation*) abgebildet und der entdeckte Ortsbestimmungstyp³ in dieser beschrieben.

(68) John go to the fridge next to the cupboard

So wird für Beispiel 68 eine Ortsbeziehung zwischen der Objektentität *fridge* und der Objektentität *cupboard* erzeugt, welche den Positionstyp (*PositionType*) *NEXT_TO* darstellt.

6.2.4. Welt- und Domänenwissenanreicherung

Zur Umsetzung der Welt- und Domänenwissenanreicherung (*KnowledgeConnector*), also der in Abschnitt 5.2.3.3 beschriebenen Funktionalität, werden die im Entitätenerkennung und Aktionserkennung identifizierten und damit im Kontext enthaltenen Kontextindividuen mittels einer Anbindung der Domänenontologie (siehe Abschnitt 3.4) sowie Wissen über die Hyper- und Hyponym-Hierarchie aus WordNet mit übergeordneten Konzepten versehen.

Dafür wird zunächst für jedes Kontextindividuum überprüft, ob ein passendes Domänenontologieindividuum existiert (für einen Überblick über den Ablauf siehe Abbildung 6.7). Hierbei muss der Typ des Kontextindividuum zum jeweiligen Domänenontologieindividuum passen⁴ sowie der Name des Kontextindividuum nach der Jaro-Winkler-Distanz eine Ähnlichkeit von über 92%⁵ aufweisen. Zur Berechnung der Jaro-Winkler-Distanz wird die Bibliothek *java-string-similarity* [Deb] verwendet. Wird ein passendes Ontologieindividuum gefunden, wird zunächst überprüft, ob es bereits ein Konzept im aktuellen Kontext gibt, welches den gleichen Namen trägt bzw. das gleiche Ontologieindividuum über die im Konzept gespeicherte Ontologieindividuumrepräsentation darstellt. Ist dies der Fall,

³Typen beschrieben in Tabelle 5.1

⁴`ObjectEntity = Object; SubjectEntity = System; Action = Method`

⁵Die Voranalyse auf den Beschreibungen des Korpus ergab diesen Wert als gutes Maß für Ähnlichkeit

werden die im gefundenen Konzept enthaltenen Informationen um die im Kontextindividuum repräsentierten Informationen erweitert. Wird kein Konzept gefunden, wird aus den generellen Informationen des Kontextindividuum und den Informationen des Ontologieindividuum ein neues Konzept erstellt. Existiert keine Übereinstimmung mit einem Domänenontologieindividuum und dem Namen des Kontextindividuum können ebenso die Synonyme des Kontextindividuum zum Abgleich herangezogen werden, sowie bei einer Nennung eines Teilkonzepts (*door* auf `Fridge.Door`) eines in der Domänenontologie beschriebenen Individuum, die in Abschnitt 5.2.3.1 beschriebene Zuordnung über das Konzept der nächst gelegenen infrage kommenden Entität. In jedem Fall erhält das Kontextindividuum eine Konzeptbeziehung (`EntityConceptRelation` bzw. `ActionConceptRelation`) zum es repräsentierenden Konzept, welches die Ähnlichkeit aus der Jaro-Winkler-Distanz als Wahrscheinlichkeit enthält.

Die Erstellung eines neuen Konzepts aus einem in der Domänenontologie gefundenen Individuum lässt sich in die drei Typen (`ObjectConcept`, `ActionConcept` und `SubjectConcept`) aufteilen. Für jeden Typ gibt es unterschiedliche Informationen in der Domänenontologie sowie in den jeweiligen Entitäten. Abhängig von diesen werden die Attribute der verschiedenen Ausprägungen des `AbstractConcept` gefüllt. Die jeweiligen Konzepte zu den synonymen, hyperonymen und meronymen Ontologieindividuen werden ebenso erstellt, falls diese nicht bereits vorhanden sind. Außerdem wird für die Zustände der Objekte (`Object`) sowie die Zustandsübergänge der Methoden (`Method`) eine weitere Form von Konzepten eingefügt, die Zustände repräsentiert durch die Klasse `State`, welche ihre assoziierten Zustände kennt. Diese Zustandskonzepte werden immer dann erstellt und in den `Context` eingebaut, wenn ein Objektkonzept (`ObjectConcept`) oder Aktionskonzept (`ActionConcept`) erstellt wird, welches mit einem Domänenontologieindividuum verknüpft wird, das einen Zustand beschreibt.

Wird kein Domänenontologieindividuum zum identifizierten Kontextindividuum gefunden, wird nach einem passenden das Kontextindividuum repräsentierenden Lemma in WordNet gesucht⁶ und ein Konzept erstellt⁷, welches sich auf die allgemeinen Informationen aus dem Kontextindividuum selbst sowie den Informationen aus WordNet stützt. Da die Informationen aus WordNet allerdings, wie bereits in Abschnitt 6.2.1 beschrieben, nicht unbedingt korrekt sein müssen, erhält die entsprechende `Relation` zwischen dem Kontextindividuum und dem Konzept zunächst nur die Wahrscheinlichkeit 10%, welche bei jedem weiteren Auftreten eines Kontextindividuum, welches eine Ausprägung des gleichen Konzepts darstellt, um 10% erhöht wird. Hierdurch lässt sich ermitteln, ob das identifizierte Konzept, ein im aktuellen Kontext auftretendes Konzept darstellt. Außerdem werden nur dann die Informationen über Hyper-/ Hyponyme bzw. Mero-/ Holonyme in das Konzept überführt, wenn bereits ein Konzept existiert, welches mit diesen übereinstimmt und somit direkt ein Konzept als Beziehung angegeben werden kann und nicht erst erstellt werden muss. Dies sorgt dafür, dass die Konzeptebene nicht mit falschen Informationen überfüllt wird.

Alle Konzepte, die Informationen aus WordNet enthalten, besitzen ebenso als Attribut das Lemma, welches die Informationen geliefert hat, um in späteren Schritten nachvollziehen zu können, woher die WordNet Informationen stammen. Somit resultiert der Kontext nach diesem ersten Teil der Wissensanreicherung für das Beispiel „John go to the two white fridges“ in dem in Abbildung 6.8 dargestellten Ergebnis, mit den in der Ontologie vorhandenen Individuen `Object.Fridge`, `Object.Fridge.Door`, `Method.move`, `Method.go`, `State.opened`, `State.closed`, `State.on`, `State.off`. Dabei wurden zunächst die Konzepte `Fridge` und `go` erkannt und mit der Ontologie verbunden, woraufhin aus den Informationen der Ontologieindividuen die weiteren Konzepte extrahiert und die Beziehungen

⁶Nur für Objektentitäten und Aktionen, da Subjektentitäten nicht in WordNet abgebildet werden.

⁷Vorausgesetzt es existiert nicht bereits ein passendes Konzept (Ähnlichkeit der Namen größer 92%).

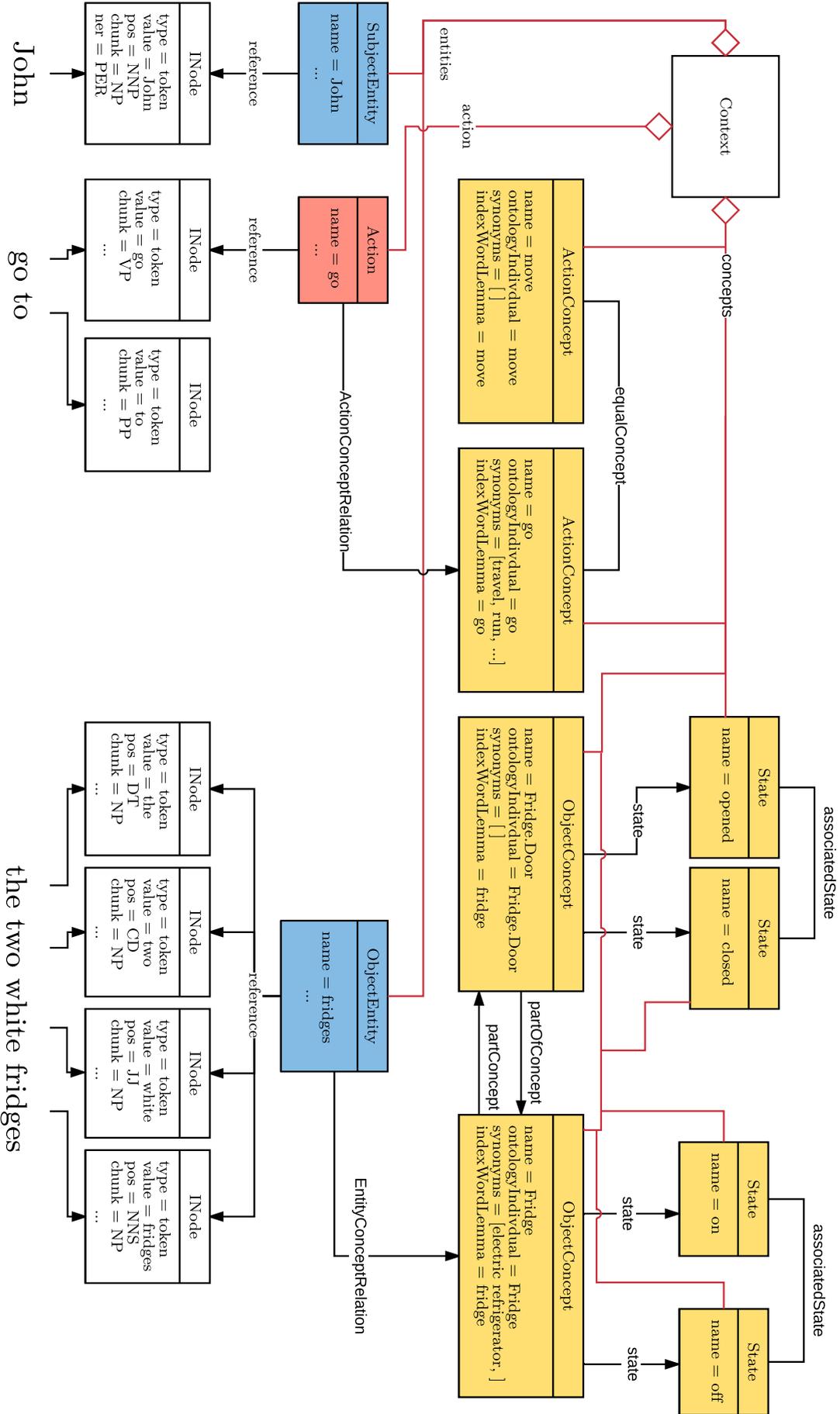


Abbildung 6.8.: Ergebnis des Kontext nach der Domänenanbindung

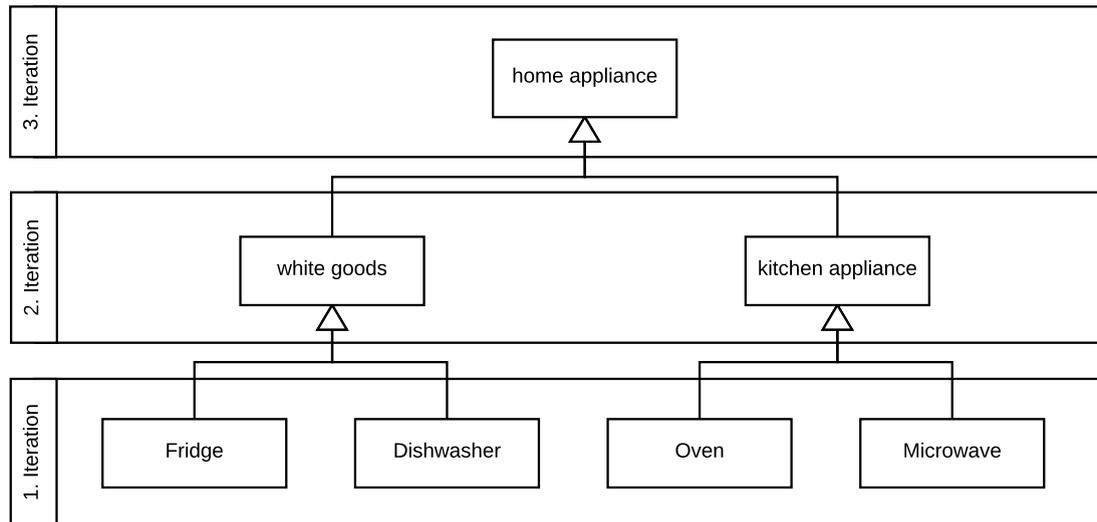


Abbildung 6.9.: Iterationsprozess der Überkonzeptbildung

abgeleitet wurden. So wurde das Teilkonzept *Fridge.Door* des Konzepts *Fridge* aus der Domänenontologie extrahiert und mit den möglichen Zuständen *opened* und *closed* versehen sowie das Synonyme-Konzept *move* zum Konzept *go* identifiziert.

Nachdem alle initial ableitbaren Konzepte identifiziert wurden, lässt sich daraufhin über alle gefundenen Objektkonzepte mit identifizierten Lemmas in WordNet iterieren und die in Abschnitt 5.2.3.3 beschriebene Lin-Metrik dafür nutzen gemeinsame Überkonzepte zu finden. Hierfür wird der niedrigste gemeinsame Überbegriff jeder Kombination der Objektkonzepte mittels *extJWNL* [Aut] bestimmt und daraufhin mittels der vorberechneten Informationsgehaltdateien des Pearl-Modules *WordNet::Similarity* [Ped] für die WordNet-Datenbank in Version 3.0 die Lin-Ähnlichkeit berechnet. Ist die Ähnlichkeit größer dem definierten Schwellwert 75%⁸, wird der niedrigste gemeinsame Überbegriff als Konzept eingefügt, sofern dieser nicht bereits existiert oder eines der beiden Objektkonzepte selbst darstellt. Ist letzteres der Fall, wird das als niedrigster gemeinsamer Überbegriff identifizierte Konzept als Überkonzept des anderen gesetzt. Somit lassen sich die Über- und Unterkonzepte der Objektkonzepte erweitern und eine mehrstufige Kontexthierarchie erstellen.

Da die Agenten auf mehrfache Ausführung ausgelegt sind, können durch dieses iterative Vorgehen bei jedem erneuten Ausführen des Agenten weitere Überkonzepte zwischen den in der vorherigen Ausführung hinzugefügten Überkonzepten entstehen und somit nach und nach die Hierarchie wachsen.

- (69) Go to the fridge next to the dishwasher and turn it on then go to the microwave and the oven and check if they are turned off

So wird mit diesem Ansatz die Hierarchie im Beispiel 69 zunächst nur die Überbegriffe *white goods* für *fridge* und *dishwasher* sowie *kitchen appliance* für *oven* und *microwave* enthalten. Im nächsten Schritt wird allerdings ebenso das Überkonzept *home appliance* für *white goods* und *kitchen appliance* gefunden. Dieser Ablauf ist in Abbildung 6.9 dargestellt.

⁸Die Voranalyse auf den Entitäten des Korpus ergab 75% als guten Wert für genügend spezifische Überkonzepte.

6.2.5. Zustandsbestimmer

Der Zustandsbestimmer (`EntityStateDeterminer`) ist der letzte Teilschritt der Kontextanalyse und setzt die in Abschnitt 5.2.3.4 beschriebene simple Heuristik um. Hierzu wird über alle sich im Kontext befindlichen Entitäten iteriert und überprüft, ob diese eine Objektentität darstellt, die einem Konzept zugeordnet ist, welches mögliche Zustände beschreibt oder es sich um eine Pronomenentität (`PronounEntity`) handelt, welche ein Objekt- oder Gruppenpronomen darstellt.

Im ersten Fall lässt sich zunächst überprüfen, ob eines der beschreibenden Adjektive der Entität mit einem der möglichen Zustände des Konzepts übereinstimmt, indem auch hier ein Ähnlichkeitsvergleich der Namen bzw. der Synonyme der Adjektive mit dem Namen des Zustands durchgeführt wird. Daraufhin wird überprüft, ob die Entität bereits einen mittels der Korreferenzanalyse identifizierten Referenten besitzt. Ist dies gegeben werden die Zustände dieser Entität ermittelt. Diese sollten bereits gegeben sein, denn es wird in Reihenfolge des Auftretens im Text über die Entitäten iteriert. Hierzu wird die letzte Pronomenentität gesucht, welche eine Korreferenzkette zum gefundenen Referenten bildet, die vor der gerade betrachteten Objektentität liegt. Dies ist nötig, um die tatsächlich letzten Zustände des Referenten zu ermitteln, ohne die in Abschnitt 2.4 bestimmte Definition der Objektidentitäten auf pronominale Anaphern auszuweiten. Ist eine solche Pronomenentität vorhanden, werden die aktuellen Zustände dieser als vorherige Zustände betrachtet. Ansonsten sind die vorherigen Zustände, die für den Referenten gefundenen.

Sind diese Informationen ermittelt, können Zustandsübergänge der Aktionen, an denen die betrachtete Entität als behandeltes Objekt teilnimmt, ermittelt werden, indem die Aktionen betrachtet werden. Existiert hierbei eine Aktion, welche mit einem Konzept verknüpft ist, das Zustandsübergänge beschreibt, lassen sich diejenigen Zustände, die mit den möglichen Zuständen der betrachteten Entität übereinstimmen, ermitteln. Hat man nun die vorherigen, die aus den Adjektiven extrahierten sowie die über die Aktion ermittelten Zustände, lassen sich diese zu den aktuellen Zuständen der Entität nach Ausführung der Aktionen zusammenführen. Dabei bleiben die vorherigen Zustände bestehen, die nicht durch die Adjektive oder die Zustandsübergänge der Aktionen widerlegt werden, also diejenigen die nicht einen assoziierten Zustand zum ermittelten Zustand aus den Adjektiven darstellen. Außerdem behält die Entität die Adjektivzustände, welche ebenso nicht durch die Zustandsübergänge der Aktionen widerlegt werden. Das heißt die ermittelten Zustände der Entität stellen die ermittelbaren Zustände nach der Ausführung der Aktionen an denen sie teilnimmt dar.

Im Fall einer Pronomenentität fällt der Schritt der Ermittlung von Adjektiven induzierten Zuständen weg, denn Personalpronomen werden nicht mit Adjektiven beschrieben. Außerdem muss keine komplexe Suche des vorherigen Referenten durchgeführt werden, da die Referenten von Pronomenentitäten immer die zuvorige Nennung der Entität darstellen. Das heißt die Zustände des vorherigen Gliedes in der Korreferenzkette werden als vorherige Zustände angenommen. Somit müssen nur die vorherigen mit den Zuständen aus den Aktionen erweitert oder widerlegt werden.

Befindet sich die Nennung der Entität im Bedingungsteil einer Anweisung, werden die Zustandsübergänge der Aktionen nicht betrachtet, sondern nur die vorherigen Zustände übernommen. Wie in Abschnitt 5.2.3.4 beschrieben bilden diese Informationen zumeist nur die für die Bedingung zu überprüfenden Zustände ab und stellen keine neuen Zustände dar.

Sind die Zustände einer Entität bestimmt, wird für jeden angenommenen Zustand eine Beziehung (`EntityStateRelation`) zwischen der Entität und dem Zustand abgebildet.

Nachdem alle Teilschritte der Kontextanalyse durchgeführt wurden und damit das aktuelle Abbild des Kontext generiert wurde, wird dieser zurück in den Graph geschrieben bzw.

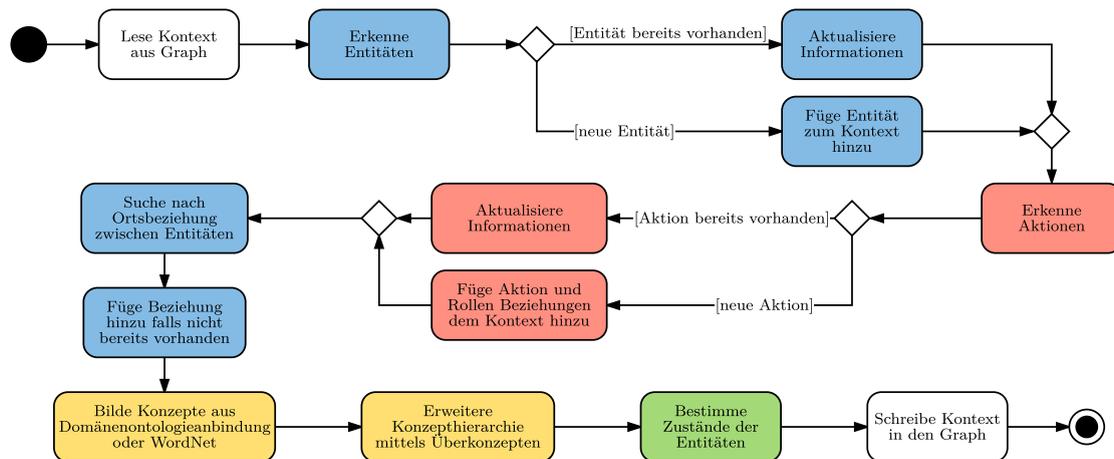


Abbildung 6.10.: Ablauf der Kontextanalyse

die Teile des Graphs aktualisiert, die sich zur letzten Ausführung des Kontextanalyserers geändert haben. Somit lässt sich der Ablauf der Kontextanalyse wie in Abbildung 6.10 dargestellt beschreiben.

6.3. Korreferenzanalyse

Zur Umsetzung der in Abschnitt 5.3 beschriebenen kontextbasierten Korreferenzanalyse wurde der Agent `CorefAnalyzer` erstellt. Hierzu wird ihm, wie dem `ContextAnalyzer`, das aktuelle Abbild des Graphen übergeben. Da es sich bei der Umsetzung um einen auf dem im Kontextanalyser erstellten Kontextwissen aufbauenden Ansatz handelt, wird zunächst der Graph nach einer Repräsentation des Kontextes durchsucht und falls vorhanden mittels der `readFromGraph`-Methode der `Context`-Klasse eingelesen.

Zunächst wurde ebenfalls in Betracht gezogen, sofern kein Kontextwissen vorliegt, erste Korreferenzinformationen mittels des *Stanford Deterministic Coreference Resolution Systems* (siehe Abschnitt 2.4.2) zu erzeugen und diese später für die kontextbasierte Korreferenzanalyse zu nutzen. Da allerdings die Ergebnisse des Stanford-Systems bedingt durch die inkorrekten Syntaxbäume nicht verlässlich verwendbar waren, wurde dieser Ansatz verworfen und nur auf den kontextbasierten Ansatz gesetzt. Da es aber auch ohne ein vorhandenes Kontextwissen möglich sein sollte erste Ergebnisse der Korreferenzanalyse generiert zu können, wurde eine Basis-Variante des Entitätenerkenners (siehe Abschnitt 6.2.1) in den Agenten eingebaut, welcher vorhandenen Entitäten bestimmt und als Kontext speichert. Damit jederzeit ein weiteres Korreferenzanalyse-Werkzeug, mit eventuell anderer Ausrichtung oder anderen Voraussetzungen, hinzugefügt werden kann, wurde die Schnittstelle `ICorefAnalyzer` eingeführt, welche die Methode `analyze(IGraph graph, Context context)` enthält. Die einzige momentane Implementierung dieser stellt aber der `FullContextAnalyzer` dar, in welchem die kontextbasierte Korreferenzanalyse durchgeführt wird. In diesem werden die drei in Abschnitt 5.3 identifizierten Arten von Korreferenzen einzeln auf den sie betreffenden Entitäten des Kontextes durchgeführt. Hierzu wurden drei die Schnittstelle `IResolution` mit der Methode `getCandidates(entity:Entity, graph:IGraph, context:Context):List<ReferentCandidate>` implementierende Analyse-Systeme eingeführt: Die `AnaphoraResolution`, die sich mit pronominaler Anapher-auflösung beschäftigt (siehe Abschnitt 5.3.3), die `SubjectIdentityResolution`, welches die Identitätsauflösung von Subjektentitäten übernimmt (siehe Abschnitt 5.3.5.2) und das `ObjectIdentityResolution`-System, dass die Identitätsauflösung der Objektentitäten durchführt (siehe Abschnitt 5.3.5.3).

Dabei werden jeweils die für das Analyse-System relevanten Entitäten nach ihrer Reihenfolge in der Äußerung sortiert und sukzessive zur Auflösung dem entsprechenden System übergeben, welches daraufhin die Kandidaten für eine Referenz der gerade betrachteten referierenden Entität zurückliefert.

Umgesetzt sind die Analyse-Systeme als mehrschichtige Filter, welche in Anlehnung an das von Stanford benutzte Mehrwegesieb als Siebe (**Sieve**) bezeichnet werden. Diese Wahl wurde getroffen, da bereits in der Analyse eine Anzahl von harten aber auch unsicheren Auswahlkriterien für Kandidaten der einzelnen referierenden Ausdrücke identifiziert wurden (siehe Abschnitt 5.3.4). Mithilfe der verschiedenen Eigenschaften erlaubt es die Siebarchitektur, schrittweise die Menge der Kandidaten für eine Referenz einzugrenzen. Es werden also durch die Siebe nach und nach Kandidaten „ausgesiebt“. Somit lässt sich pro Eigenschaft ein Sieb entwickeln, welches die Menge der Kandidaten hinsichtlich dieser Eigenschaft überprüft und entweder die Kandidatenmenge reduziert oder, im Fall einer unsicheren Eigenschaft, Wahrscheinlichkeiten für die Kandidaten vergibt. Die Wahl Wahrscheinlichkeiten zu benutzen wurde getroffen, um die unsicheren Eigenschaften trotzdem nutzen zu können, aber die potenziellen weiteren Kandidaten, die diese Eigenschaft nicht erfüllen ebenso in der Betrachtung und späteren Ausgabe zu behalten. Dies ermöglicht es zum einen, dass sich mehrere der Wahrscheinlichkeiten-vergebenden Siebe gegenseitig verbessern, indem ein Kandidat, der durch das eine Sieb als weniger wahrscheinlich eingestuft wurde, durch ein anderes wieder wahrscheinlicher wird. Zum anderen können so in einem anderen Teil des Projektes Fehler in den Referenzen entdeckt und z.B. durch ein Dialogsystem die weiteren möglichen Kandidaten überprüft werden. Außerdem bieten Wahrscheinlichkeiten für die Kandidaten die Möglichkeit abschließend eine Bewertung der gleich wahrscheinlichen Kandidaten mittels ihres Abstands zum betrachteten referierenden Ausdruck durchzuführen (Abschnitt 5.3.3.11). Somit reduzieren die einzelnen Siebe sukzessive die Anzahl der Kandidaten und bewerten diese Anhand ihrer Eigenschaften, was dazu führt, dass nach durchlaufen aller Siebe eine Wahrscheinlichkeitsverteilung für die möglichen Kandidaten entstanden ist, oder aber kein passender Kandidat existiert.

Die Siebarchitektur bietet zudem die Möglichkeit die einzelnen Siebe für mehrere Auflösungssysteme wiederzuverwenden sowie die durchgeführten Siebe und deren Reihenfolge mittels einer Konfigurationsdatei⁹ von außen einstellbar zu machen. Hierdurch lässt sich die Korreferenzanalyse variabel anpassen, was es erlaubt die Analyse auf die Voraussetzungen der Domäne zuzuschneiden und jederzeit weitere Siebe mit anderen Bewertungsschemata oder Eigenschaften einzuführen. Außerdem lässt sich in der Konfigurationsdatei festlegen, ob die Siebe in einem harten oder weichen Modus¹⁰ genutzt werden sollen. Im weichen Modus werden alle möglichen Kandidaten beibehalten, ihre Wahrscheinlichkeit aber auf Null gesetzt falls sie ausgeschlossen werden können. Im harten Modus hingegen werden die ausgeschlossenen Kandidaten komplett entfernt. Die Siebe, welche generell nur eine unpräzise Eigenschaft darstellen oder eine Bewertung durchführen, agieren immer im weichen Modus.

Somit besteht die Rückgabe der Analyse-Systeme aus einer Liste von Referenzkandidaten (**ReferentCandidate**), welche die Entität oder im Fall von mit einer Konjunktion verbundenen Entitäten, die zusammen betrachteten Entitäten, und ihre Wahrscheinlichkeit als möglichen Kandidaten enthält. Die Siebe implementieren die Schnittstelle **ISieve**, welche die Methode folgende Methode spezifiziert:

```
sieve(current:Entity, candidates>List<ReferentCandidate>)
:List<ReferentCandidate>
```

Diese erhält die aktuell betrachtete referierende Entität sowie die möglichen Referenzkan-

⁹Konfigurationsdatei liegt im Ressourcenverzeichnis der Implementierung (siehe Anhang Abschnitt C)

¹⁰SIEVE_MODUS = HARD oder SOFT

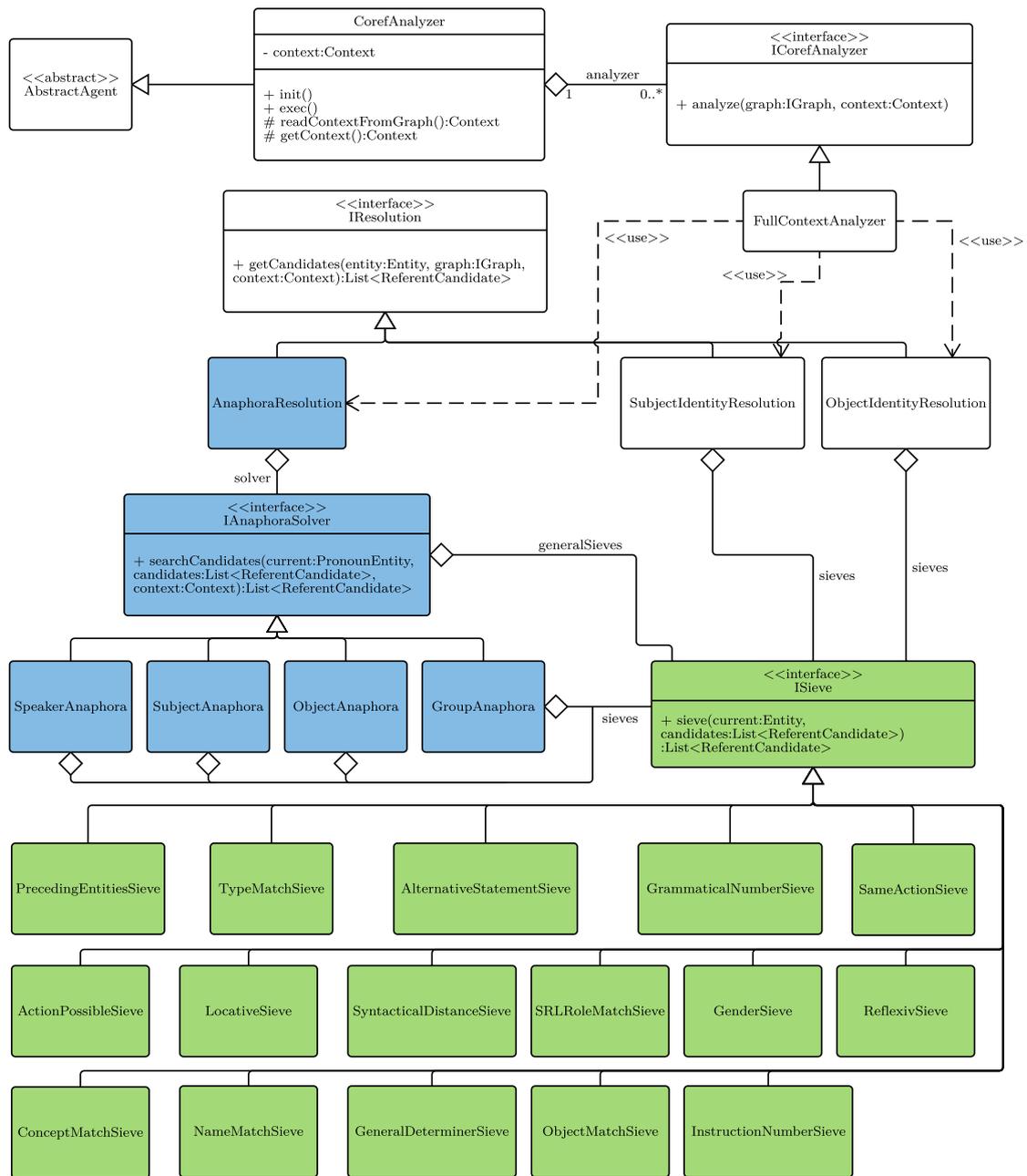


Abbildung 6.11.: Aufbau des Korreferenzanalyserers

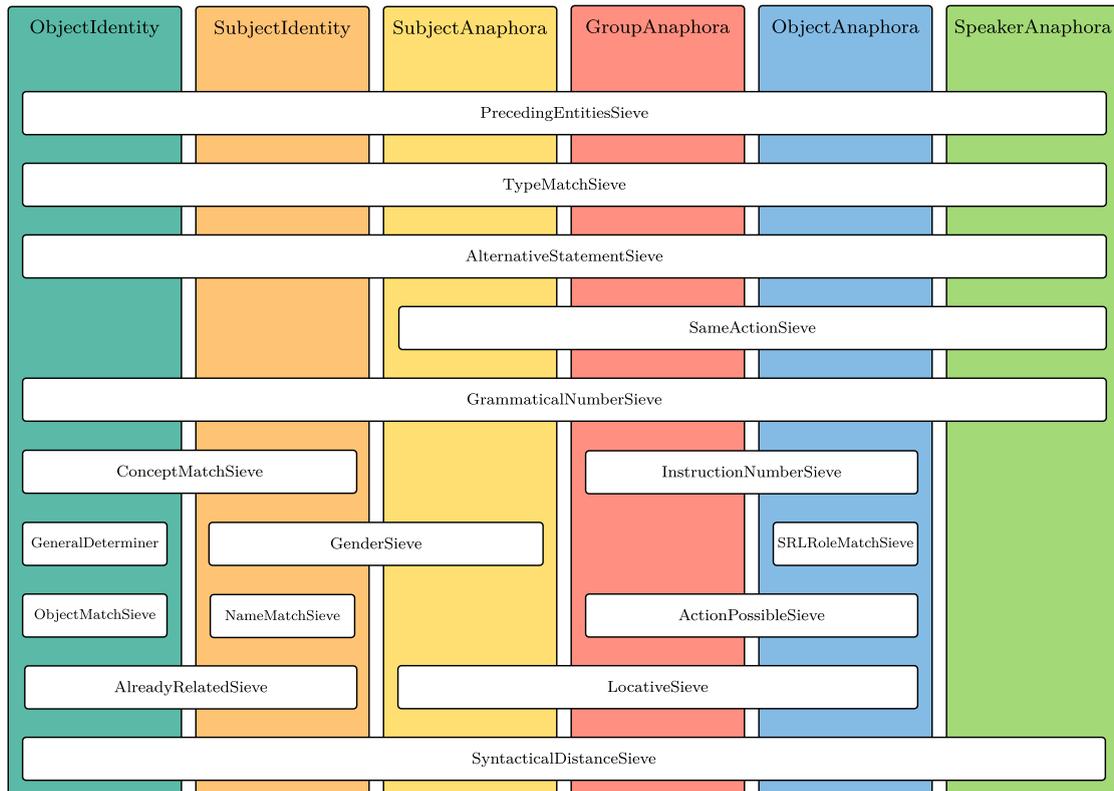


Abbildung 6.12.: Übersicht der Siebe und ihrer Reihenfolge für die einzelnen Analyse-Systeme

didaten und gibt die nach Ausführung des Siebs verbleibenden Referenzkandidaten mit ihren gesetzten Wahrscheinlichkeiten zurück. Die Siebe können mittels der Klasse `SieveFactory` und der in der Konfigurationsdatei spezifizierten ID instanziiert werden und sind somit jederzeit erweiterbar.

Bedingt durch die Pronomentyp-Unterscheidung in der pronominalen Anapherauflösung (siehe Abschnitt 5.3.3) ist das Analyse-System `AnaphoraResolution` in zwei Stufen unterteilt. Zunächst lassen sich generelle für alle Typen von Pronomen geltende Siebe definieren, welche zuerst ausgeführt werden. Daraufhin wird anhand des Pronomentyps eine Unterscheidung in unterschiedliche Anapherauflösungen durchgeführt. Hierzu wurde die Schnittstelle `IAnaphoraSolver` eingeführt, für die Ausprägungen für die einzelnen Pronomentypen existieren. Die Schnittstelle spezifiziert die Methode:

```
searchCandidate(current:PronounEntity,
                candidates:List<ReferentCandidate>,
                context:Context):List<ReferentCandidate>
```

So existieren dabei die Klassen `SpeakerAnaphora`, `SubjectAnaphora`, `ObjectAnaphora` und `GroupAnaphora`, welche jeweils die Methode der Schnittstelle implementieren und ihre eigenen spezifischen Siebe aus der Konfigurationsdatei auslesen. Somit ergibt sich der Aufbau des Agenten wie in Abbildung 6.11 dargestellt. Die in dieser Arbeit eingesetzten Siebe sowie die Reihenfolge in der sie von den einzelnen Analyseschritten verwendet werden, ist in Abbildung 6.12 abgebildet, wobei oben der Anfang der Auflösung und unten das Ergebnis nach der Auflösung dargestellt ist. Eine Übersetzung der in Abschnitt 2.4 identifizierten Eigenschaften auf die einzelnen Siebe ist in Tabelle 6.1 gegeben, sowie die Information, ob es sich bei dem Sieb um ein hartes oder weiches Sieb handelt.

¹¹ Abhängig vom Abstand zum referierenden Ausdruck

Eigenschaft	Sieb	Modi
Vorangegangene Entitäten (Abschnitt 5.3.3.1)	PrecedingEntitiesSieve	HARD
Typübereinstimmung (Abschnitt 5.3.3.6)	TypeMatchSieve	HARD
Numerus (Abschnitt 5.3.3.2)	GrammaticalNumberSieve	HARD
Bedingungen (Abschnitt 5.3.3.4)	AlternativeStatementSieve	HARD
Gleiche Aktion (Abschnitt 5.3.3.5)	SameActionSieve	HARD
Bezugsrahmen (Abschnitt 5.3.3.3)	InstructionNumberSieve	HARD
Geschlecht (Abschnitt 5.3.3.7)	GenderSieve	HARD
Aktion ausführbar (Abschnitt 5.3.3.8)	ActionPossibleSieve	SOFT mit Gewicht 0.66
Ortsbeziehung (Abschnitt 5.3.3.9)	LocativeSieve	SOFT mit Gewicht 0.25
Übereinstimmung semantischer Rollen (Abschnitt 5.3.3.10)	SRLRoleMatchSieve	SOFT mit Gewicht 0.7
Abstand (Abschnitt 5.3.3.11)	SyntacticalDistanceSieve	SOFT mit abhängigem Gewicht ¹¹
Konzeptübereinstimmung (Abschnitt 5.3.5.1)	ConceptMatch	HARD
Unspezifischer Determinierer (Abschnitt 5.3.5.3)	GeneralDeterminerSieve	HARD
Übereinstimmung der Objekteigenschaften (Abschnitt 5.3.5.3)	ObjectMatchSieve	HARD
Namensübereinstimmung (Abschnitt 5.3.5.2)	NameMatchSieve	HARD
Aussieben von bereits in Beziehung stehenden Entitäten	AlreadyRelatedSieve	HARD
Bezugsrahmen von Reflexivpronomen (Abschnitt 5.3.3.12)	ReflexivSieve	HARD
Teil-Ganzes-Beziehung Possessivpronomen (Abschnitt 5.3.3.12)	PossessiveMeronymSieve	SOFT mit Gewicht 0.75

Tabelle 6.1.: Umgesetzte Eigenschaften und die entsprechenden Siebe

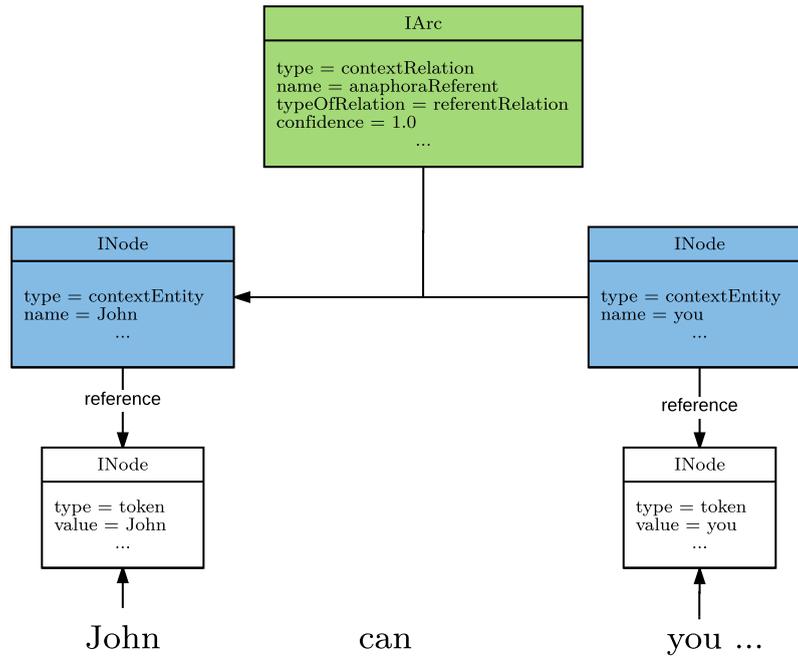


Abbildung 6.13.: Beispielhafte Repräsentation der Referenzen im Graph

Letztlich werden die pro Entität gefundenen Kandidaten, welche eine höhere Wahrscheinlichkeit als 0.0 erhalten haben, als Referenzbeziehung (**ReferentRelation**) den in Beziehung stehenden Entitäten hinzugefügt, welche die identifizierte Wahrscheinlichkeit enthält und über die `printToGraph` Methode der `Context`-Klasse in den Graph geschrieben werden kann. Somit ist eine Korreferenzbeziehung im Graph als eine Kante vom Typ `contextRelation` mit Namen `referent` repräsentiert die zwischen den beiden Knoten verläuft, welche die Entitäten repräsentieren (siehe Abbildung 6.13).

6.3.1. Pronominale Anapherauflösung

Wie zuvor bereits erläutert, wurde die pronominale Anapherauflösung in zwei Schritte unterteilt, eine generelle Verarbeitung der Kandidaten sowie eine darauf aufbauende für den Pronomentyp spezifische Behandlung. Hierbei werden die zu betrachtenden Pronomenentitäten einzeln nach ihrem Auftreten in der Äußerung sortiert an das Analyse-System übergeben und die Kandidaten jeweils einzeln für diese bestimmt und zurückgegeben. Zunächst werden dabei die generellen Siebe auf den Kandidaten ausgeführt, welche die Kandidatenmenge einschränken. Daraufhin wird anhand des in der Pronomenentität enthaltenen Namens des Pronomens eine Bestimmung des Pronomentyps durchgeführt. Diese geschieht anhand eines Abgleichs mit den in Tabelle 5.2 dargestellten Wortlisten. Hierbei werden Personal-, Reflexiv- und Possessivpronomen gleich behandelt, was es erlaubt die Struktur des Analyse-Systems für alle drei Formen gleich zu belassen. Eine Anpassung auf die besonderen Eigenschaften der Reflexivpronomen (siehe Abschnitt 5.3.3.12) ist dabei nur im Sieb *gleiche Aktion* (**SameActionSieve**) nötig, da die umgesetzte Eigenschaft nicht für Reflexivpronomen gilt. Außerdem wurde zusätzlich ein generelles Sieb eingeführt, welches nur im Fall von Reflexivpronomen den geringeren Bezugsrahmen dieser umsetzt (**ReflexiveSieve**). Die Possessivpronomen hingegen werden im Zuge der Objektidentitätsauflösung der Objektentität durchgeführt, zu der sie einen beschreibenden Aspekt darstellen. Genauereres hierzu folgt in Abschnitt 6.3.3.

Nach der Bestimmung des jeweiligen Pronomentyps wird das für den Typ zuständige Analyse-System (Ausprägung des **IAphoraSolver**) aufgerufen, welches die in der Kon-

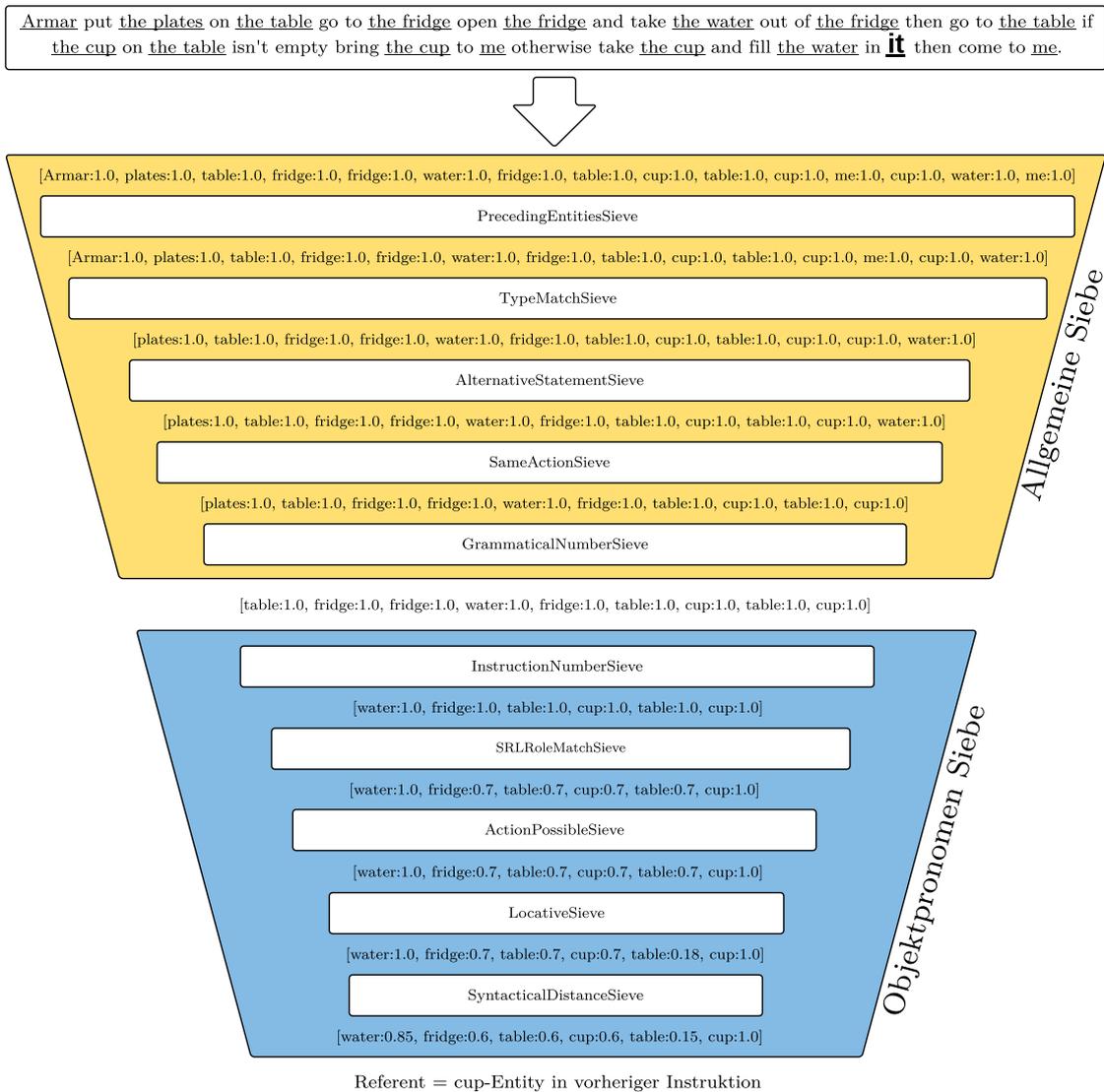


Abbildung 6.14.: Beispiel einer Objektpronomen-Anapherauflösung mit den entsprechenden Sieben

figurationsdatei beschriebenen spezifischen Siebe für den entsprechenden Pronomentyp durchführt. Hierbei stellt die Auflösung von Sprecherpronomen (**SpeakerAnaphora**) einen Sonderfall dar, da wie bereits in Abschnitt 5.3.3 beschrieben, Sprecherpronomen in Anweisungen zumeist keinen Referenten in der Äußerung besitzen und deshalb eine virtuelle Sprecherentität (**SpeakerEntity**) eingeführt wird auf die sich das erste auftretende Sprecherpronomen beziehen kann. Eine weitere leicht abweichende Behandlung wird außerdem in der Auflösung der Gruppenpronomen (**GroupAnaphora**), noch vor Ausführung eines der spezifischen Siebe, durchgeführt. Gruppenpronomen können sich nicht nur auf einzelne Entitäten, sondern ebenso auf mit Konjunktionen verbundene Entitäten, beziehen. Deshalb werden die nach den allgemeinen Sieben verbliebenen Kandidaten nach Konjunktionsbeziehungen zwischen ihnen untersucht. Falls Beziehungen gefunden wurden, werden die verbundenen Kandidaten unter einem Referenzkandidaten (**ReferentCandidate**) zusammengefasst, was es mit der Siebarchitektur erlaubt diese wie einen einzigen Kandidaten zu behandeln und ihnen die gleiche Wahrscheinlichkeit zuzuweisen. Deshalb können aus einem Referenzkandidaten mehrere Korreferenzkanten entstehen.

In Abbildung 6.14 ist der Ablauf der pronominalen Anapherauflösung des Pronomen *it* mit den Ergebnissen der entsprechenden Siebe im Modus HARD dargestellt. Hierbei wird das Pronomen *it* in der Anweisung „Armar put the plates on the table go to the fridge open the fridge and take the water out of the fridge then go to the table if the cup on the table isn't empty bring the cup to me otherwise take the cup and fill the water in it then come to me“ aufgelöst. Zunächst werden dabei alle Entitäten betrachtet und als Referenzkandidaten mit der Wahrscheinlichkeit 1.0 versehen. Daraufhin sortiert das erste Sieb die nach der betrachteten Entität *it* folgenden Pronomen aus. Daraufhin werden vom *Typübereintimmungssieb* (**TypeMatchSieve**) alle Entitäten aussortiert, die nicht entweder ein Objektpronomen oder eine Objektentität sind, wodurch *Armar* und *me* ausgesiebt werden. Anschließend werden die Entitäten vom *Bedingungssieb* (**AlternativeStatementSieve**) aussortiert, die durch den Bedingungsanalyse-Agenten als in einem alternativen Ausführungszweig zur betrachteten Entität stehend identifiziert wurden, was in diesem Fall die Entität *cup* aus „bring the cup to me“ ist. Da es sich bei dem *it* um kein Reflexivpronomen handelt, werden die Entitäten ausgeschlossen, die an der gleichen Aktion teilhaben wie die Entität *it*. Dies ist nur die Entität *water*. Daraufhin werden die Entitäten hinsichtlich ihres Numerus überprüft. Da die Entität *plates* im Plural steht, aber *it* ein Singular-Pronomen darstellt, wird diese ausgesiebt. Anschließend werden die Kandidaten bezüglich ihres Bezugsrahmens und damit ihrer Instruktionsnummer im Verhältnis zum betrachteten referierenden Ausdruck untersucht. Die Entitäten *table*, *fridge* und *fridge* aus den ersten drei Instruktionen werden ausgeschlossen, da diese mehr als 6 Instruktionen vom betrachteten Ausdruck entfernt stehen. Hierauf folgen nun die weichen Siebe. Zuerst werden im *Semantische-Rollen-Sieb* (**SRLRoleMatchSieve**) die semantischen Rollen der Kandidaten mit der Rolle des Pronomen abgeglichen. Hierbei haben nur die Entitäten *water* und *cup* ebenso die semantische Rolle *Theme*, welche auch die Entität *it* innehat, was dazu führt, dass alle anderen Entitäten eine um 30% verringerte Wahrscheinlichkeit erhalten. Das Sieb *Aktion ausführbar* (**ActionPossibleSieve**) hingegen hat in diesem Fall keine Auswirkung, da keine Information über einen Zustand der Entitäten mit der simplen Heuristik aus der Kontextanalyse bestimmt werden kann. Dahingegen führt das *Ortsbestimmungssieb* (**LocativeSieve**) dazu, dass die letztere *table*-Entität als Beschreibung der vorherigen *cup*-Entität angesehen wird und somit eine um 75% verringerte Wahrscheinlichkeit erhält. Abschließend werden dann die verbleibenden Entitäten auf gleich wahrscheinliche Kandidaten mit der höchsten Wahrscheinlichkeit untersucht und entsprechend ihres Abstands zum referierenden Ausdruck als geringer wahrscheinlich eingestuft. Die Wahrscheinlichkeiten der weiteren Entitäten werden mit dem für die letzte zum gleich wahrscheinlichen Block gehörenden Entität verwendeten Reduktionwert verringert. Hieraus ergibt sich dann die Entität *cup*, welche am nächsten zum referierenden Ausdruck steht, als wahrscheinlichster Kandidat für eine

Referenz.

Da Pronomen immer eine Referenz auf eine bereits im sprachlichen Kontext genannte Entität bilden (siehe Abschnitt 2.4) wird zusätzlich nach der Ausführung jeder für den Pronomentyp spezifischen Siebstufe überprüft, ob keine Kandidaten mehr für die Referenz übrig bleiben. Ist dies der Fall wird das zuletzt ausgeführte Sieb ignoriert und mit den Kandidaten des vorherigen Siebes fortgesetzt. Ohne diese Funktion würde sich zwar die Präzision des Werkzeugs erhöhen, dafür aber die Ausbeute deutlich sinken. Dies möchte man vermeiden, da in späteren Schritten des Projektes somit die eventuell nicht korrekten Kandidaten z.B. durch ein Dialogsystem validiert werden können. Ohne die möglichen Kandidaten wäre dies nicht umsetzbar.

6.3.2. Subjektidentitätsauflösung

Für die Umsetzung der Subjektidentitätsauflösung (siehe Abschnitt 5.3.5.2) können viele der in der Anapherauflösung verwendeten Siebe wiederverwendet werden, da die Eigenschaften in vielen Fällen ähnlich sind. Einige der Siebe werden aber bewusst nicht verwendet, da sie unsichere Eigenschaften abdecken, die für eine Auflösung von Identitäten, nicht benötigt werden. Diese sind durch die weniger mehrdeutigen Kandidaten bzw. die spezifischeren Informationen der referierenden Ausdrücke auch ohne diese Siebe gut auflösbar. Daher wird das Sieb *gleiche Aktion* sowie die weichen Siebe, bis auf das *Abstandsbewertungssieb*, für die Identitätsauflösungen nicht verwendet, da die von ihnen abgedeckten Eigenschaften nicht für die Identitäten gelten oder zu unsicher sind. Dafür wird das Geschlechtsinformationssieb aus der Subjektpronomen-Auflösung wiederverwendet und zusätzlich die in Abschnitt 5.3.5.1 beschriebene Konzeptübereinstimmung im Konzeptübereinstimmungssieb (`ConceptMatchSieve`) überprüft. Für die Subjektidentität werden des Weiteren unterschiedliche Benamungsvarianten einer Person im *Benennungssieb* (`NameMatchSieve`) überprüft. Außerdem werden für beide Identitätsauflösungen im *Bereits-Verbunden-Sieb* (`AlreadyRelatedSieve`) diejenigen Kandidaten entfernt, die schon als Referent einer ebenso unter den Kandidaten befindlichen Entität bekannt sind. Dies ist sinnvoll, um nicht durch Fehler inkonsistente Korreferenzketten zu erzeugen. Letztlich bezieht also das Analyse-System `SubjectIdentityResolution` die in Abbildung 6.12 dargestellten Siebe aus der Konfigurationsdatei und grenzt die Kandidaten dabei analog zur pronominalen Anapherauflösung ein.

6.3.3. Objektidentitätsauflösung

Auch die Umsetzung der Objektidentitätsauflösung (siehe Abschnitt 5.3.5.3) in der Klasse `ObjectIdentityResolution` bezieht die Siebe aus der Konfigurationsdatei und verwendet viele der Siebe wieder (vgl. Abbildung 6.12). Allerdings wird hier nicht nach dem Geschlecht ausgesiebt, wie bei der Subjektidentitätsauflösung, sondern zunächst eine Betrachtung der Determinierer, wie in Abschnitt 5.3.5.3 beschrieben, durchgeführt. Außerdem wird der ebenso dort beschriebene Abgleich der Adjektive, Quantitäten, Namen bzw. Synonyme und Hyperonyme durchgeführt (`ObjectMatchSieve`). Hierbei werden nur dann die Hyperonyme der aktuell betrachteten Objektentität für die Betrachtung herangezogen, wenn kein Kandidat über einen direkten Namensabgleich oder eines seiner Synonyme gefunden werden konnte. Dies sorgt dafür, dass die nicht immer präzisen Hyperonyminformationen aus WordNet nur dann verwendet werden, falls es keinen präziseren Kandidaten gibt.

Des Weiteren wird im Objektidentitätsauflöser auch die Auflösung der Possessivpronomen durchgeführt. Da allerdings die Auflösung der Possessivpronomen zu großen Teilen, der von Personalpronomen gleicht, wird hierzu das Analyse-System zu pronominalen Anapherauflösung (`AnaphoraResolution`) wiederverwendet. Dazu wird eine Unterklasse der

Klasse `PronounEntity` eingeführt, die `PossessivPronounEntity`, welche zusätzlich eine Referenz auf die Objektentität enthält, dessen Beschreibung das Possessivpronomen darstellt. Dies ermöglicht es die meisten der Siebe für die Personalpronomenauflösung ohne Anpassung wiederzuverwenden, außer die Siebe `SameActionSieve`, `SRLRoleMatchSieve` und `ActionPossibleSieve`, welche im Fall von Possessivpronomen keine Kandidateneigenschaft abdecken. Zusätzlich wird das Sieb `PossessiveMeronymSieve` als letzter Schritt der Objekt- und Gruppenpronomenauflösung eingeführt. In diesem wird überprüft, ob die vom Possessivpronomen näher beschriebene Entität ein Teilkonzept der Kandidatenentität darstellt (siehe Abschnitt 5.3.3.12).

Auch bei der Identitätsauflösung ist also das Ergebnis eine Liste von Kandidaten die letztlich an den `FullContextAnalyzer` zurückgegeben wird und dort in Referenzbeziehungen (`ReferentRelation`) überführt wird. Abschließend wird dann der von der Korreferenzanalyse (`CorefAnalyzer`) um die Referenzbeziehungen erweiterte Kontext wieder in den Graph zurückgeschrieben bzw. die Teile aktualisiert, die sich in der Analyse verändert haben.

7. Evaluation

In diesem Kapitel wird überprüft, ob und wie gut die beiden Agenten (`ContextAnalyzer`, `CorefAnalyzer`) die in der Analyse identifizierten Aufgaben des Kontextaufbaus und der Auflösung der Korreferenzen umsetzen. Hierzu wird zunächst die Kontextanalyse mit Testeingaben ausgeführt und überprüft, ob der aufgebaute Kontext den Erwartungen entspricht sowie um eine bewertbare und damit fundierte Überprüfung durchführen zu können. Anschließend wird die Korreferenzanalyse mit dem aufgebauten Kontext betrachtet und anhand händisch erstellter Musterlösungen bewertet. Die Korreferenzanalyse wird hierbei anhand der Metriken Präzision, Ausbeute und F1-Maß evaluiert und mit dem *Stanford Deterministic Coreference Resolution System* (siehe Abschnitt 2.4.2) als Referenzsystem verglichen. Außerdem wird der Einfluss des Kontextes bewertet indem das entwickelte Korreferenzsystem mit einer Variante seiner selbst ohne ein umfassendes Kontextwissen durchgeführt wird.

In den folgenden Abschnitten werden der Evaluationskorpus und das Vorgehen bei der Bewertung der Ausgaben genauer erläutert sowie abschließend die Ergebnisse diskutiert.

7.1. Evaluationskorpus

Um die entstandenen Werkzeuge evaluieren zu können, wird eine Basis von möglichen Anweisungen an ein System, für welches Quelltext erzeugt werden soll, benötigt. Hierzu wurde bereits in der Arbeit von Günes [Gün15] ein Sprachkorpus mit 22 Sprachaufnahmen pro Szenario und entsprechenden händischen Transkriptionen des Gesagten erstellt. Des Weiteren wurden aufbauend auf den verwendeten Szenarien weitere 14 Sprachaufnahmen pro Szenario gesammelt, welches die Anzahl der für diese ersten drei Szenarien zur Verfügung stehenden Aufnahmen auf je 36 angehoben hat. Da diese Szenarien allerdings nicht die Verwendung von Bedingungen, welche in Anweisungen an Systeme auftreten können, unterstützt haben, wurden in der Arbeit von Steurer [Ste16] zwei weitere Szenarien mit je 19 Probanden aufgenommen, welches den Gesamtkorpus auf 146 Aufnahmen und Transkriptionen erweitert hat. Dieser Korpus dient als Bewertungsgrundlage zur Evaluation, ist allerdings für eine umfassende Analyse des Kontextes und eine Evaluation der Korreferenzanalyse nur bedingt geeignet, da vor allem die ersten drei Szenarien relativ kurze Beschreibungen mit wenigen Entitäten und damit einfachen Korreferenzen erzeugt haben. Deshalb wurden für die Evaluation dieser Arbeit zwei weitere Szenarien erstellt, welche sich zwar an den Voraussetzungen aus der Arbeit von Günes orientieren, aber deutlich mehr Aktionen erfordern und außerdem mehr Entitäten miteinbeziehen. So wurde versucht die Szenarien so zu entwickeln, dass sie die Benutzung von Pronomen, durch viele

Tabelle 7.1.: Informationen über neue Sprachaufnahmen

	Sz. Fill the cup	Sz. Prepare a meal	Gesamt
Anzahl Aufnahmen	10	10	20
Anzahl Wörter Transkriptionen	734	811	1545
Anzahl Entitäten	199	233	432
Anzahl Aktionen	120	154	274
Anzahl Korreferenzen	96	138	234

Aktionen auf gleichen Objekten, fördern und viele verschiedene Objekte und Aktionen für ihre Ausführung benötigen. So kann gezeigt werden, wie ein Aufbau von Kontextwissen bei dem Verständnis gerade längerer Anweisungssequenzen helfen kann Zusammenhänge und Aussagen zu verstehen.

Die hierbei entstandenen Szenarien sind in Abschnitt D des Anhangs dargestellt. Szenario *Fill the cup* erfordert es den Roboter Armar anzuweisen einen grünen Becher mit Wasser aus dem Kühlschrank zu füllen und die roten Becher aus der Spülmaschine in den Schrank zu räumen. Im Szenario *Prepare a meal* hingegen soll Armar einen Teller aus der Spülmaschine holen, ihn abwaschen und ihn mit einem Fertiggericht aus dem Kühlschrank füllen. Dieses soll er dann in der Mikrowelle erhitzen und auf den Tisch stellen.

Die Anweisungen der Probanden zu den beiden Szenarien wurden daraufhin von 10 Probanden aufgenommen. Die Aufnahme eines Probanden wurde auf seinen Wunsch wiederholt, da er eine der im Szenario geforderten Aktionen vergessen hatte, was dazu führt, dass 21 Sprachaufnahmen erstellt wurden. Anschließend wurden die entstandenen 21 Sprachaufnahmen anhand der in der Arbeit von Günes [Gün15] definierten Transkriptionsregeln in händisch transkribiert. Diese können als zusätzliche Evaluationsgrundlage für die Evaluation der beiden Werkzeuge genutzt werden, ohne von den Fehlern eines Spracherkenners abhängig zu sein.

In Tabelle 7.1 wird ein Überblick über die Daten der Szenarien *Fill the cup* und *Prepare a meal* dargestellt. Diese 21 Transkriptionen enthalten dabei nur 60 Korreferenzen weniger als die 146 Transkriptionen zuvor, was darauf hindeutet, dass die Konzeption der Szenarien das Ziel mehr Korreferenzen zu fördern erfüllt hat. Außerdem sind die Transkriptionen durch das komplexere Szenario im Durchschnitt deutlich länger als die zuvorigen Texte und bieten eine angemessene Zahl von Entitäten (hierbei sind auch Pronomen eingeschlossen) und Aktionen zur Evaluation der Kontextanalyse.

7.2. Vorgehensweise

Das Vorgehen bei der Evaluation der beiden Werkzeuge unterscheidet sich leicht: Für die Kontextanalyse lassen sich nur schwer vollständige Musterlösungen erstellen, da eine Definition von vollständigem Kontext eines Textes nur schwer zu bestimmen ist. Außerdem ist eine Betrachtung des Korpus hinsichtlich der Kontextanalyse sehr aufwendig. Aus diesem Grund wird für die Evaluation der Kontextanalyse nur auf die Transkriptionen der beiden neuen Szenarien zurückgegriffen. Für die Korreferenzanalyse hingegen lassen sich leicht Musterlösungen für den ganzen Korpus erstellen. Somit lässt sich auch das Werkzeug auf dem vollständigen Korpus evaluieren.

Um die Kontextanalyse zu evaluieren werden also die Transkriptionen der beiden neuen Szenarien mit dem Vorverarbeitungsfießband verarbeitet und daraufhin mit den Musterlösungen der Korreferenzen versehen, um auch eine vollständige Evaluation der Zustandsbestimmung (siehe Abschnitt 6.2.5) durchführen zu können. Daraufhin wird außerdem der Bedingungserkennungs-Agent (siehe Abschnitt 3.3) ausgeführt um die Informationen über Bedingungssätze bei der Evaluation nutzen zu können. Außerdem wird die in Abschnitt B.2

des Anhangs dargestellte Version der Domänenontologie als Grundlage verwendet. Abschließend wird der Kontextanalyse-Agent sooft ausgeführt, bis der resultierende Kontext sich nicht mehr verändert und dieses Ergebnis dann hinsichtlich Fehlern und Unvollständigkeiten bewertet. Hierbei können allerdings nicht alle Aspekte hinsichtlich Vollständigkeit überprüft werden, da diese für manche Informationen nicht eindeutig bestimmbar sind. Daher werden die einzelnen Teilschritte der Kontextanalyse hinsichtlich unterschiedlicher Aspekte evaluiert.

Allgemein werden die Metriken Präzision, Ausbeute und das F1-Maß zur Evaluation herangezogen. Die Präzision (*Precision*) gibt hierbei an wie korrekt die gelieferten Ergebnisse des Werkzeugs zu einem Aspekt sind. Sie wird mit der folgenden Formel berechnet:

$$\text{Präzision} = \frac{tp}{tp + fp} \quad (7.1)$$

Hierbei beziffert tp (*true positive*) die Ergebnisse, die richtig eingeordnet wurden und fp (*false positives*) die Ergebnisse, welche fälschlicherweise in die betrachtete Kategorie eingeordnet wurden. Was dies in der eigentlichen Evaluation eines Aspekts bedeutet, wird in den entsprechenden Evaluationskapiteln genauer erläutert.

Die Ausbeute (*Recall*) hingegen gibt an wie viele der einzuordnenden Informationen letztlich durch das Werkzeug eingeordnet wurden. Die Formel hierzu ist:

$$\text{Ausbeute} = \frac{tp}{tp + fn} \quad (7.2)$$

Die fn (*false negative*) sind dabei die Ergebnisse, die nicht in die erwartete Kategorie eingeordnet wurden.

$$F1 = 2 * \frac{\text{Präzision} * \text{Ausbeute}}{\text{Präzision} + \text{Ausbeute}} \quad (7.3)$$

Das F1-Maß (*F1-Score*) ist das harmonischen Mittel zwischen Präzision und Ausbeute und bewertet damit das Zusammenspiel von Präzision und Ausbeute.

Für die Evaluation der Korreferenzanalyse wird der gleiche Vorgang wie in der Kontextanalyse durchgeführt, nur das diesmal der gesamte Korpus betrachtet wird und keine Korreferenzinformationen manuell eingetragen werden. Vielmehr werden der Kontextanalyse- und der Korreferenzanalyse-Agent abwechselnd solange ausgeführt, bis sich der als Ergebnis resultierende Kontext nicht mehr ändert. Dieses Ergebnis wird dann hinsichtlich der gefundenen Korreferenzketten mit den Musterlösungen verglichen. Der Aufbau der Musterlösungen ist in Abschnitt E des Anhangs erläutert.

7.3. Kontextanalyse

Die Evaluation der Kontextanalyse wird in die im Entwurf und der Implementierung herausgestellten Teilschritte unterteilt. Hierbei wird in Evaluation des Entitätenerkenners (siehe Abschnitt 6.2.1) evaluiert, ob die im sprachlichen Kontext enthaltenen Entitäten korrekt erkannt werden. In der Evaluation des Aktionserkenners (siehe Abschnitt 6.2.2) wird überprüft, ob die beschriebenen Aktionen erkannt und die extrahierten Informationen zu den semantischen Rollen korrekt sind. Die Evaluation des Ortsbeziehungserkenners (siehe Abschnitt 6.2.3) beschäftigt sich mit der Bewertung der erkannten Ortsbeziehungen und die der Welt- und Domänenwissenanreicherung (siehe Abschnitt 6.2.4) damit, ob die extrahierten Konzepte korrekt zugeordnet und die in ihnen enthaltenen Informationen korrekt sind. Letztlich wird der Zustandserkenners (siehe Abschnitt 6.2.5) hinsichtlich der Korrektheit der zu erwartenden Zustände evaluiert.

Tabelle 7.2.: Ergebnisse der Teilschritte der Kontextanalyse

Teilaspekt	Präzision	Ausbeute	F1-Maß
Entitätenerkennung	0,972	0,975	0,973
Aktionserkennung	0,852	0,762	0,804
Ortsbeziehungsanreicherung	0,945	0,793	0,862
Konzeptbildung	0,986	0,974	0,981
Konzeptbeziehungsbildung	0,683	0,845	0,756
Zustandsbestimmung	0,854	0,627	0,723

7.3.1. Entitätenerkennung

Der Entitätenerkennung wird hinsichtlich seiner Leistung in Bezug auf das Erkennen von Entitäten im sprachlichen Kontext bewertet. Dies schließt zum einen die korrekte Anzahl und Ausprägung der die Entität bildenden Wörtern sowie zum anderen die für die Textstelle spezifischen Informationen wie einen sinnvollen Namen der Entität, den Numerus, die Form des Determinierers oder die erkannten Adjektive mit ein. Zusätzlich wird überprüft, ob Konjunktionsbeziehungen zwischen Entitäten korrekt erkannt werden. Somit ergibt sich eine erkannte Entität als richtig positives Ergebnis (*true positive*), wenn sie sowohl den korrekten Bezug als auch korrekte Informationen über die Textstelle enthält. Dahingegen sind falsch positive Ergebnisse (*false positive*) dieser Betrachtung diejenigen Entitäten, die falsche Informationen oder Bezüge enthalten und falsch negative Ergebnisse (*false negative*) Entitäten, die zu erwarten gewesen wären aber nicht erkannt wurden.

Anhand dieser Einordnung ergeben sich die in Tabelle 7.2 dargestellten Werte für Präzision (0,972), Ausbeute (0,975) und F1-Maß (0,973). Es lässt sich feststellen, dass die Erkennung der Entitäten sehr zuverlässig funktioniert und sowohl Präzision als auch vor allem die Ausbeute sehr hoch ist. Dies ist deshalb wichtig, da für den Aufbau des Kontextes gerade das Erkennen aller auftretenden Entitäten wichtig ist, um darauf aufbauend die weiteren Teile des Kontextwissens zu generieren. Die zwei gänzlich nicht erkannten der 432 Entitäten resultieren aus Formulierungen wie z.B. der Wortwiederholung „Hallo Armar hello Armar“, welche durch den Phrasenerkennung als eine Nominalphrase erkannt wird und durch den Entitätenerkennung zu einer *Armar*-Entität zusammengefasst wird, anstatt zwei Entitäten wie es korrekt gewesen wäre. Damit tritt dieser Fehler ebenso als falsch positives Ergebnis im Resultat auf, da die Referenz der erkannten *Armar*-Entität über beide erwarteten Entitäten einschließt. Es wurde also die Referenz [Hallo, Armar, hello, Armar] erkannt anstatt der erwarteten [Armar] und [Armar] für zwei unterschiedliche Entitäten. Die weiteren nicht erkannten Entitäten resultieren aus falsch erkannten Entitäten. Dies sind Entitäten, die zwar zum Teil aus einem richtigen Bezug gebildet wurden, aber ebenso falsche Bezugswörter beinhalten. Diese fallen also ebenso in die Kategorie nicht erkannt (falsch negativ), da die erwartete Entität nicht in der erwarteten Form erkannt wurde, als auch falsch positiv, da eine Entität erkannt wurde die so nicht erwartet wurde. Die weiteren falsch positiven Ergebnisse resultieren zumeist aus Fehlern des Wortartmarkierers und des Phrasenerkenners und stellen vor allem als Nomen erkannte Verben dar, welche daraufhin als Entitäten erkannt werden. So wird z.B. *press* in „[...] and *press* start“ als Nominalphrase deklariert und damit die Entität *press* generiert sowie *start* als Verbalphrase und Verb und damit als Aktion eingeordnet obwohl dies ebenfalls nicht korrekt ist.

Ein weiterer Fehler der auftritt sind inkorrekte Namen der Entitäten durch Versprecher der Probanden, wie es bei „the mac microwave“ der Fall ist. Hierbei wird die Entität *mac microwave* basierend auf den Nomen der Nominalphrase erstellt. Diese Entitäten spiegeln aber immer noch die korrekte Textstelle wieder und können somit für den weiteren Kontextaufbau genutzt werden. Außerdem existiert ein Fall in dem eine Entität in einen falschen Typ eingeordnet wurde. Bei der Nominalphrase „the Apollinaris bottle“ wur-

Tabelle 7.3.: Ergebnisse der Evaluation der Aktionserkennung

Teilaspekt	Präzision	Ausbeute	F1-Maß
Prädikaterkennung (Senna)	0,939	0,799	0,863
Prädikaterkennung (Aktionserkenner)	0,919	0,883	0,901
Rollenzuordnung (Senna)	0,799	0,678	0,734
Gesamt (Senna)	0,854	0,726	0,785
Gesamt (Aktionserkenner)	0,852	0,762	0,804

de durch den Eigennamenerkenner das Wort Apollinaris als *Person* eingestuft, was dazu führt, dass die resultierende Entität eine `SubjectEntity` mit Namen *Apollinaris* darstellt anstatt der erwarteten Entität *Apollinaris bottle*.

Insgesamt funktioniert aber die Entitätenerkennung sehr gut und bietet eine solide Basis für die weiteren Schritte des Kontextaufbaus.

7.3.2. Aktionserkennung

Um die Aktionserkennung korrekt bemessen zu können muss zunächst ein vollständiges Bild der Qualität der Informationen des Semantische-Rollen-Markierers und somit des Werkzeugs Senna und der Anbindung an die Datenbanken PropBank, VerbNet und FrameNet existieren, welche maßgeblich den Informationsgehalt der Aktionen bestimmen. Hierbei wurde die Evaluation in zwei Teilaspekte unterteilt. Zum einen wird evaluiert, wie gut Senna die im Text beschriebenen Prädikate erkennt. Unter diesem Aspekt ist ein Prädikat richtig positiv (*true positive*) falls es erwartet und vollständig ist, falsch positiv (*false positive*) sofern es nicht in der Form erwartet wurde und falsch negativ (*false negative*) wenn es erwartet aber nicht erkannt wurde. Zum anderen wird die Erkennung der semantischen Rollen bewertet. Dabei ist eine gefundene Rolle nur dann korrekt, wenn vom passenden Prädikat auf das passende Argument verwiesen wird und die genannte Rolle korrekt ist. Falsch positiv sind hierbei die Rollen, die erkannt aber deren Rolle oder die Zuordnung nicht erwartet wurden und falsch negativ diejenigen die dementsprechend nicht erkannt wurden obwohl sie erkennbar gewesen wären.

Anhand dieser Aspekte ergab die Evaluation des Semantische-Rollen-Markierers die in Tabelle 7.3 dargestellten Werte. Festzustellen ist hier, dass die Präzision der Prädikaterkennung (0,939) relativ hoch ist, allerdings die Ausbeute (0,799) noch verbessert werden könnte. Dies liegt vor allem daran, dass Senna die meisten Formen von *to be* nicht als Prädikat einstuft und somit für Formulierungen, wie z.B. „the water is in the fridge“ kein Prädikat und ebenso keine semantischen Rollen erkennt. Außerdem entstehen weitere nicht erkannte Ergebnisse aus Prädikaten, welche textuelle Repräsentationen besitzen, die sowohl als Verb als auch als Adjektiv oder Nomen eingeordnet werden können, wie z.B. *open* oder *press*. In diesen Fällen kann es dazu kommen dass Senna die eigentlichen Prädikate nicht erkennt. Die falsch positiven Ergebnisse der Prädikaterkennung resultieren nahezu ausnahmslos aus dem Phänomen, dass Senna bei einer Kombination von *please* und einem Verb, wie bei „Armar please grab the cup“, als Prädikat das Verb *please* erkennt, aber nicht das eigentlich in dieser Aussage auftretende Verb *grab*. *Please* stellt in diesem Zusammenhang nämlich eigentlich kein Verb dar, sondern ein Adverb. Dies führt also zu einem falsch positiven Prädikat *please* und einem falsch negativen nicht erkannten Prädikat *grab*.

Die Ergebnisse der Rollenzuordnung sind für eine Erkennung semantischer Rollen in gesprochener Sprache relativ gut, denn Senna erreicht selbst auf geschriebener Sprache mit Satzzeichen und ohne Wortwiederholungen und Versprechern nur ein F1-Maß von 0,754¹.

¹siehe [CWB⁺11]

Viele der falsch positiven Ergebnisse basieren auf falsch als zusammengehörig eingestuften Argumenten, was durch die fehlenden Satzzeichen zu erklären ist. Ein weiteres Problem stellen Verbalphrasen dar, die nur aus einer Passivkonstruktion wie „it is finished“ bestehen. Hier erkennt Senna zwar häufig das im passiv stehende beschreibende Verb (im Beispiel *finish*), aber meist keine Rollen, welche das Prädikat mit der Entität (im Beispiel *it*) in Verbindung bringen. Hinzu kommen noch die falsch negativen Rollen, welche durch die nicht erkannten Prädikate entstehen, was die Ausbeute der Rollenzuordnung noch weiter verringert. Insgesamt stellt aber der Semantische-Rollen-Markierer eine valide erste Lösung der Zuordnung semantischer Rollen dar und könnte mit einer gesonderten Betrachtung der Gegebenheiten in gesprochenen Anweisungen und einem Wissen über häufige Probleme wie das Beispiel mit *please* noch weiter verbessert werden.

Ein Aspekt der aber durch diese Arbeit noch verbessert werden konnte, ist die Ausbeute in der Erkennung der Prädikate. Denn die Formen von *to be* und auch einige der ebenso als Nomen oder Adjektiv verwendbaren Oberflächenformen konnten aus den Verbalphrasen extrahiert werden und somit trotz der fehlgeschlagenen Erkennung von Senna aus ihnen eine Aktion erzeugt werden. Die semantischen Rollen fehlen allerdings trotzdem, da diese allein aus Senna extrahiert werden, aber es lässt sich zumindest für den weiteren Kontextaufbau erkennen, dass eine Aktion mit diesem Namen an dieser Stelle der Äußerung auftritt. Ein neues Problem was hierbei auftritt sind Verben, die ein Teil von Kontraktionen wie z.B. *it's* oder *i'm* sind. Diese werden zwar nun durch den Aktionserkennung erkannt, aber nur mit Namen wie *'s* benannt, welche sich nicht auf die Form von (*to*) *be* auflösen lässt. Die Kontraktion *'s* kann sowohl *is* als auch *has* bedeuten und ist somit nicht eindeutig. Die Aktionserkennung erreicht die in Tabelle 7.3 dargestellten Werte für Präzision, Ausbeute und F1-Maß, welche sich aus den Fehlern der Rollenzuordnung aus Senna (eins zu eins in Aktionen übertragen) und der verbesserten Prädikaterkennung ergeben. Hier ist zu erkennen, dass vor allem die Ausbeute (0,762) erhöht werden konnte, was für den Nutzen der Aktionen als Basis für einen weiteren Kontextausbau von Vorteil ist.

Insgesamt lässt sich also festhalten, dass die Aktionserkennung sehr abhängig von den Ergebnissen der Vorverarbeitungsschritte ist, aber trotzdem in vielen Fällen ein valides Abbild der im Kontext auftretenden Aktionen bietet und sich auch gerade wegen der hohen Ausbeute der Prädikaterkennung gut als Basis für die weiteren Schritte der Kontextanalyse verwenden lässt.

7.3.3. Ortsbeziehungserkennung

Der Ortsbeziehungserkennung wird bezüglich der Korrektheit der von ihm identifizierten Ortsbeziehungen bewertet. Dies bedeutet, dass alle in den Äußerungen beschriebenen Ortsbeziehungen dahingehend untersucht werden, ob sie vom Werkzeug erkannt wurden oder nicht. Richtig positive Ergebnisse (*true positive*) sind also Ortsbeziehungen die erwartet und auch erkannt wurden, falsch positive Ergebnisse (*false positive*) Ortsbeziehungen, die fälschlicherweise erkannt wurden und falsch negative Ergebnisse (*false negative*) Ortsbeziehungen, die erwartet aber nicht erkannt wurden.

Diese Kriterien führen zu dem in Tabelle 7.2 dargestellten Ergebnis für Präzision (0,945), Ausbeute (0,793) und F1-Maß (0,862). Es ist klar ersichtlich, dass der Fokus der für die Erkennung eingesetzten einfachen Heuristik auf Präzision und nicht Ausbeute lag. Dies ist sinnvoll, da somit bei einer erkannten Beziehung davon ausgegangen werden kann, dass sie korrekt ist. Die nicht erkannten Beziehungen stellen komplexe Fälle dar, die nur schwierig präzise aufgelöst werden könnten, was bei zu vielen falsch positiven Ergebnissen geführt hätte. Die wenigen falsch positiven Ergebnisse resultieren nahezu gänzlich (drei der vier falsch positiven Ergebnisse) aus Formulierungen in denen mehrere Ortsbeziehungen ineinander geschachtelt wurden, wie es z.B. bei „put the instant meal from the fridge on

it“. Hierbei versagt die Heuristik, da nur die zwei nächsten Entitäten betrachtet werden. In diesem Fall wäre für beide Beziehungen die Entität *instant meal* der Bezug aber es wird die Beziehung [*instant meal* $\xrightarrow{\text{from}}$ *fridge*] sowie die Beziehung [*fridge* $\xrightarrow{\text{on}}$ *it*] erkannt, welche eindeutig falsch ist. Hieraus resultieren auch einige der falsch negativen Ergebnisse, da die eigentliche Ortsbeziehung zwischen *instant meal* und *it* nicht erkannt wird. Die weiteren nicht erkannten Ortsbeziehungen sind zumeist Beziehungen die durch nicht betrachtete Formulierungen wie z.B. „open the microwave and put in the meal“ oder „open it and put the plate inside“ gebildet werden und ebenfalls von der Heuristik nicht abgedeckt sind, da eine Aktion zwischen den Entitäten liegt.

Es lässt sich aber feststellen, dass selbst mit dieser simplen Heuristik bereits ein sehr gutes Ergebnis für die Erkennung von Ortsbeziehungen zwischen Entitäten in Anweisungen erzielt werden kann.

7.3.4. Welt- und Domänenwissenanreicherung

Die Evaluation der Welt- und Domänenwissenanreicherung lässt sich ebenso in zwei Teilaspekte unterteilen. Zum einen ist zu beurteilen, ob die Konzeptbildung aus den Kontextindividuen korrekt funktioniert und damit auch die Anbindung an die Domänenontologie korrekt ist. Zum anderen ist eine Bewertung der aus WordNet den Konzepten hinzugefügten Beziehungen interessant, da hier der eigentliche Mehrwert der Weltwissenanreicherung liegt.

Bei ersterem werden die Beziehungen zwischen den Kontextindividuen und ihren übergeordneten Konzepten überprüft und evaluiert ob das erstellte Konzept ein valides Konzept für das Individuum darstellt sowie sofern existent das Domänenontologieindividuum zum Konzept gefunden wurde. Damit sind in diesem Fall richtig positive Ergebnisse (*true positive*) solche Beziehungen, die auf das richtige Konzept zeigen und deren Informationen korrekt abstrahiert wurden (Konzeptname, WordNet-Lemma, Geschlecht, Domänenontologieindividuum, usw.). Falsch positive Ergebnisse (*false positive*) hingegen Beziehungen, welche auf Konzepte zeigen, die nicht als übergeordnetes Konzept des Kontextindividuum in Frage kommen und falsch negative Ergebnisse (*false negative*) diejenigen Beziehungen auf Konzepte, die eigentlich erwartet wurden aber gefunden wurden.

Hiermit ergeben sich die in Tabelle 7.2 dargestellten Werte für Präzision (0,986), Ausbeute (0,974) und F1-Maß (0,981) der Konzeptbildung. Die sehr hohe Präzision lässt sich zum Teil auf die vielen passenden Individuen in der Domänenontologie zurückführen und geben einen Hinweis darauf, dass Personen häufig die Formulierungen verwenden, die sie bereits zuvor gehört oder gelesen haben. Dies lässt sich schließen, weil in der Domänenontologie die Objekte und Methoden so benannt sind, wie sie auch in der Szenariobeschreibung auftauchen und ebenso häufig von den Probanden verwendet wurden. Außerdem stellen vor allem die Methoden häufig verwendete simple Formulierungen in Anweisungen dar, was darauf hinweist, dass Personen die mit Systemen kommunizieren häufig auf möglichst einfache Formulierungen zurückgreifen. Darüber hinaus lässt sich die Präzision aber auch mit dem gut funktionierenden Abgleich von Synonymen mit Domänenontologieindividuen erklären. So wurde z.B. die *give*-Aktion mit dem Konzept *hand* verknüpft, da es sich um ein aus WordNet extrahiertes Synonym von *hand* handelt. Ebenso funktionierte dies bei *place* \Rightarrow *put* und *rinse* \Rightarrow *wash*. Allerdings traten auch zwei Fälle auf in denen das Vorgehen über die Synonyme aus WordNet zu falsch positiven Ergebnissen führte. Zum einen wurde die Aktion *see* mit dem Konzept *take* verknüpft, da eine der möglichen Bedeutungen von *see* in WordNet das Synonym *take in* enthält, welches eine Jaro-Winkler Ähnlichkeit von 91% zu *take* aufweist. Zum anderen wurde die Aktion *do* mit dem Konzept *come* verbunden, weil hier ebenso eine Bedeutung (proceed or get along) das Synonym *come* aufweist. Hätte man in beiden Fällen eine Information über die Wortbedeutung, so könnte dieser Fehler behoben werden. Die als falsch negativ eingestuften Beziehungen hingegen gehen

zum größten Teil daraus hervor, dass für die Aktionen gebildet aus Kontraktionen (z.B. *it's* \Rightarrow 's-Aktion) kein Lemma in WordNet gefunden werden konnte und dadurch kein Konzept gebildet wird.

Die aus WordNet zwischen den Konzepten aufgebauten Beziehungen lassen sich hingegen nur schwierig hinsichtlich Vollständigkeit überprüfen, da es schwierig ist zu definieren, welche Beziehungen erwartet werden. Daher werden nur die erkannten Beziehungen bewertet und damit nur für die gefundenen Beziehungen und Überkonzepte falsch negative Beziehungen unter den ansonsten gefundenen Konzepten gesucht. Dies bedeutet, dass eine Beziehung (Sub-, Super-, Teil-, Ganzes-, Equal- oder Antonym-Konzept) dann ein richtig positives Ergebnis darstellt, wenn sie eine sinnvolle Beziehung für die im Kontext vorhandenen Konzepte darstellt. Ein falsch positives Ergebnis ist somit also eine Beziehung, die eine nicht sinnvolle Aussage unter Berücksichtigung der Bedeutungen trifft und ein falsch negatives Ergebnis die Beziehungen, die zu erwarten gewesen wären, wenn man die anderen aufgebauten Beziehungen betrachtet und die Konzepte in der Konzeptebene hinzuzieht (z.B. eine Beziehung [*fridge* $\xrightarrow{\text{super}}$ *thing*] wenn es das angereicherte Konzept *thing* gibt und ebenso eine [*table* $\xrightarrow{\text{super}}$ *thing*] Beziehung existiert). Mit dieser Definition ergeben sich die in Tabelle 7.2 unter Konzeptbeziehungsbildung abgebildeten Werte für Präzision (0,683), Ausbeute (0,845) und F1-Maß (0,756).

Es lässt sich feststellen, dass einige sehr interessante Zusammenhänge gebildet werden konnten. So wurde für die beiden Konzepte *Fridge* und *Dishwasher* das gemeinsame Oberkonzept *white goods* gefunden oder für *InstantMeal* das Oberkonzept *meal* und für dieses zusammen mit *water* das Oberkonzept *nutrient*. Auch konnte in einigen Fällen trotz der fehlerhafte Erkennung von Konzepten durch Wortwiederholungen, wie das Konzept *meal instant meal*, ein korrektes Oberkonzept (in diesem Fall *meal*) identifiziert werden. Allerdings werden diese Oberkonzepte immer nur dann aufgebaut, wenn zwei ähnliche Kandidaten in der Konzeptebene vorhanden sind. Dies war eine Entwurfsentscheidung, die noch Potenzial für spätere Verbesserungen bietet, da es auch möglich wäre aus einzelnen Konzepten Oberkonzepte zu bilden. Diese Entscheidung wurde getroffen, da es schwierig gewesen wäre zu bestimmen, ab welchem Punkt und welcher Spezifität ein Oberkonzept sinnvoll in den Kontext übertragen werden sollte. Ebenso konnten in einigen Fällen Antonyme der Aktionen bestimmt werden, wie z.B. *go* als Antonym-Konzept zu *come* und *close* als Antonym-Konzept zu *open*.

Leider sind allerdings auch einige der Beziehungen bedingt durch die unpräzisen Bedeutungsinformationen aus WordNet nicht korrekt. So wird dem *take*-Konzept immer das Antonym *turnOn* zugeordnet, da *take* in einer seiner Bedeutungen das Antonym *turn down* besitzt, welches zu *turnOn* eine Jaro-Winkler-Ähnlichkeit von 95% aufweist. Ein weiterer Fall stellen die Kombinationen *Dishes* und *Meal* sowie *Plate* und *Table* dar. Hier wird bedingt durch die Bedeutung von *dish* als Essensgang *Dishes* als Teilkonzept von *Meal* erkannt und ebenso *plate* in seiner Bedeutung als (Tisch-)Platte als Teil des Konzepts *Table* erkannt. Hätte man in diesen Fällen eine Information über Wortbedeutungen ließen sich diese fehlerhaften Beziehungen vermeiden.

In dieser Evaluation wurden gewollt die einfachen Informationen über die semantischen Relationen der Konzepte, welche zwar gespeichert aber nicht in Beziehungen umgesetzt wurden ausgespart, da diese ohne große Überprüfung aus WordNet übernommen werden und damit nicht sinnvoll zu evaluieren sind. Hier bietet sich aber noch eine große Möglichkeit zur Erweiterung des als erste Herangehensweise zu sehenden Vorgehen dieser Arbeit, da sich über das aufgebaute Kontextwissen die Bedeutungen der Entitäten über z.B. ein gemeinsames Thema und die Bedeutungen der Aktionen z.B. über die semantischen Rollen eingrenzen lassen könnten und sich dadurch die semantischen Relationen aus WordNet validieren ließen.

Insgesamt lässt sich aber festhalten, dass auch so aus WordNet einige sehr interessante Zusammenhänge extrahiert werden konnten und damit der Aufbau der Konzeptebene weitere neue Erkenntnisse zum Wissen über den Kontext beitragen konnte.

7.3.5. Zustandsbestimmung

Auch die Zustandsbestimmung lässt sich hinsichtlich ihrer Leistung beurteilen. Allerdings ist es hier schwierig zu bemessen was zu erwartende Zustände und vollständige Zustandsabbilder darstellen. Daher werden als Grundlage für die Evaluation nur die in der Domänenontologie beschriebenen Zustände überprüft. Dies bedeutet, dass ein Zustand immer dann korrekt erkannt wurde, wenn er auch für diese Entität erwartet wurde. Dies bezieht zum einen die initialen Zustände mit ein aber zum anderen auch die Entitäten, welche über Korreferenzen Zustände erben. Falsch Positive Ereignisse sind dabei Zustände die erkannt wurden aber nicht erwartet wurden und falsch negative diejenigen, die erwartet und nicht erkannt wurden.

Hiermit ergeben sich die in Tabelle 7.2 dargestellten Ergebnisse für Präzision (0,854), Ausbeute (0,627) und F1-Maß (0,723). Die Werte zeigen eine deutlich höhere Präzision als Ausbeute, was auf die Abhängigkeit der Heuristik von den Ergebnissen des Semantische-Rollen-Markierers zurückzuführen ist. Ein Zustand wird nur dann ermittelt, wenn eine Aktion die betrachtete Entität als behandeltes Objekt einschließt und einen passenden Zustandsübergang beschreibt. Da allerdings, wie bereits beschrieben, viele der *close-* oder *open-*Aktionen durch Überdeckung in Senna keine semantischen Rollen erhalten haben oder durch einen Fehler in der Wortartmarkierung gar nicht erst als Aktionen erkannt wurden, konnte für diese auch kein Zustand bestimmt werden. Außerdem entstehen hieraus durch das Erben der Zustände über die Korreferenzen weitere falsch positive Ergebnisse.

(70) [...] open the fridge [...] close the fridge [...] next to the fridge [...]

So erben die beiden letzteren *fridge-*Entitäten in Beispiel 70 den Zustand *opened* falls die *close-*Aktion nicht erkannt wird, was dazu führt das sowohl die jeweiligen *opened-*Zustände falsch positive Ergebnisse als auch die beiden fehlenden *closed-*Zustände als falsch negative Ergebnisse in die Auswertung mit einfließen. Ein weiterer Fall, der zu einigen falsch positiven Ergebnissen führt, sind Formulierungen wie „fill it with water“ bei denen die *water-*Entität vom Semantische-Rollen-Markierer die VerbNet-Rollen *Co-Theme* und *Theme* erhält, was dazu führt, dass sowohl die *it-* als auch die *water-*Entität den Zustand *full* zugeordnet bekommt, was offensichtlich falsch ist. Wäre hier nur die thematische Rolle *Co-Theme* angegeben, hätte die Heuristik, welche den Zustand den Argumenten hinzufügt, die die Rolle *Theme* einnehmen, das korrekte Ergebnis geliefert. Außerdem existieren Formulierungen, die zwar auf einen in der Domänenontologie vorhandenen Zustand hinweisen, aber nicht mit einem Verb beschrieben wurden, welches sich einem Aktionskonzept zuordnen ließ, dass diesen Zustandsübergang durchführt. So wird für die Formulierung „activate the microwave“ nicht der Zustand *on* ermittelt, da *activate* nicht mit dem **turnOn**-Konzept verbunden werden konnte. Allerdings gibt es auch Fälle in denen dies durch die Synonyme abgefangen wird. So wird für die Formulierung „rinse it“ der Zustand *cleaned* gefolgert, da *rinse* ein Synonym von *wash* ist und dieses Konzept in der Domänenontologie vorhanden ist und den entsprechenden Zustandsübergang beschreibt.

Insgesamt lässt sich also sagen, dass die Zustandsbestimmung gut funktioniert aber stark von der Qualität der Erkennung der Aktionen abhängt, was bedingt durch die Konzeption der Heuristik (siehe Abschnitt 6.2.5) zu erwarten war. Hätte man also eine zuverlässigere Aktionserkennung, würde auch die Zustandsbestimmung zuverlässiger.

7.4. Korreferenzanalyse

Um sowohl die Leistung der in dieser Arbeit erstellten Korreferenzanalyse bezüglich aktueller Standard-Werkzeuge als auch hinsichtlich der Auswirkungen des Kontextwissens beurteilen zu können wird die Evaluation der Korreferenzanalyse in mehreren Stufen durchgeführt. Zunächst wird als Referenz das Ergebnis des *Stanford Deterministic Coreference Resolution Systems* (siehe Abschnitt 2.4.2) betrachtet. Daraufhin wird das in dieser Arbeit entworfene Korreferenzverfahren mit einem minimalen Kontextaufbau, bestehend nur aus einer Erkennung der Entitäten, untersucht, welches die Informationen emulieren soll die Korreferenzwerkzeugen normalerweise zur Verfügung stehen. Abschließend wird dann das Verfahren mit dem vollständigen Kontext gebildet durch die komplette Kontextanalyse evaluiert.

Bei allen drei Varianten, wird sowohl der Alt-Korpus als auch die neuen Transkriptionen als Grundlage verwendet und das Ergebnis mit den Musterlösungen wie in Abschnitt E im Anhang beschrieben validiert. Da ein Teil des Alt-Korpus zur Entwicklung der Werkzeuge betrachtet wurde, wird dieser gesondert in der Evaluation geführt, um den Effekt der Anpassung auf bestimmte Formulierungen betrachten zu können. Dieser Teil besteht aus den ersten 10 Probanden der ersten drei Szenarien aus der Arbeit von Günes [Gün15] sowie den ersten 9 Probanden der zwei Szenarien aus der Arbeit von Steurer [Ste16], umfasst also insgesamt 48 der 146 Transkriptionen des Alt-Korpus.

Zur Bewertung der Ergebnisse werden die folgenden Definitionen für die Ergebniseinordnungen verwendet:

- **Richtig positives Ergebnis** (*true positive* kurz *tp*): Die Referenz mit der höchsten Wahrscheinlichkeit, die von einem referierenden Ausdruck ausgeht stimmt mit der erwarteten Referenz überein. Das heißt für eine Musterlösungsreferenz mit Start 2 und Ende 0 wird eine Entität gefunden, welche Start- bzw. Ziel-Wort beinhaltet und die **ReferentRelation** mit der höchsten Wahrscheinlichkeit, die von der Start-Entität ausgeht zeigt auf die Ziel-Entität.
- **Falsch positives Ergebnis** (*false positive* kurz *fp*): Die Referenz mit der höchsten Wahrscheinlichkeit, die von einem referierenden Ausdruck ausgeht stimmt nicht mit der erwarteten Referenz überein oder es handelt sich um eine Referenz von einer Entität ausgehend, die keine Referenzen bildet.
- **Falsch negatives Ergebnis** (*false negative* kurz *fn*): Es wurde eine Referenz zwischen zwei Entitäten erwartet, die entweder nicht die mit der höchsten Wahrscheinlichkeit darstellt oder gar nicht gefunden wurde.

Im Fall der Evaluation des Korreferenzwerkzeugs von Stanford lässt sich das Konzept der Entität auf die dort verwendete **Mention** übertragen sowie die **ReferentRelation** mit der **CorefRelation** ersetzen. Außerdem existieren in diesem Fall keine Wahrscheinlichkeiten, was die Auswertung einfacher gestaltet.

7.4.1. Referenzkorreferenzen

Um das Ergebnis des *Stanford Deterministic Coreference Resolution Systems* (siehe Abschnitt 2.4.2) mit dem Ergebnis der Korreferenzanalyse dieser Arbeit vergleichen zu können, wurde das Korreferenzwerkzeug von Stanford mit dem Ergebnis des Vorverarbeitungsfließband ausgeführt. Hierzu wurden vor allem die Wortart-Markierungen übernommen. Die Ergebnisse des Eigennamenerkenners (siehe Abschnitt 6.1.1) ließen sich allerdings nicht wiederverwenden, da Stanford eine umfassendere Kategorisierung der Eigennamen erwartet, weshalb die Eigennamenerkennung aus der Werkzeugsammlung *Stanford CoreNLP* [sta] vorgeschaltet wurde. Außerdem benötigt das Korreferenzwerkzeug von Stanford das Ergebnis der Syntaxanalyse des *Stanford Parser*, welcher ebenso aus der Werk-

Tabelle 7.4.: Ergebnisse der Evaluation des Stanford Korreferenzers

Korpusenteil	Präzision	Ausbeute	F1-Maß
Alt-Korpus (Entwicklung)	0,743	0,491	0,591
Alt-Korpus (Rest)	0,773	0,544	0,639
Neue Texte	0,728	0,607	0,662
Gesamt (Alt-Korpus)	0,757	0,520	0,617
Gesamt	0,746	0,561	0,640

zeugsammlung *Stanford CoreNLP* übernommen wurde. Die Evaluation des Referenzkorreferenzers ergab das in Tabelle 7.4 dargestellte Ergebnis für Präzision (0,746), Ausbeute (0,561) und F1-Maß (0,640). Dieses Ergebnis liegt zwar über dem F1-Maß, welches das Werkzeug auf Nachrichtentexten erreicht (0,577 siehe Abschnitt 2.4.2) und ist somit kein schlechtes Ergebnis, aber selbst für die zum Teil sehr einfachen Beschreibungen im Alt-Korpus ist das Ergebnis nicht zufriedenstellend. Gerade in diesen Texten ohne viele Objekte und Bezugsmöglichkeiten wäre ein besseres Ergebnis zu erwarten gewesen. Außerdem ergab die Evaluation ein weiteres Problem, welches das Werkzeug durch den Einsatz der Syntaxbäume hat. Da die Syntaxbäume auf gesprochener Sprache zum Teil fehlerhaft sind resultieren viele der erkannten *Mention* als ebenso fehlerhaft und überspannen mehrere Phrasen, wie es bei der *Mention* „hi robo please look at the table“ der Fall ist. Somit lässt sich für diese Formen von *Mention* kein präziser Referent bestimmen. Da die in der Evaluation eingesetzte Suche nach Start- und Ziel-*Mention* allerdings somit trotzdem die unpräzisen Ergebnisse findet, kann hierdurch sogar das Evaluationsergebnis zum positiven verfälscht werden.

Dieses Ergebnis zeigt außerdem sehr eindrücklich, dass bei gesprochener Sprache auf andere Heuristiken, als Syntaxbäume, zurückgegriffen werden muss, um robuste Korreferenzanalyse-Werkzeuge erstellen zu können.

7.4.2. Basis-Korreferenzanalyse

Um den Einfluss des Kontextwissens auf die Korreferenzanalyse bemessen zu können, wurde der Korreferenz-Agent um eine Basis-Variante des Entitätenerkenners erweitert, welche ausgeführt wird falls kein Kontext vorhanden ist. Dieser stellt die Menge an Informationen dar, die Korreferenzwerkzeuge normalerweise aus externen Werkzeugen, wie im Fall von Stanford dem *Stanford Parser*, generieren. Ansonsten wird die Auflösung der Referenzen genauso umgesetzt wie in Abschnitt 6.3 beschrieben. Dies ermöglicht es ohne vollen Zugriff auf das Kontextwissen das Ergebnis des entstandenen Korreferenzanalyse-Werkzeugs zu bewerten. Dieser *BasicEntityRecognizer* setzt das in Abschnitt 6.2.1 beschriebene Verhalten um, allerdings ohne Informationen aus jeglichen Welt- oder Domänenwissensquellen. Dadurch beinhaltet er natürlich auch einen Teil des Kontextwissens und ist somit nicht komplett kontextunabhängig, bietet aber eine einfache Möglichkeit zumindest den Einfluss der Aktionserkennung, der Konzeptualisierung und der Zustandsbestimmung zu bemessen.

Tabelle 7.5.: Ergebnisse der Evaluation der Basis-Variante der Korreferenzanalyse

Korpusteil	Präzision	Ausbeute	F1-Maß
Alt-Korpus (Entwicklung)	0,922	0,848	0,884
Alt-Korpus (Rest)	0,841	0,786	0,812
Neue Texte	0,797	0,786	0,791
Gesamt (Alt-Korpus)	0,872	0,810	0,840
Gesamt	0,837	0,799	0,818

Die Ergebnisse der Evaluation mit dieser Basis-Variante sind in Tabelle 7.5 dargestellt und deuten bereits eine deutliche Verbesserung zum Stanford-Ergebnis an (Präzision 0,837, Ausbeute 0,799, F1-Maß 0,818). Diese lässt sich zum einen auf die an die Gegebenheiten von Anweisungen und deren Formulierungen angepasste Lösung dieses Werkzeugs zurückführen, aber zum anderen auch mit der Auswahl der Eigenschaften begründen, welche unter dem Aspekt einer gegenüber gesprochener Sprache robusten Deutung der Referenzen ausgewählt wurden. So liegt der Fokus stark auf dem Abstand zum referierenden Ausdruck, was als Eigenschaft in gesprochener Sprache sehr gut zu funktionieren scheint.

Es lässt sich außerdem festhalten, dass der für die Entwicklung des Werkzeugs betrachtete Teil des Korpus zwar wie erwartet das beste Ergebnis liefert aber bedingt durch die hauptsächlich einfachen Formulierungen und kurzen Anweisungsfolgen mit wenigen Entitäten in diesem Teil des Korpus das bessere Ergebnis nicht nur auf eine Überanpassung zurückzuführen ist. Die im Korpus enthaltenen Anweisungen bestehen häufig aus nur einem Typ von Pronomen und zumeist nicht mehr als zwei bis drei Entitäten, was eine Auflösung der dort beschriebenen Referenzen wegen einer geringeren Zahl von möglichen Kandidaten einfacher macht. Außerdem stellen die 48 Texte nur 112 der 294 erwarteten Korreferenzen im Alt-Korpus und vor allem einen noch kleineren Teil der 528 insgesamt betrachteten Korreferenzen.

7.4.3. Kontextbasierte Korreferenzanalyse

Die letzte evaluierte Variante stellt die in Abschnitt 6.3 beschriebene Korreferenzanalyse mit dem vollständigen Kontextwissen aus dem Kontextanalysierer dar. Dabei werden beide Werkzeuge abwechselnd ausgeführt und aus dem resultierende Kontext, sobald er sich nicht mehr verändert, die Referenzbeziehungen (`ReferentRelation`) ausgelesen und diese mit den Musterlösungen (siehe Abschnitt E) verglichen. Die abwechselnde Ausführung simuliert den späteren parallelen Betrieb der Agenten.

Die Einordnung der Referenzen bezüglich der Musterlösung führte zu dem in Tabelle 7.6 dargestellten Ergebnis (Präzision 0,881, Ausbeute 0,881, F1-Maß 0,881). Hierbei ist ersichtlich, dass der volle Kontext gerade was die Ausbeute angeht noch einmal einen deutliches

Tabelle 7.6.: Ergebnisse der Evaluation der kontextbasierten Korreferenzanalyse

Korpusteil	Präzision	Ausbeute	F1-Maß
Alt-Korpus (Entwicklung)	0,936	0,920	0,928
Alt-Korpus (Rest)	0,890	0,885	0,887
Neue Texte	0,848	0,859	0,854
Gesamt (Alt-Korpus)	0,907	0,898	0,903
Gesamt	0,881	0,881	0,881

Plus gegenüber der Basisvariante erreicht sowie in allen Bereichen eindeutig die beste Leistung der betrachteten Varianten liefert. Die erhöhte Ausbeute lässt sich vor allem auf die Anreicherung mit Welt- und Domänenwissen zurückführen, die es ermöglicht z.B. zuverlässig *Armar* als Subjektentität zu klassifizieren oder aber auch Referenzen über Synonyme und Hyperonyme, wie z.B. zwischen *bottle* und *water bottle* zu erkennen. Aber auch die Präzision, gerade auf den neuen Texten, ließ sich durch das umfassende Kontextwissen nochmals erhöhen.

Es lässt sich feststellen, dass gerade die Identitätsauflösung sehr präzise funktioniert, was durch die detaillierteren Informationen im sprachlichen Kontext auch zu erwarten war. So werden nur fünf Identitäten falsch erkannt, allerdings hingegen 46 pronominale Anaphern. Von diesen werden alle bis auf zwei vom Pronomen *it* gebildet. Dies war zu erwarten, da das Pronomen *it* auch die meisten möglichen Kandidaten zur Auflösung in den Szenarien besitzt, weil es auf Objekte verweist. Diese fehlerhaften Referenzen basieren häufig auf Ausnahmen der Eigenschaften *Ortsbeziehung* und *Übereinstimmung semantischer Rollen*. So wird bei der Formulierung „get the plate from the dishwasher and close it“ die Entität *dishwasher* als unwahrscheinlicher eingestuft, da sie als beschreibende Ortsbeziehung zu *plate* betrachtet wird. Diese Einschätzung ist für diesen Fall falsch, trifft aber in vielen anderen Fällen zu. Eigentlich ließe sich diese fehlerhafte Referenz außerdem vermeiden, wenn man sicher feststellen könnte, dass *plate* nicht an einer Aktion teilhaben kann, die einen Zustandsübergang nach *opened* durchführt, *dishwasher* aber schon. Allerdings kann bei einer Entität, die nicht den Zustand *opened* enthält nicht automatisch darauf geschlossen werden, dass diese nicht trotzdem als Referent in Frage kommt, denn sie könnte entweder aus WordNet extrahiert worden sein und deshalb keine Zustände enthalten oder das Lemma der Aktion lässt mehrere Bedeutungen zu, was es schwierig macht eine definitive Entscheidung zu fällen. Deshalb ist auch das Sieb *Aktionausführbar* (siehe Abschnitt 5.3.3.8) nur als Ausschlusskriterium von Entitäten gedacht bei denen die Informationen vollständig sind aber nicht in Fällen wie hier, bei denen keine sichere Aussage getroffen werden kann. Ein Fall bei dem das Sieb *semantische Rollen* fehlschlägt stellt die Formulierung „[...] place the plate somewhere on the table go to the fridge and open it [...]“ dar. Hier wird die *plate*-Entität als wahrscheinlicherer Referent eingestuft, weil sie ebenso die thematische Rolle *Theme* einnimmt wie das Pronomen *it*, der eigentliche Referent *fridge* hingegen die Rolle *Destination*. Auch dies ist ein Fall bei dem eine unsichere Eigenschaft nicht auf den richtigen Referenten hinweist, aber wiederum ist diese Eigenschaft eine Abwägung zwischen vielen Fällen in denen sie funktioniert und einigen Fällen, in denen die Einordnung fehlerhaft ist. Diese beiden Beispiele deuten aber darauf hin, dass man ein genaueres Wissen über Zustände und Zustandsübergänge benötigt, um sicherer entscheiden zu können, welche Referenten in Frage kommen, was eine gute Motivation für eine weitere Verbesserung der Zustandsbestimmung und des Kontextaufbaus darstellt.

Von den Identitätsreferenzen die falsch positive Ergebnisse darstellen, resultieren allerdings einige (drei der fünf) aus Fehlern der Weltwissenanreicherung (siehe Abschnitt 7.3.4). So wurde z.B. die Referenz *table* → *plate* deshalb erkannt, weil *plate* fälschlicherweise als Unterbegriff von *table* bestimmt wurde. Außerdem sind einige der falsch negativen Ergebnisse ebenso auf diesen Fall zurückzuführen. Ein Beispiel hierfür ist, dass *juice* in WordNet kein Überbegriff für *orange juice* ist sondern *orange juice* ein *fruit juice* ist, welcher in keiner Weise mit *juice* verknüpft ist. Dadurch lässt sich keine Hyperonym-Beziehung zwischen *juice* und *orange juice* erkennen, was dazu führt, dass das *Objektübereinstimmungssieb* (siehe Abschnitt 6.3.3) keinen passenden Kandidaten finden kann.

Zusätzlich treten noch einige Fehler auf, die nicht durch das Fehlverhalten der Korreferenzanalyse zu erklären sind, sondern auf fehlerhaften Formulierung der Probanden beruhen. So existieren die Formulierungen „take the plates from the dishwasher and fill it“ und „Hello Armar bring me the popcorn from the table he stands between Vitalis and cup“, in

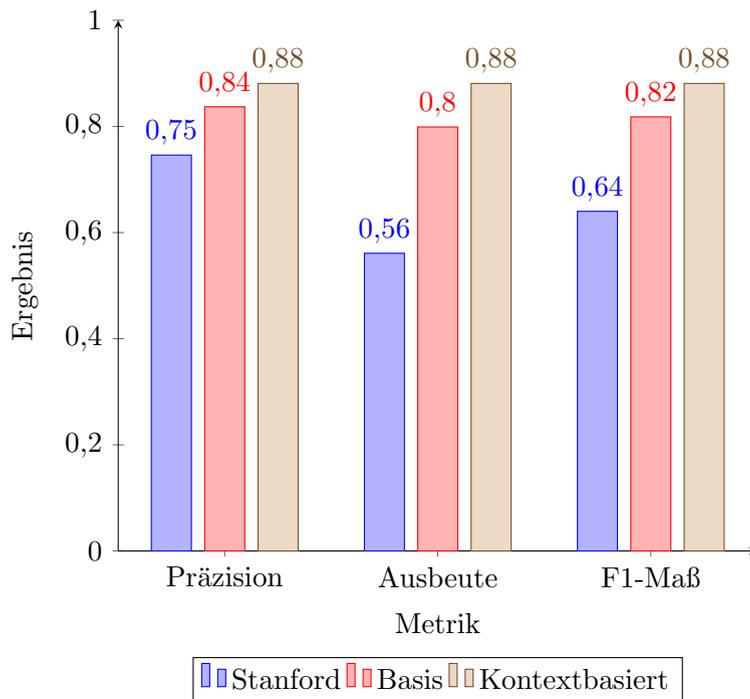


Abbildung 7.1.: Übersicht der Ergebnisse der Korreferenzvarianten

welchen jeweils das falsche Pronomen für den Bezug verwendet wurde. Zum einen wurde *it*, welches eigentlich nur für ein Neutrum im Singular als Platzhalter dient, für die Entität *plates* im Plural verwendet und zum anderen das Pronomen *he* als Referenz auf das Objekt *popcorn* verwendet, welches nicht Maskulin ist. Solche Fälle lassen sich nur schwer mit einer automatisierten Koreferenzanalyse aufdecken, da sie grammatikalisch falsch sind und es zumeist anderen Kandidaten gibt die grammatikalisch korrekt zu den Pronomen passen. Des Weiteren resultieren auch einige der Fehler aus der Entitäten- und Aktionserkennung in fehlerhaften Referenzen. So kann dadurch, dass einige Zustände durch das Fehlen der Aktionen nicht erkannt werden, die Eigenschaft *Aktion ausführbar* in einigen Fällen nicht sinnvoll angewendet werden. Außerdem können fehlerhaft erkannte Aktionen, wie die fälschlicherweise als Entitäten deklarierten *press* und *wait* (siehe Abschnitt 7.3.1), als potenzielle Kandidaten für Referenten gesehen werden, obwohl diese eigentlich als Aktion keine Rolle bei der Korreferenzanalyse darstellen würden.

Auch die kontextbasierte Korreferenzanalyse erreicht ihr bestes Ergebnis auf dem Alt-Korpus, was auch hier mit den dort einfacheren Formulierungen und kürzeren Anweisungen zu erklären ist. Allerdings lässt sich feststellen, dass die erreichte Verbesserung der Metriken auf dem Alt-Korpus und den neuen Texten in etwa gleich ist. Dies deutet darauf hin, dass die eingesetzten kontextbasierten Eigenschaften sowohl auf kurzen einfachen Formulierungen als auch bei komplexeren längeren Anweisungen ähnlich gut funktionieren. Insgesamt lässt sich festhalten, dass die kontextbasierte Korreferenzanalyse eine valide Lösung für das Auflösen von Referenzen in gesprochenen Anweisungen darstellt und in 465 von 528 Fällen die korrekten Referenzen liefert. Außerdem bietet sie durch das Speichern der ebenso möglichen aber als unwahrscheinlicher klassifizierten Kandidaten die Möglichkeit durch spätere genauere Informationen oder auch ein Nachfragen über ein Dialogsystem, die fehlerhaften Referenzen zu verbessern.

7.5. Bewertung der Ergebnisse

Insgesamt bietet die Kombination aus Kontextwissensaufbau und Korreferenzanalyse eine gute Möglichkeit, um Referenzen in gesprochener Sprache aufzulösen. Außerdem bietet die Kontextanalyse die Möglichkeit auch andere Teile des Projektes und somit der Umsetzung von gesprochenen Anweisungen in Quelltext zu unterstützen. Auch wenn einige Teile des Kontextaufbaus noch Potenzial für Verbesserungen haben, bietet der aufgebaute Kontext ein deutlich umfassenderes Abbild und somit Verständnis der ausgedrückten Anweisungen.

Bei der Evaluation ließ sich außerdem feststellen, dass gerade bei längeren Szenarien und Anweisungsfolgen deutlich mehr Entitäten und Konzepte eingeführt werden und die Menge der Korreferenzen steigt. Gerade in diesen Fällen wäre es ohne einen Aufbau eines Kontextes nur schwer möglich ein Verständnis des Gesagten zu bilden sowie die auftretenden Referenzen korrekt aufzulösen. Außerdem bietet die gesprochene Sprache weitere Herausforderungen für den Aufbau von Wissen und auch das Auflösen von Referenzen, da grammatikalische Fehler und Versprecher dazu führen, dass ohne ein Wissen über Zusammenhänge viele Mehrdeutigkeiten nur schwer auflösbar sind. Gerade unter diesen Herausforderungen leidet der Ansatz von Stanford, welcher auf korrekte Syntaxbäume setzt. In Abbildung 7.1 sind die Metriken der drei Varianten der Korreferenzanalyse noch einmal gegenübergestellt und es wird deutlich, dass der Ansatz von Stanford nicht geeignet für gesprochene Sprache ist, aber auch, dass ein umfassendes Kontextwissen einen weiteren Zuwachs der Qualität des Ergebnisses erreicht. Dieser Zuwachs wäre sogar noch größer, wenn die Fehler der Kontextanalyse verringert werden könnten und damit die kontextabhängigen Siebe des Korreferenz-Agenten noch besser arbeiten könnten. Gerade die Zustandserkennung und die Anreicherung mit Weltwissen haben hier noch viel Potenzial für Verbesserungen, zeigen aber zunächst auch was mit den einfachen Heuristiken bereits möglich ist.

Außerdem lässt sich festhalten, dass gerade die Identitätsauflösung sehr vom Kontext profitiert, denn um Identitäten vollständig auflösen zu können benötigt man zwangsläufig Informationen über Beziehungen zwischen Ausdrücken wie Hyperonym-Beziehungen und Synonyme, wenn man mehr als nur direkte Wortwiederholungen erkennen möchte.

8. Zusammenfassung und Ausblick

Das Forschungsgebiet natürlichsprachliche Programmierung, in welchem diese Arbeit angesiedelt ist, hat das Potenzial gerade in einer immer mehr von Software durchdrungenen Welt, Menschen die Interaktion mit Softwaresystemen zu erleichtern. Hierdurch benötigen sie keine Kenntnisse über komplexe Programmiersprachen, sondern nutzen die ihnen bekannte Form der Kommunikation: die natürliche Sprache. Um diesem Ziel näher zu kommen beschäftigt sich das Projekt *Programming Architecture for Spoken Explanations (PARSE)* mit der Umsetzung von in gesprochener natürlicher Sprache gegebenen Anweisungen in Quelltext, der von einem Zielsystem interpretiert werden kann. Gerade die gesprochene natürliche Sprache stellt die Umsetzung vor weitere Herausforderungen, die über die Anforderungen für die Umsetzung textuell gegebener Anweisungen hinausgehen. Sie stellt allerdings auch die natürlichste Form der Kommunikation dar und bietet gerade im Hinblick auf Systeme im Alltag, wie z.B. Haushaltsroboter, in naher Zukunft viel Potenzial.

Diese Arbeit hat sich im Rahmen des Projektes mit zwei für die Umsetzung wichtigen Aufgaben befasst: Der Kontext- und der Korreferenzanalyse. Zunächst wurde, um ein Abbild der in einer Anweisung gegebenen Zusammenhänge zu erhalten, Kontextwissen generiert. Dies bietet die Möglichkeit die Beziehungen von verschiedenen Elementen der Äußerung untereinander besser zu verstehen und für die Betrachtung weiterer Äußerungen auf bereits geäußerte Informationen zurückgreifen zu können. Hierzu wurden die in einer Äußerung gegebenen Entitäten und Aktionen erkannt und ihr Zusammenspiel mittels semantischer Rollen dargestellt. Außerdem wurde eine einfache Heuristik zur Abbildung von Ortsbeziehungen zwischen Entitäten eingeführt, um ein Bild der gegebenen Ortsinformationen zu erhalten. Daraufhin wurden die extrahierten Informationen in übergeordneten, von der Textstelle unabhängigen, Konzepten zusammengefasst und diese mit Wissen über die Domäne aus der Domänenontologie angereichert. Diese Konzepte wurden zusätzlich um Weltwissen aus WordNet erweitert, indem Überkonzepte gebildet sowie Synonyme- und Antonyme-Konzepte erkannt werden. Diese Informationen versuchen das dem Menschen durch jahrelanges Lernen zur Verfügung stehende Zusammenhangswissen zumindest in einer Basisform zu ersetzen. Als letzten Schritt der Kontextanalyse wurde versucht die erkannten Entitäten um Zustände, die sie zu einem Zeitpunkt innehaben, zu erweitern. Hierbei wurde auf die in der Domänenontologie modellierten Zustände zurückgegriffen und versucht aus den beschriebenen Aktionen und Adjektiven ein erstes Bild der Zustände zu generieren.

Die zweite Aufgabe befasst sich mit der Auflösung von Korreferenzen und stellt ein gutes

Anwendungsbeispiel für das aufgebaute Kontextwissen dar, denn auch der Mensch nutzt zur Auflösung von Referenzen sein Wissen über die Zusammenhänge des zuletzt Gesagten. Hierbei wurde eine auf die Auflösung von Referenzen in gesprochenen Anweisungen angepasste Korreferenzanalyse implementiert, welche sich den aufgebauten Kontext zu Nutze macht, um ein besseres Ergebnis zu erreichen.

Die Evaluation der Kontextanalyse hat ergeben, dass die Erkennung der Entitäten sehr zuverlässig funktioniert (F1-Maß 0,97) und auch die Aktionserkennung mit 0,80 ein gutes Ergebnis liefert, obwohl die Ausbeute von 0,68 bei der Erkennung der semantischen Rollen durch das verwendete Werkzeug *Senna* noch viel Raum für Verbesserungen bietet. Die Erkennung der Ortsbeziehungen erreichte trotz der einfachen Heuristik bereits eine Präzision von 0,95 bei einer Ausbeute von 0,79 (F1-Maß 0,86) und könnte gerade durch eine Auflösung der nicht betrachteten verketteten Ortsbeziehungen noch weiter verbessert werden. Auch die Bildung der Konzepte aus den Informationen der Entitäten und Aktionen im Zusammenspiel mit der Domänenontologie erreichte mit 0,98 F1-Maß ein sehr gutes Ergebnis. Die Anreicherung von Konzeptbeziehungen, abgeleitet aus WordNet, könnte hingegen mit 0,76 F1-Maß (Präzision 0,68) noch verbessert werden. Hier spielen vor allem die teils sehr unterschiedlichen Bedeutungen eines Wortes eine große Rolle, welche ohne eine Wortbedeutungsanalyse nur schwer zu unterscheiden sind. Die Zustandsbestimmung erreichte eine Präzision von 0,85, allerdings bei einer Ausbeute von 0,63, was zumeist auf die fehlenden oder fehlerhaften semantischen Rollen der Aktionen zurückzuführen ist.

Die Evaluation der Korreferenzanalyse hat gezeigt, dass eine auf gesprochene Sprache angepasste Korreferenzanalyse gegenüber dem als Referenz eingesetzten *Stanford Deterministic Coreference Resolution System* [LPC⁺11] eine Verbesserung von 0,18 auf 0,82 im F1-Maß erreicht. Durch das aufgebaute Kontextwissen ließ sich das Ergebnis der Korreferenzanalyse noch einmal um 0,06 erhöhen und insbesondere die Ausbeute konnte von 0,80 auf 0,88 angehoben werden.

Trotz dieser guten Ergebnisse lassen sich noch einige Punkte zur Verbesserung und vor allem Erweiterung der vorgestellten Lösungen bestimmen. So stellt die vorgestellte Kontextanalyse nur den ersten Schritt eines möglichen Kontextaufbaus dar und kann in einigen Punkten noch verbessert und erweitert werden. Zunächst wäre hier eine auf gesprochene Sprache angepasste Erkennung von semantischen Rollen zu nennen, welche dadurch auch das Ergebnis der Zustandsbestimmung verbessern würde. Auch die Zustandsbestimmung selbst deckt momentan nur binäre in der Domänenontologie beschriebene Zustände ab, könnte aber mit einem umfassenderen Weltwissen noch erweitert werden. Außerdem gibt es einige zustandsbeschreibende Formulierungen, wie z.B. Relativsätze, die bisher nicht betrachtet werden. Dies gilt auch für die Bestimmung von Ortsbeziehungen bei denen noch weitere Formulierungen und Formen betrachtbar wären. Hier ließen sich außerdem die bereits erwähnten nicht betrachteten zusammengesetzten Ortsbestimmung wie „the cup on the table next to the popcorn“ oder Ortsbeziehungen die über mehrere Entitäten hinweg formuliert sind analysieren, um das Ergebnis noch weiter zu verbessern.

Aber gerade die Anreicherung der Konzepte mit Weltwissen bietet das größte offene Potenzial dieser Arbeit. Wie bereits beschrieben ließe sich mit einer Wortbedeutungsanalyse, welche sich auch auf die Informationen der Aktionserkennung und den identifizierten Überbegriffen stützen könnte, die Zahl der falsch erkannten Beziehungen zwischen Konzepten verringern. Außerdem könnte der Aufbau von Überkonzepten von der Suche nach kleinsten gemeinsamen Überbegriffen zweier Konzepte auf hinreichend spezifische Überbegriffe einzelner Konzepte übertragen werden und zwischen diesen wiederum Zusammenbegriffe gesucht werden, was zu einer noch umfassenderen Konzepthierarchie führen würde.

Der aufgebaute Kontext bietet über die Anwendung in der Korreferenzanalyse hinaus noch weitere Möglichkeiten. So könnte das aufgebaute Wissen dazu genutzt werden eine Wortbedeutungsanalyse mit Wissen über die auftretenden Konzepte zu unterstützen oder die

Domänenontologie um weiteres Wissen zu ergänzen. Außerdem bietet das Abbild der genannten Entitäten und Aktionen eine gute Grundlage für weitere Schritte, wie eine Analyse der zeitlichen Abfolgen der genannten Aktionen oder ein Dialogsystem, zur Klärung von unsicheren Ergebnissen.

Somit ist der nächste logische Schritt im Rahmen des Projektes eine Verbesserung der Aktionserkennung und eine auf den hierdurch verbesserten Ergebnissen der Kontextanalyse basierende Erweiterung des Konzeptaufbaus. Dies würde es ermöglichen die Korreferenzanalyse nochmals zu verbessern und auch die aufgeführten weiteren Anwendungsbereiche des Kontextes zu ermöglichen.

Literaturverzeichnis

- [All95] ALLEN, James: *Natural language understanding / James Allen*. 1995. – ISBN 0–8053–0334–0 (zitiert auf den Seiten 25 und 29).
- [All01] ALLAN, Keith: *Natural Language Semantics*. Wiley, 2001. – ISBN 978–0–631–19297–8 (zitiert auf Seite 15).
- [All03] ALLEN, James F.: Natural Language Processing. In: *Encyclopedia of Computer Science*. Chichester, UK : John Wiley and Sons Ltd., 2003. – ISBN 978–0–470–86412–8, S. 1218–1222 (zitiert auf Seite 7).
- [ARA⁺06] ASFOUR, T. ; REGENSTEIN, K. ; AZAD, P. ; SCHRODER, J. ; BIERBAUM, A. ; VAHRENKAMP, N. ; DILLMANN, R.: ARMAR-III: An Integrated Humanoid Platform for Sensory-Motor Control. In: *2006 6th IEEE-RAS International Conference on Humanoid Robots*, 2006, S. 169–175 (zitiert auf Seite 18).
- [Aut] AUTAYEU, Aliaksandr: *extJWNL (Extended Java WordNet Library)*. <http://extjwnl.sourceforge.net/>, (zitiert auf den Seiten 71 und 77).
- [BFL98] BAKER, Collin F. ; FILLMORE, Charles J. ; LOWE, John B.: The Berkeley FrameNet Project. In: *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics - Volume 1*. Stroudsburg, PA, USA : Association for Computational Linguistics, 1998 (ACL '98), S. 86–90 (zitiert auf Seite 15).
- [BL06] BERGSMA, Shane ; LIN, Dekang: Bootstrapping path-based pronoun resolution. In: *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, 2006, S. 33–40 (zitiert auf den Seiten 36 und 71).
- [Bos08] BOS, Johan: Wide-coverage Semantic Analysis with Boxer. In: *Proceedings of the 2008 Conference on Semantics in Text Processing*. Stroudsburg, PA, USA : Association for Computational Linguistics, 2008 (STEP '08), S. 277–286 (zitiert auf Seite 24).
- [BUCW10] BORDES, Antoine ; USUNIER, Nicolas ; COLLOBERT, Ronan ; WESTON, Jason: Towards understanding situated natural language. In: *International Conference on Artificial Intelligence and Statistics*, 2010, S. 65–72 (zitiert auf den Seiten xi, 23 und 29).
- [Buß02] BUSSMANN, Hadumod: *Lexikon der Sprachwissenschaft*. 3. aktualisierte und erweiterte Auflage. Alfred Kröer Verlag, 2002. – ISBN 3–520–45203–0 (zitiert auf den Seiten 3, 4, 5, 6 und 7).
- [CC14] COULTHARD, Malcolm ; COULTHARD, Margaret: *An Introduction to Discourse Analysis*. Routledge, 2014. – ISBN 978–1–317–86921–4 (zitiert auf Seite 4).

- [cona] *CoNLL-2005 Shared Task: Semantic Role Labeling*. <http://www.cs.upc.edu/~srlconll/home.html>, (zitiert auf Seite 10).
- [conb] *CoNLL-2011 Shared Task: Modeling Unrestricted Coreference in OntoNotes*. <http://conll.cemantix.org/2011/introduction.html>, (zitiert auf Seite 12).
- [CRAar] CORCOGLIONITI, Francesco ; ROSPOCHER, Marco ; APROSIO, Alessio P.: Frame-based Ontology Population with PIKES. In: *IEEE Transactions on Knowledge and Data Engineering* (to appear). <http://dx.doi.org/10.1109/TKDE.2016.2602206>. – DOI 10.1109/TKDE.2016.2602206 (zitiert auf Seite 66).
- [CRF03] COHEN, William W. ; RAVIKUMAR, Pradeep ; FIENBERG, Stephen E.: A Comparison of String Distance Metrics for Name-Matching Tasks, 2003, S. 73–78 (zitiert auf Seite 42).
- [CSDS10] CHEN, Desai ; SCHNEIDER, Nathan ; DAS, Dipanjan ; SMITH, Noah A.: SEMAFOR: Frame Argument Resolution with Log-linear Models. In: *Proceedings of the 5th International Workshop on Semantic Evaluation*. Stroudsburg, PA, USA : Association for Computational Linguistics, 2010 (SemEval '10), S. 264–267 (zitiert auf Seite 24).
- [CWB⁺11] COLLOBERT, Ronan ; WESTON, Jason ; BOTTOU, Léon ; KARLEN, Michael ; KAVUKCUOGLU, Koray ; KUKSA, Pavel: Natural Language Processing (Almost) from Scratch. In: *Journal of Machine Learning Research* 12 (2011), Nr. Aug, S. 2493–2537. – ISSN 1533–7928 (zitiert auf den Seiten 65 und 93).
- [Deb] DEBATTY, Thibault: *Java-String-Similarity*. <https://github.com/tdebatty/java-string-similarity>, (zitiert auf Seite 74).
- [DG13] DEL CORRO, Luciano ; GEMULLA, Rainer: ClausIE: Clause-based Open Information Extraction. In: *Proceedings of the 22Nd International Conference on World Wide Web*. Republic and Canton of Geneva, Switzerland : International World Wide Web Conferences Steering Committee, 2013 (WWW '13). – ISBN 978–1–4503–2035–1, S. 355–366 (zitiert auf den Seiten xi, 24 und 29).
- [Dow91] DOWTY, David: Thematic Proto-Roles and Argument Selection. In: *Language* 67 (1991), Nr. 3, S. 547–619. <http://dx.doi.org/10.2307/415037>. – DOI 10.2307/415037. – ISSN 0097–8507 (zitiert auf Seite 9).
- [Fel98] FELLBAUM, Christiane: *WordNet: An Electronic Lexical Database*. Cambridge : MIT Press, 1998 (zitiert auf Seite 13).
- [Fel13] FELIX, Sascha W.: *Kognitive Linguistik: Repräsentation und Prozesse*. Springer-Verlag, 2013. – ISBN 978–3–663–05399–6 (zitiert auf Seite 1).
- [FR05] FLEISCHMAN, Michael ; ROY, Deb: Intentional Context in Situated Natural Language Learning. In: *Proceedings of the Ninth Conference on Computational Natural Language Learning*. Stroudsburg, PA, USA : Association for Computational Linguistics, 2005 (CONLL '05), S. 104–111 (zitiert auf Seite 22).
- [Fra96] FRANCK, Dorothea: Kontext und Kotext. In: *Sprachphilosophie*. Walter de Gruyter, 1996 (zitiert auf Seite 6).
- [GOS09] GUARINO, Nicola ; OBERLE, Daniel ; STAAB, Steffen: What Is an Ontology? In: STAAB, Steffen (Hrsg.) ; STUDER, Rudi (Hrsg.): *Handbook on Ontologies*. Springer Berlin Heidelberg, 2009 (International Handbooks on Information Systems). – ISBN 978–3–540–70999–2 978–3–540–92673–3, S. 1–17 (zitiert auf Seite 15).

- [Gru93] GRUBER, Thomas R.: A Translation Approach to Portable Ontology Specifications. In: *Knowl. Acquis.* 5 (1993), Juni, Nr. 2, S. 199–220. <http://dx.doi.org/10.1006/knac.1993.1008>. – DOI 10.1006/knac.1993.1008. – ISSN 1042–8143 (zitiert auf Seite 16).
- [Gün15] GÜNES, Zeynep: *Aufbau eines Sprachkorpus zur Programmierung autonomer Roboter mittels natürlicher Sprache*, Karlsruher Institut für Technologie (KIT) – IPD Tichy, Bachelor’s Thesis, Mai 2015 (zitiert auf den Seiten 18, 89, 90 und 98).
- [i2b] *i2b2: Informatics for Integrating Biology & the Bedside: Coreference Resolution Challenge*. <https://www.i2b2.org/NLP/Coreference/>, (zitiert auf Seite 26).
- [JM09] JURAFSKY, Daniel ; MARTIN, James H.: *Speech and Language Processing (2Nd Edition)*. Upper Saddle River, NJ, USA : Prentice-Hall, Inc., 2009. – ISBN 0–13–187321–0 (zitiert auf den Seiten 5, 9, 10 und 11).
- [JR13] JINDAL, Prateek ; ROTH, Dan: Using domain knowledge and domain-inspired discourse model for coreference resolution for clinical narratives. In: *Journal of the American Medical Informatics Association* 20 (2013), März, Nr. 2, S. 356–362. <http://dx.doi.org/10.1136/amiajnl-2011-000767>. – DOI 10.1136/amiajnl-2011-000767. – ISSN 1067–5027, 1527–974X (zitiert auf den Seiten 25 und 29).
- [Kam81] KAMP, Hans: A Theory of Truth and Semantic Representation. In: PORTNER, P. (Hrsg.) ; PARTEE, B. H. (Hrsg.): *Formal Semantics - the Essential Readings*. Blackwell, 1981, S. 189–222 (zitiert auf Seite 5).
- [Kar] KARLSRUHER INSTITUT FÜR TECHNOLOGIE (KIT) – IPD TICHY: *PARSE Projekt Homepage*. <https://code.ipd.kit.edu/weigelt/parse/wikis/home>, (zitiert auf Seite 2).
- [Kie16] KIESEL, Viktor: *Optimierung von POS-Tagger-Ergebnissen*, Karlsruher Institut für Technologie (KIT) – IPD Tichy, Bachelor’s Thesis, April 2016. https://code.ipd.kit.edu/weigelt/parse/wikis/Theses/kiesel_ba (zitiert auf Seite 19).
- [KKRP06] KIPPER, Karin ; KORHONEN, Anna ; RYANT, Neville ; PALMER, Martha: Extending VerbNet with novel verb classes. In: *Proceedings of LREC* Bd. 2006, Citeseer, 2006, S. 1 (zitiert auf Seite 14).
- [KM03] KLEIN, Dan ; MANNING, Christopher D.: Accurate Unlexicalized Parsing. In: *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 1*. Stroudsburg, PA, USA : Association for Computational Linguistics, 2003 (ACL ’03), S. 423–430 (zitiert auf Seite 24).
- [Koc15] KOCYBIK, Markus: *Projektion von gesprochener Sprache auf eine Handlungsrepräsentation*, Karlsruher Institut für Technologie (KIT) – IPD Tichy, Bachelor’s Thesis, Juli 2015 (zitiert auf den Seiten 19, 65 und 67).
- [Lin98] LIN, Dekang: An information-theoretic definition of similarity. In: *ICML* Bd. 98, Citeseer, 1998, S. 296–304 (zitiert auf Seite 45).
- [LL05] LIU, Hugo ; LIEBERMAN, Henry: Metafor: Visualizing Stories As Code. In: *Proceedings of the 10th International Conference on Intelligent User Interfaces*. New York, NY, USA : ACM, 2005 (IUI ’05). – ISBN 1–58113–894–6, S. 305–307 (zitiert auf Seite 21).

- [Loc05] LOCKWOOD, David: *Syntactic Analysis and Description: A Constructional Approach*. A&C Black, 2005. – ISBN 978-1-4411-6256-4 (zitiert auf Seite 34).
- [LPC⁺11] LEE, Heeyoung ; PEIRSMAN, Yves ; CHANG, Angel ; CHAMBERS, Nathanael ; SURDEANU, Mihai ; JURAFSKY, Dan: Stanford's multi-pass sieve coreference resolution system at the CoNLL-2011 shared task. In: *Proceedings of the Fifteenth Conference on Computational Natural Language Learning: Shared Task*, Association for Computational Linguistics, 2011, S. 28–34 (zitiert auf den Seiten 12 und 106).
- [MMS93] MARCUS, Mitchell P. ; MARCINKIEWICZ, Mary A. ; SANTORINI, Beatrice: Building a Large Annotated Corpus of English: The Penn Treebank. In: *Comput. Linguist.* 19 (1993), Juni, Nr. 2, S. 313–330. – ISSN 0891-2017 (zitiert auf den Seiten 7 und 13).
- [Mor38] MORRIS, Charles W.: *Foundations of the Theory of Signs*. Monograph Collection (Matt - Pseudo), 1938 (zitiert auf Seite 4).
- [MTLS15] MISRA, Dipendra K. ; TAO, Kejia ; LIANG, Percy ; SAXENA, Ashutosh: Environment-Driven Lexicon Induction for High-Level Instructions. In: *ACL (1)* (2015), S. 992–1002 (zitiert auf den Seiten xi, 21, 22 und 29).
- [Pas15] PASKARAN, Dinesh: *Evaluation unterschiedlicher Spracherkennungssysteme in der Domäne Humanoide Robotik*, Karlsruher Institut für Technologie (KIT) – IPD Tichy, Bachelor's Thesis, November 2015 (zitiert auf Seite 18).
- [PDG12] PRESUTTI, Valentina ; DRAICCHIO, Francesco ; GANGEMI, Aldo: Knowledge Extraction Based on Discourse Representation Theory and Linguistic Frames. In: *Proceedings of the 18th International Conference on Knowledge Engineering and Knowledge Management*. Berlin, Heidelberg : Springer-Verlag, 2012 (EKAW'12). – ISBN 978-3-642-33875-5, S. 114–129 (zitiert auf den Seiten 24 und 29).
- [Ped] PEDERSEN, Ted: *WordNet::Similarity*. <http://wn-similarity.sourceforge.net/>, (zitiert auf Seite 77).
- [PGK05] PALMER, Martha ; GILDEA, Daniel ; KINGSBURY, Paul: The Proposition Bank: An Annotated Corpus of Semantic Roles. In: *Comput. Linguist.* 31 (2005), März, Nr. 1, S. 71–106. <http://dx.doi.org/10.1162/0891201053630264>. – DOI 10.1162/0891201053630264. – ISSN 0891-2017 (zitiert auf Seite 13).
- [PK08] PFISTER, Beat ; KAUFMANN, Tobias: *Sprachverarbeitung: Grundlagen und Methoden der Sprachsynthese und Spracherkennung*. Springer-Verlag, 2008. – ISBN 978-3-540-75910-2. – Google-Books-ID: uacnBAAAQBAJ (zitiert auf Seite 4).
- [Sch15] SCHNEIDER, Michael: *Entwurf einer Handlungsrepräsentation für gesprochene Sprache*, Karlsruher Institut für Technologie (KIT) – IPD Tichy, Bachelor's Thesis, Mai 2015 (zitiert auf Seite 18).
- [Sch16] SCHEU, Sven: *Aufbereitung von Spracherkennerausgaben*, Karlsruher Institut für Technologie (KIT) – IPD Tichy, Master's Thesis, Juni 2016. https://code.ipd.kit.edu/weigelt/parse/wikis/Theses/scheu_ma (zitiert auf Seite 19).
- [SM03] STRUBE, Michael ; MÜLLER, Christoph: A Machine Learning Approach to Pronoun Resolution in Spoken Dialogue. In: *Proceedings of the 41st Annual*

- Meeting on Association for Computational Linguistics - Volume 1*. Stroudsburg, PA, USA : Association for Computational Linguistics, 2003 (ACL '03), S. 168–175 (zitiert auf Seite 51).
- [sta] *Stanford CoreNLP – a suite of core NLP tools*. <https://stanfordnlp.github.io/CoreNLP/>, (zitiert auf den Seiten 12, 71 und 98).
- [Ste16] STEURER, Vanessa: *Strukturerkennung von Bedingungen in gesprochener Sprache*, Karlsruher Institut für Technologie (KIT) – IPD Tichy, Bachelor's Thesis, April 2016. https://code.ipd.kit.edu/weigelt/parse/wikis/Theses/steuerer_ba (zitiert auf den Seiten 19, 71, 89 und 98).
- [TA06] TETREAULT, Joel R. ; ALLEN, James F.: Dialogue Structure and Pronoun Resolution. 2006. – Forschungsbericht (zitiert auf Seite 51).
- [TB00] TJONG KIM SANG, Erik F. ; BUCHHOLZ, Sabine: Introduction to the CoNLL-2000 Shared Task: Chunking. In: *Proceedings of the 2Nd Workshop on Learning Language in Logic and the 4th Conference on Computational Natural Language Learning - Volume 7*. Stroudsburg, PA, USA : Association for Computational Linguistics, 2000 (ConLL '00), S. 127–132 (zitiert auf Seite 8).
- [TD03] TJONG KIM SANG, Erik F. ; DE MEULDER, Fien: Introduction to the CoNLL-2003 Shared Task: Language-independent Named Entity Recognition. In: *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003 - Volume 4*. Stroudsburg, PA, USA : Association for Computational Linguistics, 2003 (CONLL '03), S. 142–147 (zitiert auf Seite 9).
- [Van76] VAN DIJK, Teun A.: Philosophy of action and theory of narrative. In: *Poetics* 5 (1976), Dezember, Nr. 4, S. 287–338. [http://dx.doi.org/10.1016/0304-422X\(76\)90014-0](http://dx.doi.org/10.1016/0304-422X(76)90014-0). – DOI 10.1016/0304-422X(76)90014-0. – ISSN 0304-422X (zitiert auf Seite 4).
- [WT15] WEIGELT, S. ; TICHY, W.F.: Poster: ProNat: An Agent-Based System Design for Programming in Spoken Natural Language. In: *Software Engineering (ICSE), 2015 IEEE/ACM 37th IEEE International Conference on* Bd. 2, 2015, S. 819–820 (zitiert auf den Seiten xi und 18).

Anhang

A. Tagsets und Rolesets

A.1. Penn Treebank Tagset

Tag	Description
CC	Coordinating conjunction
CD	Cardinal number
DT	Determiner
EX	Existential <i>there</i>
FW	Foreign word
IN	Preposition or subordinating conjunction
JJ	Adjective
JJR	Adjective, comparative
JJS	Adjective, superlative
LS	List item marker
MD	Modal
NN	Noun, singular or mass
NNS	Noun, plural
NNP	Proper noun, singular
NNPS	Proper noun, plural
PDT	Predeterminer
POS	Possessive ending
PRP	Personal pronoun
PRP	Possessive pronoun
RB	Adverb
RBR	Adverb, comparative
RBS	Adverb, superlative
RP	Particle
SYM	Symbol
TO	<i>to</i>
UH	Interjection
VB	Verb, base form
VBD	Verb, past tense
VBG	Verb, gerund or present participle
VBN	Verb, past participle

Tag	Description
VBP	Verb, non-3rd person singular present
VBZ	Verb, 3rd person singular present
WDT	Wh-determiner
WP	Wh-pronoun
WP	Possessive wh-pronoun
WRB	Wh-adverb

A.2. VerbNet Thematic Roles

Thematic Role	Example classes that use them
Actor:	used for some communication classes (e.g., Chitchat-37.6, Marry-36.2, Meet-36.2) when both arguments can be considered symmetrical (pseudo-agents).
Agent:	generally a human or an animate subject. Used mostly as a volitional agent, but also used in VerbNet for internally controlled subjects such as forces and machines.
Asset:	used for the Sum of Money Alternation, present in classes such as Build-26.1, Get-13.5.1, and Obtain-13.5.2 with „currency“ as a selectional restriction.
Attribute:	attribute of Patient/Theme refers to a quality of something that is being changed, as in (The price)att of oil soared. At the moment, we have only one class using this role Calibratable cos-45.6 to capture the Possessor subject Possessor-Attribute Factoring Alternation. The selectional restriction „scalar“ (defined as a quantity, such as mass, length, time, or temperature, which is completely specified by a number on an appropriate scale) ensures the nature of Attribute.
Beneficiary:	the entity that benefits from some action. Used by such classes as Build-26.1, Get-13.5.1, Performance-26.7, Preparing-26.3, and Steal-10.5. Generally introduced by the preposition „for“, or double object variant in the benefactive alternation.
Cause:	used mostly by classes involving Psychological Verbs and Verbs Involving the Body.
Location, Destination, Source:	used for spatial locations.
Destination:	end point of the motion, or direction towards which the motion is directed. Used with a „to“ prepositional phrase by classes of change of location, such as Banish-10.2, and Verbs of Sending and Carrying. Also used as location direct objects in classes where the concept of destination is implicit (and location could not be Source), such as Butter-9.9, or Image impression-25.1.
Source:	start point of the motion. Usually introduced by a source prepositional phrase (mostly headed by „from“ or „out of“). It is also used as a direct object in such classes as Clear-10.3, Leave-51.2, and Wipe instr-10.4.2.
Location:	underspecified destination, source, or place, in general introduced by a locative or path prepositional phrase.

Experiencer:	used for a participant that is aware or experiencing something. In VerbNet it is used by classes involving Psychological Verbs, Verbs of Perception, Touch, and Verbs Involving the Body.
Extent:	used only in the Calibratable-45.6 class, to specify the range or degree of change, as in The price of oil soared (10%) <i>ext</i> . This role may be added to other classes.
Instrument:	used for objects (or forces) that come in contact with an object and cause some change in them. Generally introduced by a „with“ prepositional phrase. Also used as a subject in the Instrument subject Alternation and as a direct object in the Poke-19 class for the Through/With Alternation and in the Hit-18.1 class for the With/Against Alternation.
Material and Product:	used in the Build and Grow classes to capture the key semantic components of the arguments. Used by classes from Verbs of Creation and Transformation that allow for the Material/Product Alternation.
Material:	start point of transformation.
Product:	end result of transformation.
Patient:	used for participants that are undergoing a process or that have been affected in some way. Verbs that explicitly (or implicitly) express changes of state have Patient as their usual direct object. We also use Patient1 and Patient2 for some classes of Verbs of Combining and Attaching and Verbs of Separating and Disassembling, where there are two roles that undergo some change with no clear distinction between them.
Predicate:	used for classes with a predicative complement.
Recipient:	target of the transfer. Used by some classes of Verbs of Change of Possession, Verbs of Communication, and Verbs Involving the Body. The selection restrictions on this role always allow for animate and sometimes for organization recipients.
Stimulus:	used by Verbs of Perception for events or objects that elicit some response from an experiencer. This role usually imposes no restrictions.
Theme:	used for participants in a location or undergoing a change of location. Also, Theme1 and Theme2 are used for a few classes where there seems to be no distinction between the arguments, such as Differ-23.4 and Exchange-13.6 classes.
Time:	class-specific role, used in Begin-55.1 class to express time.
Topic:	topic of communication verbs to handle theme/topic of the conversation or transfer of message. In some cases, like the verbs in the Say-37.7 class, it would seem better to have „Message“ instead of „Topic“, but we decided not to proliferate the number of roles.

B. PARSE

B.1. Domänenontologie-Schnittstelle

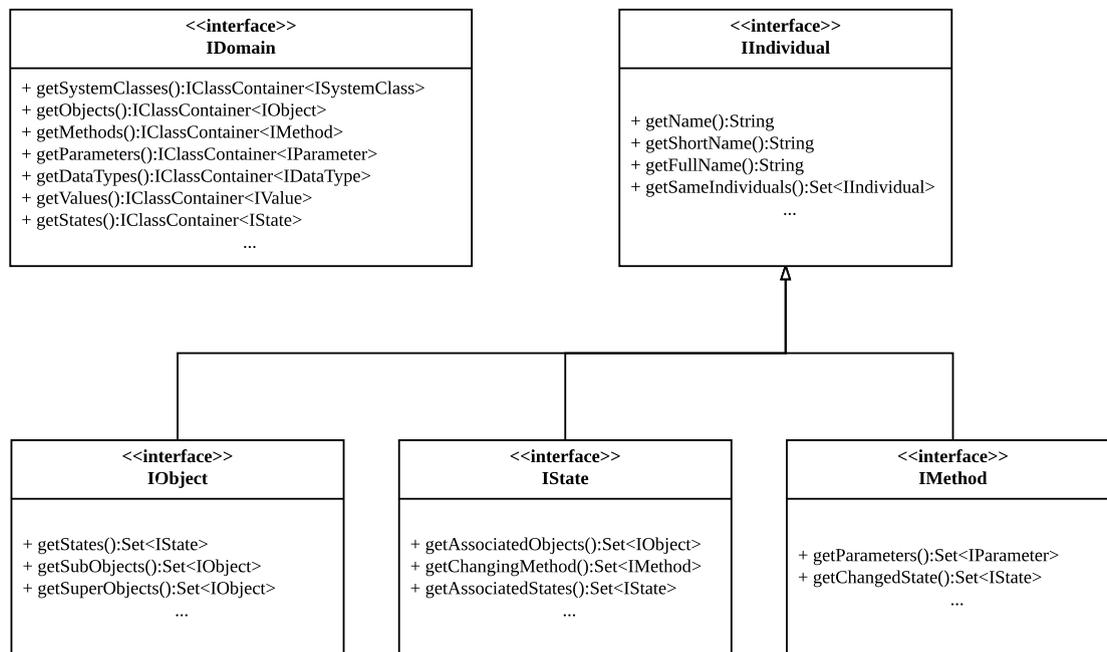


Abbildung B.1.: Ausschnitt aus der Domänenontologie-Schnittstelle

B.2. Domänenontologie-Abbild

Hier ist ein Abbild der zu Evaluation genutzten Domänenontologie mit den beinhalteten Ausprägungen und den für den Kontextaufbau relevanten Relationen dargestellt:

Konzeptname	Ausprägungen	Beschreibung
Thing		Oberkonzept der Ontologie
System	Armar, Roboter, Robo	Repräsentation eines ansprechbaren Systems
Object	Fridge, Fridge.Door, Apollinaris, AppleJuice, CerealBox, CocoaBox, CookieBox, Cup, Cupboard, Cupboard.Door, Dishes, Dishwasher, Dishwasher.Door, InstantMeal, Ketchup, Lemon, Microwave, Microwave.Door, Milk, MultivitaminJuice, Nesquik, Orange, OrangeJuice, Plate, Popcorn, Table, Television, TV, VodkaBottle, Water, WaterBottle	auftretende Objekte
Method	open, close, bring, take, clean, come, fill, get, go, grab, hand, move, pour, put, show, turnOff, turnOn, warm, wash	ausführbare Methoden
State	opened, closed full, empty, cleaned, dirty, on, off, cold, warmed	mögliche Zustände
DataType	boolean, String, char, double, float, int, long, Object, short	mögliche Datentypen
Parameter		mögliche Parameter für Methoden
Value		mögliche Werte der Parameter

equalConcept-Relation

```

Television --equal--> TV
WaterBottle --equal--> Apollinaris
Apollinaris --equal--> WaterBottle
TV --equal--> Television
Nesquik --equal--> CocoaBox
CocoaBox --equal--> Nesquik
wash --equal--> clean
grab --equal--> take
take --equal--> grab
move --equal--> go
clean --equal--> wash
go --equal--> move
get --equal--> bring
bring --equal--> get

```

subObject-Relation

Struktur: Object --sub--> Object

WaterBottle --sub--> Water
 Microwave --sub--> Microwave.Door
 Fridge --sub--> Fridge.Door
 Cupboard --sub--> Cupboard.Door
 Dishwasher --sub--> Dishwasher.Door
 Dishes --sub--> Plate

SuperObject-Relation

Struktur: Object --super--> Object

Cupboard.Door --super--> Cupboard
 Dishwasher.Door --super--> Dishwasher
 Microwave.Door --super--> Microwave
 Fridge.Door --super--> Fridge
 Water --super--> WaterBottle
 Plate --super--> Dishes

HasState-Relation

Struktur: Object --hasState--> State

Television --hasState--> on
 Television --hasState--> off
 CerealBox --hasState--> empty
 CerealBox --hasState--> full
 VodkaBottle --hasState--> empty
 VodkaBottle --hasState--> full
 WaterBottle --hasState--> empty
 WaterBottle --hasState--> full
 WaterBottle --hasState--> closed
 WaterBottle --hasState--> opened
 Microwave --hasState--> dirty
 Microwave --hasState--> cleaned
 Microwave --hasState--> on
 Microwave --hasState--> off
 Fridge --hasState--> dirty
 Fridge --hasState--> cleaned
 Fridge --hasState--> on
 Fridge --hasState--> off
 Apollinaris --hasState--> empty
 Apollinaris --hasState--> full
 Apollinaris --hasState--> closed
 Apollinaris --hasState--> opened
 Plate --hasState--> empty
 Plate --hasState--> dirty
 Plate --hasState--> full
 Plate --hasState--> cleaned
 Dishwasher.Door --hasState--> dirty
 Dishwasher.Door --hasState--> cleaned

```

Dishwasher.Door --hasState--> closed
Dishwasher.Door --hasState--> opened
InstantMeal --hasState--> cold
InstantMeal --hasState--> warmed
Cupboard --hasState--> dirty
Cupboard --hasState--> cleaned
Cup --hasState--> empty
Cup --hasState--> dirty
Cup --hasState--> full
Cup --hasState--> cleaned
TV --hasState--> on
TV --hasState--> off
Cupboard.Door --hasState--> dirty
Cupboard.Door --hasState--> cleaned
Cupboard.Door --hasState--> closed
Cupboard.Door --hasState--> opened
Table --hasState--> dirty
Table --hasState--> cleaned
Microwave.Door --hasState--> dirty
Microwave.Door --hasState--> cleaned
Microwave.Door --hasState--> closed
Microwave.Door --hasState--> opened
Fridge.Door --hasState--> dirty
Fridge.Door --hasState--> cleaned
Fridge.Door --hasState--> closed
Fridge.Door --hasState--> opened
Dishwasher --hasState--> dirty
Dishwasher --hasState--> cleaned
Dishwasher --hasState--> on
Dishwasher --hasState--> off
Dishes --hasState--> dirty
Dishes --hasState--> cleaned

```

AssociatedState-Relation

Struktur: State --associatedState--> State

```

off --associatedState--> on
empty --associatedState--> full
closed --associatedState--> opened
warmed --associatedState--> cold
cold --associatedState--> warmed
cleaned --associatedState--> dirty
full --associatedState--> empty
dirty --associatedState--> cleaned
opened --associatedState--> closed
on --associatedState--> off

```

ChangesStateTo-Relation

Struktur: Method --changesStateTo--> State

```

open --changesStateTo--> opened
wash --changesStateTo--> cleaned

```

```
close --changesStateTo--> closed
turnOff --changesStateTo--> off
fill --changesStateTo--> full
clean --changesStateTo--> cleaned
pour --changesStateTo--> empty
turnOn --changesStateTo--> on
warm --changesStateTo--> warmed
```

C. Konfigurationsdatei der Korreferenzimplementierung

Die Konfigurationsdatei der Korreferenzanalyse-Implementierung enthält alle für den Korreferenzanalyse-Agenten von außen einstellbaren Parameter. Dies sind vor allem die auszuführenden Siebe in ihrer Reihenfolge sowie der Modus der Siebarchitektur.

```
#Options :

#Anaphora Resolution to be performed
ANAPHORA_RESOLUTION=true

#General Sieves to be executed for all types of Anaphora Resolution
GENERAL_SIEVES=precedingEntities , typeMatch , alternativeStatement ,
    sameAction , reflexive

#Sieves to be executed for Object Anaphora Resolution
OBJECT_SIEVES=grammaticalNumber , instructionNumber , srlRoleMatch ,
    actionPossible , locativeSieve , syntacticalDistance

#Sieves to be executed for Group Anaphora Resolution
GROUP_SIEVES=grammaticalNumber , instructionNumber , actionPossible ,
    locativeSieve , syntacticalDistance

#Sieves to be executed for Speaker Anaphora Resolution
SPEAKER_SIEVES=grammaticalNumber , syntacticalDistance

#Sieves to be executed for subject Anaphora Resolution
SUBJECT_SIEVES=grammaticalNumber , genderMatch , locativeSieve ,
    syntacticalDistance

#subject Identity Resolution to be performed
SUBJECT_IDENTITY_RESOLUTION=true
#Sieves to be executed for subject Identity Resolution
SUBJECT_IDENTITY_SIEVES=precedingEntities , typeMatch ,
    alternativeStatement , grammaticalNumber , conceptMatch , genderMatch ,
    nameMatch , alreadyRelated , syntacticalDistance

#Object Identity Resolution to be performed
OBJECT_IDENTITY_RESOLUTION=true
#Sieves to be executed for Object Identity Resolution
OBJECT_IDENTITY_SIEVES=precedingEntities , typeMatch , alternativeStatement
    , grammaticalNumber , conceptMatch , generalDet , objectMatch ,
    alreadyRelated , syntacticalDistance

#Sieve Modus (HARD -> Sieves remove Candidates completely | SOFT ->
    Sieves set Confidences)
SIEVE_MODUS=HARD

#PossessivePronoun Anaphora Resolution to be performed
POSS_ANAPHORA_RESOLUTION=true

#General Sieves to be executed for all types of Anaphora Resolution
POSS_GENERAL_SIEVES=precedingEntities , typeMatch , alternativeStatement
```

```

#Sieves to be executed for Object Anaphora Resolution
POSS_OBJECT_SIEVES=grammaticalNumber , instructionNumber , locativeSieve ,
    possessiveMeronym , syntacticalDistance

#Sieves to be executed for Group Anaphora Resolution
POSS_GROUP_SIEVES=grammaticalNumber , instructionNumber , locativeSieve ,
    possessiveMeronym , syntacticalDistance

#Sieves to be executed for Speaker Anaphora Resolution
POSS_SPEAKER_SIEVES=grammaticalNumber , syntacticalDistance

#Sieves to be executed for subject Anaphora Resolution
POSS_subJECT_SIEVES=grammaticalNumber , genderMatch , locativeSieve ,
    syntacticalDistance

```

Quelltextausschnitt 8.1: Konfigurationsdatei der Korreferenzanalyse

D. Erweiterung des Korpus

Der Sprachkorpus des *PARSE*-Projektes wurde um zwei weitere Szenarien ergänzt, welche dahingehend entworfen wurden mehr Objekte und Aktionen auf diesen Objekten zu fördern und damit einen größeren Kontext sowie mehr Korreferenzen zu erzeugen. Die Probanden wurden dabei zunächst über die Projektsituation aufgeklärt weshalb ihnen bekannt ist, dass das Zielsystem *ARMAR-III* in einer Küchenumgebung darstellt und dem Roboter grundlegende Aktionen bekannt sind.

D.1. Szenario 1: Fill the cup

In this scene you want the robot to fill your cup with water and bring it to you. The cup is located on the table among other things (Abbildung D.3a) and has the color green. The water is in the fridge (Abbildung D.2) and the door of the fridge is closed (Abbildung D.2a). Afterwards the robot should take the red cups out of the dishwasher (Abbildung D.3b) and put them into the cupboard.

D.2. Szenario 2: Prepare a meal

In this scene you want the robot to prepare your meal. Therefore he should take any of the plates, which are in the dishwasher (Abbildung D.3b). Give the plate a wash and put the instant meal from the fridge (Abbildung D.2b) on the plate. Then warm it in the microwave (Abbildung D.4). Afterwards he should put the plate on the table.



(a) The fridge (closed)



(b) The fridge (opened)



(c) The water in the fridge

Abbildung D.2.: Pictures of the fridge



(a) Objects on the table



(b) The dishwasher (opened)

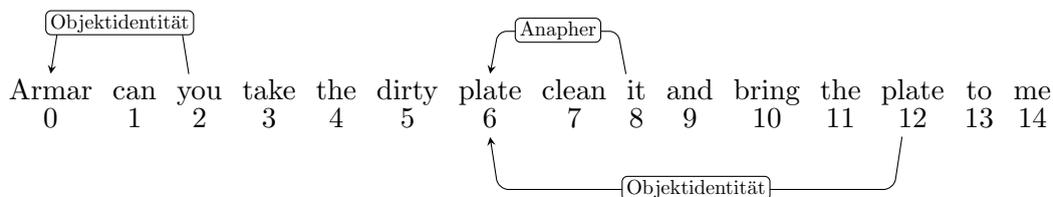
Abbildung D.3.: Table and dishwasher



Abbildung D.4.: View at the kitchen with marked microwave M

E. Musterlösung der Korreferenzanalyse

Um nachvollziehen zu können, wie die Musterlösungen für die Korreferenzen in einer Transkription gebildet wurden und wie diese im Werkzeug verwendet werden wird anhand des folgenden Beispiels die Form der Musterlösung erklärt:



Hier ist das Beispiel mit den Positionsnummern in der Aussage versehen dargestellt sowie die als Musterlösung beschriebenen Referenzen als Kanten hinzugefügt. In der Musterlösung werden die Korreferenzkanten einzeln abhängig von der Positionsnummer des Wortes definiert. Hierbei wird immer die Positionsnummer des Wortes gewählt, welche den Kopf der Phrase bildet, also entweder das zuletzt genannte Nomen der Nominalphrase oder das Pronomen selbst. Somit wäre die Angabe der Musterlösung für die Referenz *the plate* zu *the dirty plate* die Korreferenz mit dem Start 12 und dem Ende 6. Die Musterlösungen werden dabei in der XML-Datei `korpus.xml` zusammen mit den Transkriptionen beschrieben und hätten für dieses Beispiel die Form:

```
<korpus>
<text name="bsp">Armar can you take the dirty plate clean it and bring
  the plate to me
  <coref start="2" end="0"/>
  <coref start="8" end="6"/>
  <coref start="12" end="6"/>
</text>
</korpus>
```

Für den letztlichen Abgleich mit den Referenzen der Werkzeuge wird im Fall von Stanford die **Mention** bzw. für den Agenten die **Entity** gesucht, welche das Wort mit der Positionsnummer enthält und zwischen diesen nach einer Referenz in die entsprechende Richtung gesucht.