Institut für Theoretische Informatik
Interactive Systems Labs
Fakultät für Informatik
Universität Karlsruhe (TH)

Lernen von Objekten und deren Bedeutung im Dialog

Diplomarbeit
von

Cand. Inform. Daniel Neubig

Betreuer:
Prof. Dr. Alex Waibel
Dipl. Inform. Hartwig Holzapfel

19. Januar 2008

Erklärung

Hiermit erkläre ich, dass ich diese Diplomarbeit selbständig verfasst, noch nicht anderweitig für Prüfungszwecke vorgelegt, keine anderen als die angegebenen Quellen oder Hilfsmittel verwendet, sowie wörtliche und sinngemäße Zitate als solche gekennzeichnet habe.

gez. Daniel Neubig, Karlsruhe, den 18.Januar 2008

Abstract

Learning objects and their meaning in a dialogue means, that a robot has to be able to identify objects visually and has to know the type and properties of an object. Till now, research was done in all these fields for its own, but nobody has considered all three aspects all together. Only an integrated system of visual object recognition, speech recognition and object modeling can hold a conversation with a human.

This diploma thesis introduces a dialogue system, which is able to learn real objects from untrained users. An object model and an ontology are presented, which are able to model a dynamic set of known objects and a set of objects in the actual scenario. Errors are solved within the dialogue. An algorithm is introduced, which is able to learn semantic classes of objects in a short dialogue by combining an initial one shot learning with the possibility to specify the class in more detail. Even though the dialog uses little more interaction, it is better than a single one shot learning algorithm because the error rate is lower and the learned objects are classified more precise.

Danksagung

Ich danke meinem Betreuer Hartwig Holzapfel, der mir während meiner ganzen Zeit am ISL immer mit Rat und Tat zur Seite stand und von dem ich vieles lernen konnte. Auch danke ich Pedram Azad, der mir mit dem Objekterkenner eine wichtige Grundlage für meine Applikationen zur Verfügung stellte. Mein größter Dank gebührt den Testpersonen, die nicht nur freiwillig an den langen Testreihen mitgemacht haben, sondern die auch mit viel Spaß und Unterstützung lange über Sinn und Zweck von Dialogen mit Robotern diskutiert haben.

Inhalt

Abbildungsverzeichnis

Tabellenverzeichnis

1. Einleitung

Das heute vorherrschende Bild eines Roboters ist geprägt durch bekannte Roboter aus Filmen oder Romanen. Tritt eine Laie an einen real existierenden Roboter heran, erwartet er Fähigkeiten wie die eines C3POs aus „Star Wars", eines Bishops aus „Alien", eines Marvins aus „Per Anhalter durch die Galaxis" oder eines Datas aus „Enterprise The Next Generation". Diese Roboter sind immer in der Lage, Aufgaben zu erledigen, natürlich mit dem Menschen zu interagieren und solange keine menschlichen Gefühlsregungen nötig sind, gibt es kaum Grenzen. Roboter können selbstverständlich Aufgaben erkennen, Objekte manipulieren und Auskünfte geben. Dazu gehört, dass sie mit Menschen über alles reden können und sich zudem noch alles merken können. Für eine Maschine, die nichts vergessen kann und immer alles weiß kann es dann auch nicht schwer sein, sich Objekte zu merken oder diese neu zu lernen, falls es nötig ist.

Es ist nicht möglich, alle Objekte zu kennen, da es zu viele Objekte gibt und immer neue erfunden werden. Also muss der Roboter Methoden und Funktionen integrieren, die es ermöglichen, dass er auf Benutzereingaben sinnvoll und fehlerfrei reagiert und er sich an den Benutzer anpasst. Die komplette Interaktion findet, wie mit anderen Menschen über visuelle und akustische Kommunikation statt. Dabei soll der Benutzer möglichst frei mit dem Roboter reden können. Egal ob er sagt „Gib mir die Kaffeetasse", „Ich will meine Lieblingstasse" oder „Kannst du mir bitte die rote Tasse von da drüben geben?" sollte der Roboter in der Lage sein, das Objekt zu finden und dem Benutzer zu reichen. Ein Serviceroboter, mit dem Laien interagieren können sollen, muss die Dialoge variabel gestalten können. Der Roboter muss in der Lage sein, die Intention des Benutzers zu erkennen und die Aufgabe zu erledigen.

1.1. Zielsetzung

Der Sonderforschungsbereich 588 „Humanoide Roboter – Lernende und kooperierende multimodale Roboter“ (Deutsche Forschungsgemeinschaft, 2001) hat das Ziel, einen Roboter in einer natürlichen Umgebung zu integrieren, der einem Benutzer helfen kann. Die vorliegende Diplomarbeit entstand im Rahmen der Forschungen, um dieses Ziel zu erreichen und einen humanoiden Roboter zu befähigen, dass er erfolgreicher mit Benutzern interagieren kann. Aus den natürlichsprachlichen, intuitiven Äußerungen sollen symbolische Objektbezeichnungen erkannt werden, die in den aufgenommenen Bildern der Kamera Objekte referenzieren. Dabei soll das System in der Lage sein, auf sich verändernde Umgebungen und neue Objektbezeichnungen zu reagieren und online neue Objekte zu lernen. Zudem ist es nötig, dass dieses Wissen in weiteren Interaktionen variabel bleibt und sich der Roboter an verschiedene Arten, das gewünschte Objekt zu referenzieren, anpassen kann.

Die auftretenden Fehler und Schwierigkeiten, die bei dem Matching der Beschreibung und den Objekten entstehen, müssen im Dialog gelöst werden. Dazu soll untersucht werden, wie Personen reale Objekte benennen. Die Beschreibungen werden von dem Benutzer subjektiv erstellt und können sich ändern, wenn der Benutzer zu einem anderen Zeitpunkt das gleiche Objekt beschreibt, wenn sich die Situation ändert und der Benutzer andere Merkmale des Objektes beschreibt, um es von anderen Objekten abzuheben oder wenn andere Benutzer das gleiche Objekt beschreiben. Das System soll trotzdem in der Lage sein, mit wenigen Nachfragen das gewünschte Objekt zu finden und die neuen Bezeichnungen zu lernen.

Es ist nötig, dass der Roboter auch lernt, was für eine Art von Objekt vor ihm steht oder liegt, damit er besser auf den Benutzer reagieren kann. Das heißt, dass festgehalten werden muss, was man mit einem Objekt machen kann. Dadurch wird es möglich, dass der Benutzer nach etwas zu essen oder etwas zu trinken fragt. Diese semantischen Informationen sollen dazu dienen, dass der Roboter die Objekte besser verstehen kann.

Das gesamte System soll auf Armar-III portiert werden und die Benutzertests sollen auf dem zentralen Robotersystem des Sonderforschungsbereichs 588 ausgeführt werden.

1.2. Gliederung

Kapitel 2 beschreibt die Grundlagen dieser Arbeit. Es wird auf Dialogsysteme eingegangen, die es ermöglichen, die Intention des Benutzers zu ermitteln und zu verarbeiten. Spracherkenner, Sprachverstehen und Dialogführung werden vorgestellt. Zusätzlich werden die Methoden der Objekterkennung erläutert.

In Kapitel 3 werden mehrere verwandte Arbeiten vorgestellt, die alle verschiedene Aspekte des „Objekte Lernens" behandeln. Dabei werden die Grundlagen des Lernens neuer Wörter, visueller Eigenschaften und semantischen Informationen über die Objekte behandelt.

In Kapitel 4 wird vorgestellt, wie das Dialogsystem aufgebaut ist und welche Modelle im Hintergrundwissen des Roboters verwendet werden. Dabei wird das Szenario beschrieben und die Modellierung von Objekten mit Hilfe einer Ontologie wird erklärt und der Ablauf des Dialogs in seinen Einzelheiten erläutert.

Kapitel 5 befasst sich mit der Durchführung und Auswertung der Benutzertests. Die Evaluation des Dialogs wird diskutiert und die Ergebnisse werden dargestellt.

Kapitel 6 fasst die Ergebnisse zusammen und gewährt einen Ausblick, welche Folgearbeiten an diese Arbeit anschließen können.

In Kapitel 7 folgt das Literaturverzeichnis der verwendeten Quellen.

2. Grundlagen

Jeden Tag begegnen uns in unserem Alltag neue Objekte. Obwohl jeden Monat neue Modelle von Mobiltelefonen, neue Filme auf DVDs oder neue Produkte im Supermarkt auftauchen, fällt es uns leicht, diese Objekte sofort zu erkennen und uns zu merken. Eine verpackte DVD oder ein Buch erkennt man zum Beispiel sogar ohne das Geschenk zu öffnen am Gewicht, dem Format und der Art, wie sich das Taschenbuch biegen lässt, wie der Umschlag des gebundenen Buches leicht über den Inhalt hinaussteht oder die DVD Schachtel leicht unter dem Druck nachgibt. Wenn man jedoch ein bisher komplett unbekanntes Objekt vor sich hat, wie zum Beispiel ein kompliziertes Werkzeug, ist es für einen Laien schon deutlich schwerer, zu erkennen, was für eine Art dieses Objekt sein könnte. Oft hilft es, jemanden zu fragen, der das Objekt erklären kann und uns lehrt, wie das Objekt überhaupt heißt. Unterbewusst und schnell lernen wir das Cover einer DVD oder die Form und Farbe eines beliebigen Gegenstandes und den Namen des neuen Objekts. Wir verarbeiten also sowohl die visuellen Informationen als auch die akustischen Informationen und erstellen daraus eine interne Repräsentation dieses Objektes. Dabei speichern wir die Bezeichnung, um das Objekt zukünftig wieder referenzieren zu können.

Für Roboter sind alle Objekte erst einmal unbekannt und die Modellierung, was ein Objekt tatsächlich bedeutet, was man damit machen kann und wie man dieses Objekt beschreiben kann, erfordert eine genaue und langwierige Untersuchung über Objekte und Arten von Objekten. Damit der Roboter trotzdem neue Objekte lernen kann, ist es nötig, dass viele Informationen gesammelt werden. Leider sind die Sensoren, die ein Roboter zur Verfügung hat, lange nicht mit unseren Augen, Ohren und Händen zu vergleichen. Der Roboter kann sich nie sicher sein, dass das was er verstanden hat auch tatsächlich gesagt wurde und dass der Gegenstand den er erkannt hat auch wirklich das gemeinte Objekt ist.

In der Interaktion mit einem humanoiden Roboter ist es nötig, dass der Roboter möglichst natürlich agiert und reagiert, um dem Benutzer eine einfache und intuitive Bedienung zu ermöglichen. Damit ein Roboter Objekte lernen kann, sind viele verschiedene Fähigkeiten notwendig. Der Roboter muss das Objekt im Raum detektieren und erkennen, er muss verstehen, was für eine Art Objekt das ist und was an

diesem speziellen Objekt dieser Art besonders ist und er muss verstehen, was der Benutzer sagt um dieses spezielle Objekt (oder irgendein Objekt dieser Art) zu referenzieren.

Hier werden die nötigen Grundlagen erläutert, um in einem Robotersystem neue Objekte zu lernen. Erst wird darauf eingegangen, wie ein Spracherkenner funktioniert und wie neue Wörter im Dialog gelernt werden und dann wird die Erkennung von texturierten Objekten erläutert.

2.1. Spracherkennung und Sprachverstehen

Um mit einem humanoiden Roboter zu interagieren, sollte der Benutzer nicht an eine Tastatur oder eine Maus gebunden sein. Stattdessen sollte es möglich sein, dass der Benutzer vollkommen natürlich das Gespräch mit dem Roboter sucht. Der Benutzer formuliert seine Wünsche und Anweisungen frei und ohne spezielle Kommandos ohne sich an das System anpassen zu müssen.

Spracherkenner werden eingesetzt, damit Benutzer mit den Anwendungen auch in Umgebungen kommunizieren können, wenn sie beide Hände verwenden oder anderweitig beschäftigt sind. Die Wahl, wie Menschen Sprache formulieren kann variabel sein und die Versionen, die das System verstehen muss, um die Intention und die Information des Benutzers zu verstehen sind unterschiedlich. Durch die große Varianz ist es schwer für das System, die Eingabe in eine maschinenlesbare Sprache zu parsen, um darauf reagieren zu können. Ein natürlichsprachliches Dialogsystem besteht, wie in Abbildung 2-1 dargestellt, aus einem Spracherkenner, einer Komponente zum Sprachverstehen und einem Dialogsystem. Jeder Teil dieses Spracherkenners ist für die Verwendung von unbekannten Wörtern, so genannten Out Of Vocabulary Words (OOV) angepasst, damit sich der Spracherkenner an den Benutzer anpassen kann, indem er während der Ausführung neue Wörter lernt.

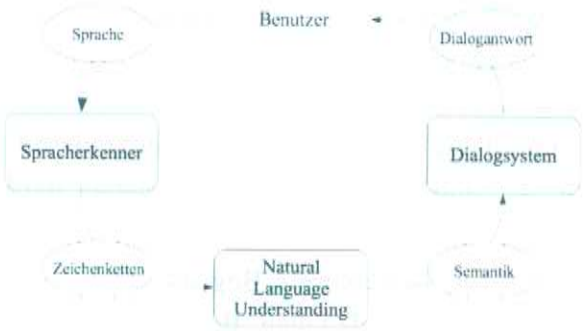

Abbildung 2-1: Zyklus eines natürlichsprachlichen Dialogsystems

Diese Einzelkomponenten werden vom Dialogmanager Tapas (Holzapfel H. , 2005) verwaltet, der eine Entwicklung von multilingualen und multimodalen Dialogsystemen ermöglicht.

Die aufgenommene Sprache wird erst von einem Segmentierer detektiert und dann von dem Spracherkenner, der aus akustischen und sprachlichen Modellen besteht, weiter verarbeitet. Dieser Ansatz ist rein statistisch und nutzt trainierte Modelle, die aus einer großen Menge von Trainingsdaten erstellt wurden. Unbekannte Wörter, die nicht in dem Vokabular enthalten sind, modelliert der Spracherkenner ebenfalls, damit er diese detektieren und erkennen kann, da er immer versucht das aufgenommene mit den trainierten Mustern zu vereinbaren. Wenn es keine Trainingsbeispiele für das Wort gibt, kann der Spracherkenner dieses eigentlich nicht verstehen. Eine weitere Einführung in

die Spracherkennung wird in (Talamazzini, 2001) und (Waibel, 1990) vorgenommen. In dieser Arbeit wird der Spracherkenner Janus verwendet, der in (Finke, 1997) und (Soltau, 2001) genauer vorgestellt wird.

Da neue Objekte gelernt werden sollen, kann es sein, dass der Benutzer Wörter verwendet, die nicht in dem Wortschatz des Erkenners vorhanden sind und somit keine semantische Repräsentation für die Weiterverarbeitung vorliegt. Es ist also nötig, dass der Spracherkenner diese Wörter erkennt oder zumindest detektiert. Die unbekannten Wörter und Objekte werden modelliert, damit der Dialog darauf reagieren kann. Der hier verwendete Spracherkenner wurde von Thomas Schaaf erweitert, um unbekannte Wörter zu detektieren und zu lernen. Die Arbeit wird in Kapitel 3.1.1 vorgestellt.

Das Sprachverstehen verwendet eine Ontologie und eine semantische Grammatik, um die Hypothese, die der Spracherkenner erstellt hat, weiter zu verarbeiten. Die Ontologie (Gruber, 1993) beschreibt Begriffe und deren Zusammenspiel, stellt also dar, wie die Begriffe miteinander in Verbindung gebracht werden können. Jeder Begriff wird als ein Knoten modelliert, der über Kanten, die logische Relationen wie Nachbarschaftsrelationen oder Vererbung darstellen, mit anderen Knoten verbunden ist. In einer Ontologie gibt es Begriffe, die Konzepte beschreiben, Instanzen, die Objekte eines Konzepts sind, Relationen, welche die Eigenschaften eines Konzepts darstellen und Axiome, die direkt Informationen darstellen, die nicht anderweitig beschrieben werden können. Konzepte können auch von anderen Konzepten erben. In dem Fall, dass Objekte mit Hilfe einer Ontologie beschrieben werden, kann man mit den Relationen die Beziehungen zwischen Objekten darstellen oder Klassen von Objekten in Konzepte zusammenfassen. Die Ontologie beschreibt das Kontextwissen, um den Begriff und damit auch das Objekt, das durch den Begriff beschrieben wird, einzuordnen und zu interpretieren. Da nicht sichergestellt ist, dass der Benutzer beim Beschreiben den gleichen Begriff verwendet wie der gelernte Begriff, muss so eine Brücke geschlagen werden zwischen den beiden verschiedenen Beschreibungen und einem tatsächlich existierenden Objekt. In dieser Arbeit wird eine Ontologie verwendet, die Mehrfachvererbung unterstützt, Objektklassen und Objektfunktionalität abbildet und somit die semantischen Klassen von Objekten darstellt. Mehr darüber wird in Kapitel 4.5.3 beschrieben.

In der Linguistik wird dieser Zusammenhang mit dem semiotischen Dreieck (Ogden & Richards, 1923) beschrieben, in dem das Symbol einen Bezug ausdrückt und dieser wiederum ein bestimmtes Bezugsobjekt bestimmt. Eine direkte Verbindung zwischen Symbol und Bezugsobjekt gibt es nicht, diese ist nur gestrichelt angedeutet, da das Symbol nur von jemandem benutzt wird, um den Referenten zu vertreten. Abbildung 2-2 zeigt das Dreieck in der Definition von Ogden und Richards.

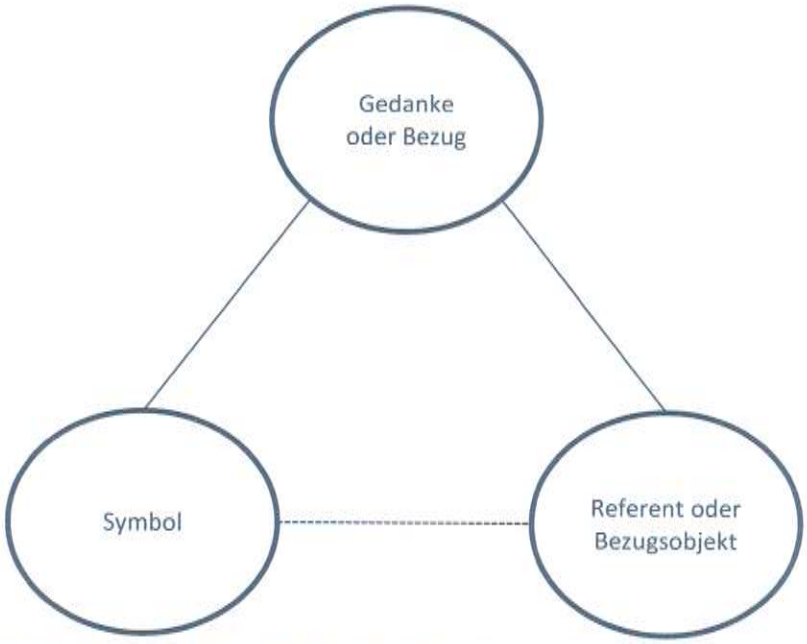

Abbildung 2-2: Das semiotische Dreieck von Ogden und Richards

Die Ontologie beschreibt also für das tatsächlich gesprochene Wort, das Symbol, in dem der Bezeichner fällt, einen Begriff, also die Semantik oder den Bezug. Damit wird ein tatsächlich existierendes Bezugsobjekt, das auf dem Tisch steht, beschrieben. Menschen können die drei Ecken des semiotischen Dreiecks normalerweise leicht miteinander verschmelzen, wenn sie von einem Objekt reden. Die Informationen über das Objekt und der direkte Bezug bilden eine Einheit. Das ist aber nur durch eine einheitliche Definition eines Begriffs unter allen am Gespräch beteiligten Personen möglich. Der Roboter, der ein reales Objekt finden soll, muss in der Lage sein, die Aussagen des Benutzers so zu interpretieren, dass er das gewünschte Objekt findet. Für jedes reale Objekt, das das System kennt, werden mögliche Klassen und Eigenschaften gespeichert, durch die dieses spezielle Objekt beschrieben werden kann. Dadurch werden im Laufe der Zeit die bereits bekannten Objekte immer genauer beschrieben, so dass einmal verwendete Bezeichnungen für dieses Objekt von da an leicht auch verwendet werden können. Diese Funktionalität wird in Kapitel 4.5.4 genauer beschrieben.

Für die Weiterverarbeitung der Benutzeräußerung durch den Dialogmanager werden die Informationen in eine normierte Form gepackt, welche die Semantik in eigens dafür vorgesehene Slots aufteilt. Die Anwendung kann diese Informationen dann weiter verwenden, um die Reaktion zu berechnen oder Fragen vorzubereiten, damit noch fehlende Informationen gesammelt werden können. Diese Merkmalsstruktur, Typed Feature Structure (TFS), wurde von (Carpenter, 1992) eingeführt. Außerdem wird eine an die Begriffsstruktur angepasste Grammatik verwendet, die mögliche Sätze umfasst, die der Benutzer verwenden kann. Aus dem wahrscheinlichsten Satz werden die semantisch wichtigen Inhalte extrahiert. So wird aus den verwendeten Verben die

red cup“ und „bring me the cup in red” werden in dem Zusammenspiel von Ontologie und Grammatik in die gleiche semantische Form umgewandelt:

```
[object_learn:act_bring
    ITEM [object_learn:obj_item
        TYPE_SLOT [„cup“]
        PROP [object_learn:prp_object
            PROP_SLOT [“red”]
        ]
    ]
]
```

Der Objekterkenner ordnet den erkannten Objekten ebenfalls eine Beschreibung zu. Diese wird zusammen mit einer Repräsentation des Objektes im Objektmodell abgespeichert, damit das Objekt gefunden werden kann.

Das Dialogsystem verwendet eine Strategie, nach der das System auf den aktuellen Zustand reagiert. Erst wird aus der TFS der Diskurs ausgelesen und zusammen mit den Dialogzielen in einen Zustand des Systems umgewandelt. Dann entscheidet die programmierte oder gelernte Strategie, welcher Schritt notwendig ist, damit mehr Informationen angesammelt oder bestehende Informationen abgesichert werden oder die gewünschte Aktion ausgeführt wird. Die Strategie versucht erst zu ermitteln, was der Benutzer möchte, also was sein Ziel ist und führt den nächsten Schritt aus. Das kann eine Antwort auf eine Frage des Benutzers sein oder eine Frage an den Benutzer. Diese Strategie ist unabhängig von der verwendeten Sprache oder Domäne und entscheidet alleine auf dem Dialogzustand, welcher Dialogschritt ausgeführt wird.

2.2. Objekterkennung

Die Hypothesen, welche Objekte in der Umgebung des Roboters vorhanden sind, werden von der Objekterkennung geliefert. In dieser Arbeit wird ein Objekterkenner verwendet, der texturierte Objekte erkennen und lernen kann. Einige vorher durch aufwendige 3D Untersuchungen gelernte einfarbige Objekte können ebenfalls erkannt werden. Diese Objekterkennung ist in das Integrated Vision Toolkit (IVT), einer Open Source Bildverarbeitungsbibliothek von Azad Pedram (Azad P. , 2007), integriert. Das IVT ermöglicht einen plattformunabhängigen und komfortablen Zugriff auf OpenCV Funktionalitäten. Die Kameras werden für die Segmentierung von Vordergrund und Hintergrund mit Hilfe eines Schachbrettmusters kalibriert.

Für die Erkennung von texturierten Objekten werden sogenannte SIFT Features detektiert und gespeichert. SIFT Features sind Scale Invariant Feature Transforms, die verwendeten Merkmale der Objekte sind weitgehend unabhängig von der Skalierung, dem Blickwinkel, der Rotation, Lichtverhältnissen und der Position im Bild. In (Lowe, 2004) und werden SIFT Features vorgestellt. Dieses Vorgehen erfordert, dass die Objekte eine Texturierung haben, damit einzelne Bereiche überhaupt solche SIFT Features darstellen. Deshalb werden einfarbige 3D Objekte, wie rote Becher oder blaue Teller durch Farbsegmentierung und Abgleich auf vorher genau gelernten Umrisse gefunden. Das online Lernen von 3D Objekten ist nicht möglich, weil diese von einem 3D Scanner untersucht werden müssen, damit ein Objektmodell erstellt werden kann.

Bei der Berechnung von SIFT Features werden zuerst mit allen Skalierungen und an allen Positionen im Bild Schlüsselpunkte gesucht, die sich zu Nachbarpunkten am meisten unterscheiden. Diese sollen sich auch bei verschiedener Rotation und Skalierung immer noch detektieren lassen. Das wird gewährleistet, indem man das Bild mit einer Gaußfunktion faltet und somit einen Skalenraum (scale space) erhält. Auf einem DoG (Difference of Gaussian) mit einem konstant verschobenen scale space werden dann alle Punkte mit ihren Nachbarn verglichen. Wenn dieser Punkt ein Extremwert ist, wird er als mögliches Merkmal gespeichert. Gerade Ecken, Kanten und andere interessante Punkte in einer Textur können solche Merkmale sein.

Im zweiten Schritt werden die Schlüsselpunkte aussortiert, die nicht genügend Kontrast zu ihren Nachbarn haben. Deshalb wird für jedes Merkmal ein Deskriptor berechnet, der zum Vergleich mit den anderen dient. Damit man Rotationsinvarianz erreicht, wird jedem Schlüsselpunkt eine Richtung zugewiesen und der Gradient wird dann relativ zu dieser Richtung aufgebaut. Zusammen mit der Richtung ergibt der Gradient den Deskriptor. Die Merkmale werden auf verschiedene Stabilitätskriterien getestet. Sie sind zum Beispiel nur stabil, wenn sie nicht auf Kanten liegen und genügend Kontrast zu ihren Nachbarn haben. Nur stabile Merkmale werden verwendet.

Beim Lernen von Objekten werden diese Merkmale gesucht und mit allen bekannten Objekten verglichen. Wenn kein bekanntes Objekt zugeordnet werden kann, und trotzdem genügend Merkmale im Bild gefunden wurden, werden diese mit dem Bezeichner für ein neues Objekt in der Datenbank abgelegt.

Um Objekte wiederzuerkennen, werden wieder Merkmale berechnet, die mit den Features der Objekte in der Datenbasis verglichen werden. Die gefundenen Objekte werden mit dem Bezeichner und der Klasse, unter der sie in der Datenbank abgelegt wurden in das Objektmodell geschrieben und zusätzlich wird festgehalten, an welcher Stelle das Objekt gefunden wurde. Falls keine bekannten Objekte gefunden wurden, wird überprüft ob sich überhaupt ein Objekt im Sichtbereich befindet. Falls genügend Merkmale gefunden werden, wird das Objekt gelernt. Abbildung 2-3 zeigt ein Objekt mit den Merkmalen, die erkannt wurden. Dabei werden Vordergrund und Hintergrund segmentiert und Merkmale detektiert. Die grünen und roten Punkte stellen die gelernten Merkmale dar. In einem zweiten Schritt erkennt ein zweiter Objekterkenner farbige 3D Objekte.

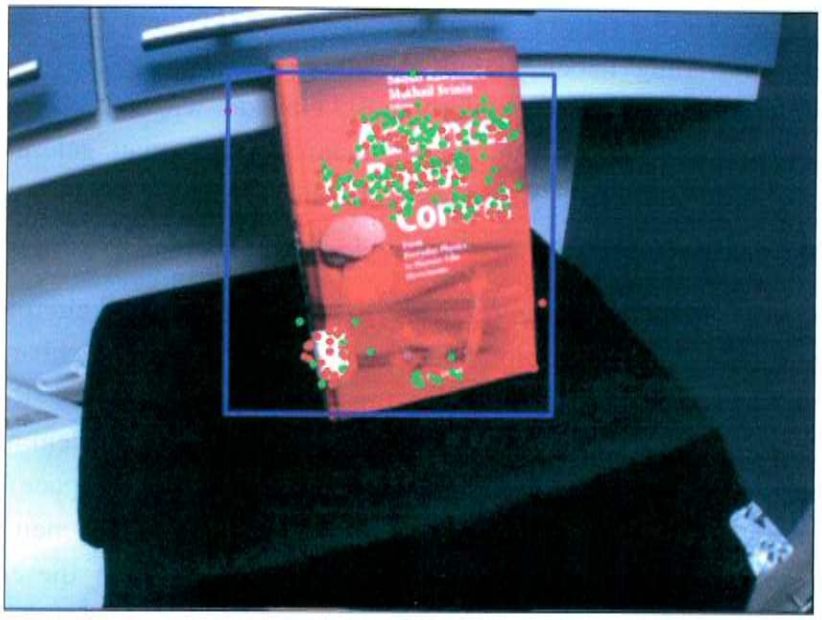

Abbildung 2-3: Erkennung eines Objektes mit Anzeige der Merkmale

Durch die große Variabilität der Skalierung, Drehung und der Lichtverhältnisse wird eine robuste Wiedererkennung in zufälligen Umgebungen, in denen die Objekte nicht speziell ausgerichtet werden gewährleistet. Es ist aber nötig, dass die Segmentierung von Vordergrund und Hintergrund möglich ist, da so Gegenstände vorsegmentiert werden, damit keine Merkmale im Hintergrund der Szene als Merkmale von Gegenständen erkannt werden.

3. Verwandte Arbeiten

Hier werden einige zentrale Arbeiten zu dem Thema Objekte lernen zusammengefasst und bewertet. Da viele verschiedene Aufgaben gelöst werden wollen, bedeutet der Begriff „Objekte lernen" für verschiedene Forscher weltweit oft etwas unterschiedliches. Je nachdem, aus welchem Bereich der Informatik sie kommen, beschäftigen sie sich mit unterschiedlichen Aspekten. Manche untersuchen das Lernen von visuellen Informationen über das Objekt, damit das Objekt in einem Bild oder Video erfolgreich detektiert und erkannt werden kann. Manche beschäftigen sich mit Spracherkennern, damit diese in der Lage sind, neue Wörter zu erkennen und in ihren Wortschatz aufzunehmen. Die dritte Gruppe möchte die Bedeutung von Objekten möglichst genau lernen, damit ein Umweltmodell entstehen kann, in dem die gefundenen Objekte auch Bedeutung tragen. Die hier vorgestellten Arbeiten behandeln alle einen oder zwei dieser Aspekte. Kein bisher laufendes System kann allerdings alle drei, zum vollständigen Objektlernen notwendigen Arbeiten, zusammen realisieren.

Im weiteren Verlauf werden erst in Kapitel 3.1 Arbeiten vorgestellt, die neue Wörter für die Spracherkennung lernen können. Dann werden in Kapitel 3.2 Systeme untersucht, die sich mit dem Lernen von visuellen Informationen von Objekten beschäftigen. Das Lernen der Bedeutung folgt in Kapitel 3.3.

3.1. Lernen von neuen Wörtern

Spracherkenner können nur Wörter und Sätze verstehen, die dem System vorher bekannt sind. Nur wenige Forschungseinrichtungen haben Methoden entwickelt, den Spracherkenner während der Laufzeit zu erweitern und neue Wörter mit Hilfe eines Dialogs aufzunehmen. Die beiden dem Autor bekannten Einrichtungen sind die Universität Karlsruhe (TH), deren Applikation in Kapitel 3.1.1 beschrieben wird und in dem hier verwendeten Spracherkenner integriert wurde und das Massachusetts Institute of Technology (MIT), deren Ansatz zum Lernen von Wörtern bei Kleinkindern in Kapitel 3.1.2 beschrieben wird.

3.1.1. Erkennen und Lernen neuer Wörter

Thomas Schaaf erforschte in (Schaaf, 2004) das Auftreten, Detektieren und Lernen von unbekannten Wörtern in Dialogen. Durch die neue Möglichkeit, durch Nachfragen des Systems während des Dialogs neue Wörter zu lernen und in den Spracherkenner zu integrieren, wurde es erst möglich, neue Objekte zu lernen. Die vorher starren Spracherkenner wurden so erweitert, dass während der Ausführung neue Wörter integriert werden können und sich das System somit an den Benutzer und die Umgebung anpassen kann.

Mit Hilfe von Head-Tail Modellen konnten unbekannte Wörter in einer Grammatik gefunden werden. Die Grammatik dient zum Verstehen der Äußerung und dort wird angegeben, wo unbekannte Wörter auftreten können. Im Kopfteil (Head) des Modells für ein Wort befindet sich ein exaktes phonetisches Modell, aber im Schwanzteil (Tail) wird nur noch ein weniger präzises akustisches Modell verwendet. Da die meisten bekannten Wörter bereits nach wenigen Phonemen eindeutig erkannt werden können, reicht es aus, unbekannte Wörter so lange genau zu beschreiben, bis klar ist, dass es ein unbekanntes Wort ist. Durch die genaue Modellierung am Anfang, wird der Suchraum stark eingeschränkt, während die ungenauere Modellierung des Restes immerhin noch die Länge und eine ungefähre Erkennung des Restes liefert und Rechenzeit spart. Schaaf teilt die unbekannten Worte im Vorfeld unüberwacht in Gruppen ein. Das so entstandene Sprachmodell hat mehrere verschiedene Klassen von unbekannten Wörtern, die durch nur zufällig vorkommende Vertreter ihrer Klasse trainiert wurden. Manche Firmennamen kommen beispielsweise häufig in dem Trainingstext vor, aber manche andere tauchen gar nicht auf. Die Firmennamen haben aber ähnliche Eigenschaften und werden durch das Clustern zusammengefasst. Durch die Ermittlung der Eigenschaften bei der Erkennung ist eine Einordnung des unbekannten Begriffs möglich. Damit lernt er bereits eine semantische Klasse des neuen Wortes.

Weiterhin wurde die Möglichkeit untersucht, den neu gelernten Wörtern eine sinnvolle Wahrscheinlichkeit für die Erkennung zuzuweisen. Damit wurde ein erster Prototyp entwickelt, bei dem es möglich ist, online neue Wörter zu erkennen und zu lernen. Während des Dialogs mit dem Benutzer kann die Klasse des neuen Wortes erkannt werden, ob es sich zum Beispiel um einen Ortsnamen oder einen Personennamen handelt. Im weiteren Verlauf des Gesprächs wird mit der Hilfe eines Phonemerkenners das Wort in seine Einzelphoneme aufgeteilt. Dieser Erkenner liefert eine Liste mit wahrscheinlichen Phonemfolgen für dieses Wort zurück. Ein Buchstabiererkenner versucht, die Schreibweise des Wortes zu lernen. Wenn all diese Punkte erfolgreich durchgeführt wurden, endet der Lerndialog. Diese Funktionalität wurde im weiteren Verlauf in den am Interactive System Labs (ISL) verbreiteten Spracherkenner Janus von (Finke, 1997) und (Soltau, 2001) integriert und kann seitdem verwendet werden. In der vorliegenden Arbeit wird er benutzt, so dass an jeder Stelle, an der ein Wort auftaucht, das der Spracherkenner noch nicht kennt, ein „Out of Vocabulary (OOV)" an den Dialogmanager weitergeleitet wird.

3.1.2. Grounded Situation Models

Am MIT (Massachusetts Institute of Technology) werden Systeme untersucht, die in der Lage sein sollen, mit menschlichen Benutzern zu interagieren. Dafür wurde die Idee eines Grounded Situation Models (GSM) entworfen, das für den Roboter die aktuelle Umgebung und Situation wiederspiegelt. Darin werden die physischen Objekte in der Umgebung, der Benutzer und der Roboter selbst modelliert. Außerdem wird der zeitliche Verlauf aller beteiligten Elemente aufgezeichnet. Aus den Wahrnehmungen des Roboters über visuelle und akustische Sensoren werden Beschreibungen für dieses Umweltmodell des Roboters erstellt. Aus den Zuständen des GSM können Aussagen über die Umwelt getroffen werden. Neben der visuellen Wahrnehmung kann der Roboter Anweisungen und Hinweise des Benutzers verarbeiten. In (Roy & Mavridis, Grounded Situation Models for Robots: Where words and percepts meet, 2006) wird dieses System genauer erläutert.

Eine mögliche Datenquelle für das Training von Robotern wird gerade in (Roy, Mavridis, Levit, & Gorniak, 2006) untersucht. Darin wird ein System vorgestellt, bei dem die ersten drei Lebensjahre eines menschlichen Kindes aufgenommen werden, um das Lernen von Sprache bei Menschen zu untersuchen. Dafür wurde das Haus einer Familie komplett mit Deckenkameras und Mikrophonen ausgestattet, die den Alltag und den ganzen Input des Kindes aufnehmen. Um besser zu verstehen, wie Kinder Sprache lernen, sollen verschiedene Lernstrategien des Roboters darauf getestet werden, ob sie realem Lernen von Menschen ähneln. Das Projekt läuft gerade und bisher sind hauptsächlich Methoden veröffentlicht, wie die Privatsphäre der Familie geregelt werden kann und wie die vollautomatische Erkennung auf der großen Menge an Daten möglich ist. Das System ist in der Lage, Objekte nach der Form und der Farbe zu

erkennen und dazu Bezeichnungen zu lernen, da angenommen wird, dass die Bezeichnung des Objektes gehäuft auftritt, wenn das Objekt im Blickfeld des Kindes ist. Dieses Lernen ist komplett unüberwacht und benötigt keine Vorsortierung der Objekte. Die semantischen Kategorien, die gelernt werden, wurden limitiert auf die Farbe und die Form der Objekte, die durch die visuellen Sensoren erkannt werden.

Das System verwendet ein Modul, das CELL (Cross-Channel Early Lexical Learning) genannt wird, um neue Wörter zu lernen. Es geht darum, herauszufinden, wie Kleinkinder lernen, welche Segmente der Sprache Bezeichnungen für Objekte sind, welche semantischen Kategorien bezeichnet werden und wie man sprachliche Einheiten mit semantischen Kategorien verbindet. Dafür wird Sprache automatisch segmentiert und Wörter werden mit visuellen Umrisskategorien in nicht transkribierten Sprach- und Videoaufnahmen verknüpft. (Roy, Learning words from sights and sounds: A computational model., 1999) beschreibt, wie CELL in der Lage ist, Vokabular über Formen und Farben zu lernen, die gerade im Bild auftauchen. Dafür werden Mütter und deren Kleinkinder aufgenommen, wenn diese spielen. Die Mütter haben eine Auswahl an Spielzeug, das sie ihrem Kind zeigen können und dabei benennen. So werden zum Beispiel „Rot“ und „Ball“ gesagt, wenn ein roter Ball gezeigt wird. Die Lernstrategie kombiniert die zeitgleiche sprachliche Äußerung mit einer Aufnahme des Umrisses in einem audiovisuellen Event in einem Kurzzeitgedächtnis. Zu beiden Teilen des audiovisuellen Events werden größtmögliche Abweichungen ermittelt, indem die wechselseitige Information der beiden Modalitäten gemessen wird. Dafür werden die Events mit Hilfe des Langzeitgedächtnisses untersucht, das eine Sammlung an Prototypen von audiovisuellen Events umfasst, die mit der Zeit weiter verfeinert werden. Mit Hilfe eines Events und der möglichen Abweichung für die sprachlichen und visuellen Merkmale werden Lexikoneinträge erstellt, die im Langzeitgedächtnis gespeichert werden. So wird mit der Zeit ein Wissen über Objekte und deren Namen aufgebaut.

3.2. Lernen von visuellen Daten

Bei genauerer Betrachtung stellt man fest, dass die meisten Wissenschaftler unter Objekten die visuellen Informationen verstehen und diese genau und trickreich lernen. Dabei wird meist kein Augenmerk darauf gelegt, was für ein Objekt tatsächlich vor dem Benutzer auf dem Tisch liegt. Die verwendete Sprache ist oft vorgegeben und die Bezeichner für Objekte werden reduziert auf „rotes Objekt" oder „Objekt eins". Um die vorliegende Arbeit zu vergleichen, wurden mehrere Systeme untersucht um Gemeinsamkeiten und Unterschiede zu erkennen.

Die hier ausführlich vorgestellten Arbeiten dienen dazu, einen Überblick über die aktuellen Systeme aufzuzeigen. Dabei werden die Arbeiten von Kirstein und Wersing in Kapitel 3.2.1 und die von Frank Lömker in Kapitel 3.2.2 beschrieben. Zum Lernen von visuellen Daten gibt es noch eine Reihe von anderen Arbeiten. Als weitere Arbeiten sind unter anderem (Arsenic, 2004), (Bekel, Bax, Heidemann, & Ritter, 2004) und (Garcia, Oliveira, Grupen, Wheeler, & Fagg, 2000) bekannt.

3.2.1. Lernen von visuellen Objekten

Bei Honda Research Institute werden zusammen mit der Universität Bielefeld intelligente Systeme entwickelt, die auch auf den berühmten Honda Humanoiden Asimo und P3 (Honda Robots, 2007) integriert werden. Dabei wird auch der Aspekt des Lernens von Objekten untersucht. Die Roboter sind durch ihre eigene Wahrnehmung und Erklärungen des Benutzers in der Lage, sich Wissen über Objekte anzueignen.

Das System nimmt mit Hilfe einer Stereokamera und Mikrophonen Objekte und einen Dialog mit dem Benutzer auf. Der Benutzer kann den Roboter mit unbekannten Objekten konfrontieren und dieser versucht, eine unifizierte Repräsentation zu lernen, die dieses Objekt beschreibt. In (Kirstein, Wersing, & Körner, 2005), (Wersing, et al., 2006) und (Wersing, et al., 2007) werden diese Ansätze näher erklärt. Dabei werden vor allem visuelle Informationen gelernt. Das Lernen geschieht, indem die Objekte vor der Kamera präsentiert werden. Der Hintergrund und das Szenario sind dabei nicht vorgegeben. Die Objekte werden aber mit vorher festgelegten Bezeichnungen wie „blue mug", „sun cream" oder „object one" bezeichnet. Diese tragen keine Bedeutung und haben somit keine Aussage für den Roboter. Ein uneingeschränktes Lernen ist also damit nicht möglich. Die Objekte werden mit hierarchischen Eigenschaftsdetektoren auf visuelle Eigenschaften hin untersucht. Das Lernen geschieht komplett online und benötigt keine weitere Überwachung nach der Initialisierung.

Für die Erkennung der Objekte werden Daten einer Stereokamera verwendet. Ein separater Erkenner vom Fokus des Benutzers segmentiert in den beiden Bildern eine Region der Aufmerksamkeit (ROI), in der das Objekt liegt. Dieser Teil aus den Originalbildern ist nur noch 144x144 Pixel groß und enthält das unbekannte Objekt.

Auf diesem Bildausschnitt werden mit Hilfe von Berechnungen über die Farbe und räumliche Tiefe Merkmale jedes Bildpunkts berechnet. Es werden adaptiven szenenabhängigen Filter (ASDF) verwendet, die eine Maske bilden, die das Objekt aus dem Bild ausschneiden. Zusätzlich wird durch Hautfarbenerkennung die Hand, welche auch im Vordergrund des Bildes auftaucht, von dem Objekt abgezogen. Bevor die Daten über das Bild von dem Objekt weiter verarbeitet werden, wird das Bild gedreht, damit die Hauptachsen des Objektes nicht schief im Bild liegen. Dadurch können die Ergebnisse verbessert werden. Diese Erkennung wird iterativ auf dem Video durchgeführt. Das Objekt wird dann mit 50 Umrissfiltern und drei Farbfiltern untersucht, um eine Repräsentation in diesen Eigenschafts-Maps zu erhalten. Die Filter sind jeweils 18x18 Pixel große Vergleichsmuster, die Eigenschaften aus kleinen Elementen des Objekts abbilden und die drei Kanäle des RBG Farbmanagements. Dadurch werden die Objekte genau klassifiziert und die erzielten Ergebnisse dieses Erkenners können sich mit anderen aktuellen Methoden messen (Wersing, et al., 2006). Diese Einordnung wird dann an das System übergeben, damit das Objekt erkannt oder gelernt werden kann. Einen Überblick über die Verarbeitung der visuellen Daten liefert Abbildung 3-1.

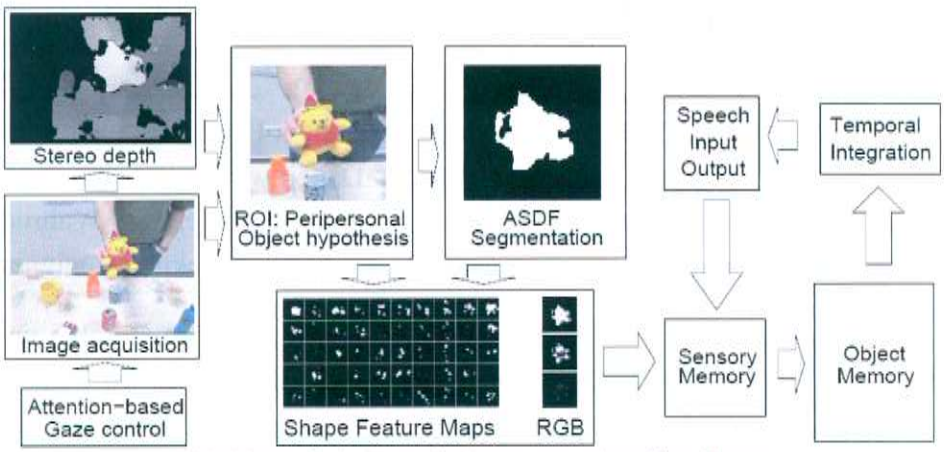

Abbildung 3-1: Verarbeitung von visuellen Daten

Insgesamt hat die bisherige Klassifikation einen sehr großen Raum, aber nur wenige Punkte, die Objekte repräsentieren. Trotz der hohen Dimension der gesamten Abbildung von (50+3)x18x18 = 17172 wird der Raum nicht reduziert, um Geschwindigkeit, Robustheit oder Generalisierungsfähigkeit zu erreichen. Die Abbildung des Inputs in einen spärlich besetzten Raum mit weit auseinanderliegenden Datenpunkten, welche die wichtigsten Strukturen des Objekts abbilden, führt dazu, dass die einzelnen Objekte trotzdem leicht gefunden und gespeichert werden können.

Das System erstellt eine Hypothese, um welches Objekt es sich handelt, indem es dieses Objekt mit den bisher bekannten Objekten vergleicht. Diese Annahme bleibt in einem temporären Modell erhalten, bis der Benutzer eine Aussage dazu abgibt. Der Benutzer

kann diese Vermutungen annehmen oder ablehnen oder durch Benennung des Objektes neue Objekte einlernen. Wenn eine Aussage des Benutzers über das Objekt vorliegt, wird diese Information mit den aktuellen visuellen Daten des Gegenstandes zusammen gespeichert. Bis dahin werden mehrere Ansichten des Objektes analysiert, damit mehrere Datenpunkte zu einem Objekt gelernt werden und die Erkennung robuster wird. Das Lernen geschieht dabei direkt und das neue Objektmodell wird sofort in die Objektdatenbank aufgenommen und ist nun bekannt. Beim Erkennen werden durch einen Nächster-Nachbar-Algorithmus Hypothesen generiert, um welches Objekt es sich handeln könnte.

Neben dem Lernen von Objekten kann dieses System auch zum Lernen von Gesichtern verwendet werden.

3.2.2. Lernen von Objektbenennungen mit visuellen Prozessen

Ein umfangreicher Dialog zum Lernen von Objekten wird in (Lömker, 2004) vorgestellt. Dabei sollen in einer kontrollierten Umgebung Objekte gefunden und neu gelernt werden. Das System verwendet Spracherkennung, um den Namen des Objektes aufzulösen, Gestenerkennung, damit Informationen aus Zeigegesten des Benutzers weiter verarbeitet werden und einen Objekterkenner, der Hypothesen über gefundene Objekte liefert. Das System ist in der Lage, Auskünfte über seine Umwelt zu geben, das Umweltmodell aus Benutzereingaben aufzubauen und visuell neue Objekte zu lernen. Dabei werden aber keine semantischen Informationen über die Objekte berücksichtigt und der Spracherkenner kann nur bereits bekannte Namen und Bezeichner verstehen. Die Umgebung des Roboters und die darin enthaltenen Objekte werden in einem Objektmodell beschrieben. Vorlage ist das menschliche Bewusstsein, in dem wir Modelle von den physikalischen Objekten unserer Umwelt speichern.

Die Arbeit hat das Ziel, ein interaktives, multimodales und visuell basiertes Lernen unbekannter Objekte zu ermöglichen. Der Schwerpunkt ist auf dem Lernen von Ansichten eines Objektes gelegt, die es ermöglichen, dass dieses Objekt zu einem späteren Zeitpunkt wieder gefunden werden kann. Ein spezieller Hintergrund und eine spezielle Domäne für die Objekte sind nicht nötig. Das System beobachtet die Szene mit mehreren Farbkameras, und nimmt die Sprache des Benutzers auf. Zeigegesten und Greifaktionen des Benutzers werden interpretiert, um das gerade im Fokus stehende Objekt besser im Bild zu finden. Abbildung 3-2 zeigt den Aufbau des Gesamtsystems. Das gesamte System ist sprachgesteuert.

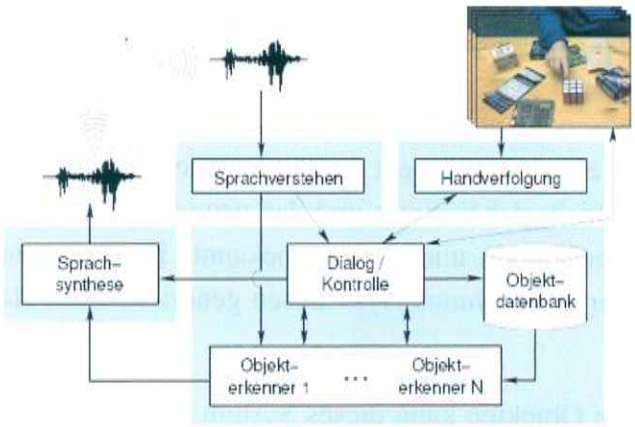

Abbildung 3-2: Architektur des Systems zum Objekte lernen

Damit das System ein Objekt neu lernen kann, oder eine weitere Ansicht eines bekannten Objektes lernen kann, wird der Benutzer aufgefordert, das gerade beschriebene Objekt aus der Szene zu entfernen. Durch den Vergleich der Aufnahmen mit und ohne Bild, kann das gewünschte Objekt detailliert segmentiert werden, da sich der Rest der Szene nicht ändert. Dadurch können auch unbekannte Hintergründe im Bild auftauchen, da sich dieser ebenfalls nicht ändert. Wenn sich doch ein Teil des Restbildes ändert, kann die Greifaktion zusätzlich Auskunft darüber geben, wo genau das Objekt im Bild war. Das System kann entweder mit Hilfe von erkannten Handbewegungen gesteuert werden oder der Benutzer kann direkt Objekte referenzieren. Es ist nötig und beabsichtigt, dass der Dialog bei falschen Annahmen von dem Benutzer korrigiert wird. Abbildung 3-3 zeigt den Dialogaufbau.

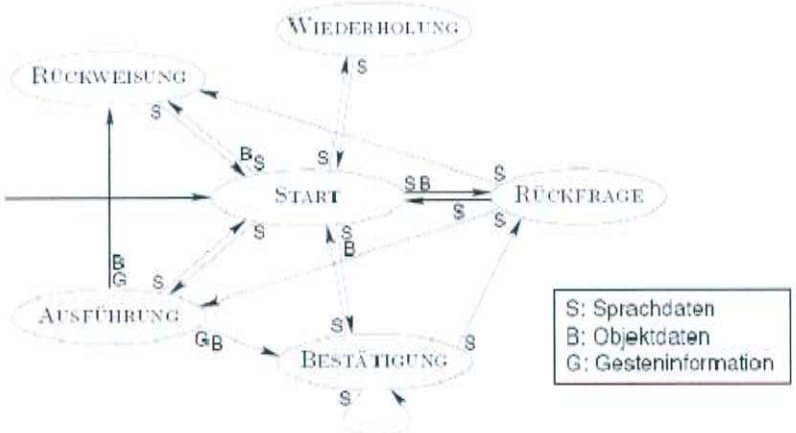

Abbildung 3-3: Dialogablauf

Das System ist in der Lage, Objekte zu lernen und zu lokalisieren, wenn es dazu aufgefordert wird. So kann zum Beispiel „Gib mir den Würfel“ von dem Benutzer gefordert werden. Dabei können nur dem Spracherkenner bekannte Bezeichner verwendet werden, das System kann keine neuen Bezeichner lernen. Für den Spracherkenner wurden Angaben zur Größe und zur Farbe von Objekten definiert, die verstanden werden können und auf semantische Informationen über das Objekt geparst werden können. Es gibt Gesten, mit denen man auf Objekte zeigen kann und solche, die das Verändern der Objektposition beschreiben. Dafür werden die Handtrajektorien ausgewertet, wobei für Zeigegesten zusätzlich eine sprachliche Information benötigt wird, mit der der Benutzer das Objekt referenziert. Durch den Dialog steuert das System, indem es den Benutzer zum Beispiel dazu auffordert, das Objekt aus der Szene zu entfernen.

Ein Objekt wird rein durch die visuellen Informationen repräsentiert. Zudem werden zwei semantische Klassifikationen verwendet, welche die Farbe und Größe des Objektes angeben. Diese Adjektive werden auf die Wahrscheinlichkeiten der verschiedenen Objekthypothesen angewandt. Dabei wird die Angabe der Größe auf alle Hypothesen ausgewertet und Farbinformationen in Abhängigkeit zu dem jeweils verwendeten Objekterkenner. Es werden drei Objekterkenner verwendet, die verschiedene Merkmale in dem Bild untersuchen und später kombiniert werden. Die Erkennung erfolgt rotations- und translationsinvariant. Der gesamte Dialog stellt einen Ansatz zum Lernen von Objekten dar, es ist aber nicht möglich neue Objektbezeichnungen zu lernen oder semantische Informationen, die nicht Farbe und Größe betreffen anzugeben.

3.3. Lernen von Bedeutung von Objekten

Die Bedeutung von Objekten zu kennen ist notwendig, damit der Roboter mit den Gegenständen interagieren kann. Das Lernen von semantischen Informationen über neue Objekte stellt einen Schwerpunkt dieser Diplomarbeit dar. Hierzu gibt es eine Reihe an Arbeiten, die sich bereits damit beschäftigt haben.

In Kapitel 3.3.1, bei der Arbeit von Jaime Carbonell, wird ein System vorgestellt, das mit Hilfe der Struktur des Satzes, Informationen über Grammatik und politischen und militärischen Fakten, Informationen über unbekannte Wörter ableitet.

In Kapitel 3.3.2 wird das System von Dusan und Flanagan vorgestellt, das Eigenschaften, von semantischen Konzepten lernt. Neue Farben und Formen können angegeben und deren Bedeutung gelernt werden.

Die Arbeit von Boris Schulz aus Kapitel 3.3.3 befasst sich mit Objektklassen, die in einer Ontologie gespeichert werden und in einem Dialog neu gelernt werden können.

Die Arbeit von Regine Becher, die in Kapitel 3.3.4 vorgestellt wird, befasst sich mit dem Modellieren und Lernen von Objekten und deren Eigenschaften, um ein detailliertes und umfangreiches Hintergrundwissen aufzubauen.

3.3.1. FOUL-UP und POLITICS

Jamie Carbonell untersuchte bereits in den Anfängen der automatischen Spracherkennung Methoden, wie unbekannte Begriffe automatisch mit ihrer Bedeutung verknüpft werden können. In (Carbonell, 1979) werden zwei Ansätze vorgestellt, die es erlauben, ein unbekanntes Wort zu interpretieren und eine mögliche semantische Bedeutung zu ermitteln.

Der erste Ansatz, das so genannte FOUL-UP verwendet eine Wissensbasis mit Aussagen über eine Domäne, um Rückschlüsse auf das unbekannte Wort zu ziehen. Dabei geht Carbonell davon aus, dass ein Wort in einer bestimmten Domäne in ähnlichen Satzgefügen und in ähnlichen Zusammenhängen mit anderen Begriffen auftaucht. Dieses Vorgehen benötigt zum einen eine große Basis an Hintergrundwissen über eine bestimmte Domäne, ist damit nur innerhalb dieser Domäne anwendbar und führt zum Teil zu falschen Schlussfolgerungen. In abgeschlossenen, relativ kleinen Domänen, wie zum Beispiel in einem Restaurant, gibt es nur wenige mögliche Äußerungen, die sich auf Bestellungen, das Essen und das Kassieren beschränken. Wenn in einem Satz wie „Ich möchte bitte das Wiener Schnitzel mit Salat" der Begriff des „Wiener Schnitzels" bisher unbekannt ist, kann mit Hilfe von anderen Beispielen von Bestellungen geschlossen werden, dass es sich um ein Gericht handeln muss. Dieser Ansatz ist in größeren Domänen aber nicht mehr verfolgbar und führt zu unvollständigen oder falschen Hypothesen.

Der zweite Ansatz, das POLITICS untersucht unbekannte Begriffe erst mit Hilfe einer syntaktischen Analyse, um herauszufinden, um welche grammatikalische Art von Wort es sich handelt. Weiterhin wird der ganze Satz, in dem das Wort auftauchte untersucht, um herauszufinden, welche semantische Bedeutung das verwendete Verb, auftretende Substantive und Adjektive haben. Durch die Relation mit den bekannten Wörtern wird eine Schlussfolgerung gezogen, um was es sich bei dem Wort handeln könnte. In dem Beispiel „Russia sent massive arms shipments to the MPLA in Angola.", in dem unklar ist, was „MPLA" ist, werden durch die Analyse, dass es sich um ein Dativobjekt handelt und dass es darum geht, dass etwas versendet wird, die Schlüsse gezogen, dass es sich um Personen oder ein Ort in Angola handeln muss. In einem zweiten Schritt werden dann Informationen aus dem Hintergrundwissen über politische Gegebenheiten verwendet, in denen gespeichert ist, dass in Angola Bürgerkrieg herrscht und welche Ziele Russland gerade verfolgt. Diese werden interpretiert und damit geschlossen, dass die MPLA eine Befreiungsbewegung in Angola sein muss.

Beide Ansätze sind zwar in der Lage, in bestimmten Konstellationen selbstständig semantische Informationen zu gewinnen, aber dazu ist es nötig, dass die verwendete Domäne genau beschrieben wurde und die Art, wie ein unbekanntes Wort verwendet wurde Schlüsse zulässt. In einem Satz wie „This is a red cup" kann man fast keine Information aus dem Satzbau und der Bedeutung der Verben und Substantive lesen. In einer alltäglichen Situation führt das System deshalb oft zu einer fehlerhaften Hypothese oder Abbruch aufgrund von fehlenden Informationen über das unbekannte Wort. Deshalb ist dieser Ansatz für das hier vorliegende Dialogsystem nicht sinnvoll.

3.3.2. Multimodale Sprachakquisition

Dusan und Flanagan beschäftigen sich in (Dusan & Flanagan, Adaptive Dialog Based upon Multimodal Language Acquisition, 2002) und (Dusan & Flanagan, An Adaptive Dialogue System Using Multimodal Language Acquisition, 2002) mit dem Lernen von neuen Begrifflichkeiten in einem natürlichsprachlichen Dialog. Dabei wird eine initiale Grammatik um neue Wörter erweitert. Das Hintergrundwissen besteht aus einer Grammatik und semantischen Informationen. In der Grammatik werden semantische Konzepte wie die Farbe, Form oder Funktionalität verwendet. Für jedes semantische Konzept gibt es mehrere Ausprägungen, wie zum Beispiel blau, rot oder grün. Ein Objekt wird entweder von dem Entwickler oder dem Benutzer in eine dieser Unterklassen einsortiert. Die Konzepte können um neue Ausprägungen erweitert werden, wenn diese in dem Dialog auftreten. Wenn der Benutzer zum Beispiel eine neue Farbe verwendet, wird er nach der semantischen Klasse gefragt und er kann in einer Farbpalette diese Farbe direkt spezifizieren. Jedes neue Objekt kann so in eine semantische Klasse einsortiert werden. Mögliche Klassen können Früchte, geometrische Formen oder Zeichen sein. Theoretisch kann aber alles, was graphisch darstellbar ist auf

diese Weise gelernt werden. Wichtig ist, dass neben neuen, semantischen Klassen auch Synonyme gelernt werden können.

Insgesamt befassen sich Dusan und Flanagan nicht nur mit der Akquisition von neuen Wörtern, sondern mit einer umfassenden multimodalen Dialogstruktur, die Objekte in ihrer Umgebung wahrnehmen kann, den Benutzer in einer natürlichen Art und Weise versteht und auf Aufträge und Fragen des Benutzers reagieren kann. Das Gesamtsystem wird in Abbildung 3-4 dargestellt.

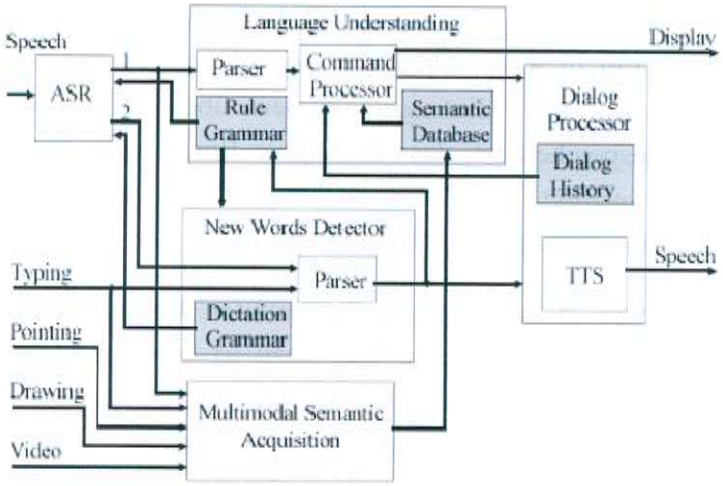

Abbildung 3-4: Blockdiagram des Gesamtsystems

Das System erlaubt das Anlegen neuer semantischer Ausprägungen, wenn deren Konzept im Vorhinein angelegt wurde. Die semantischen Informationen beschränken sich aber auf visuelle Eigenschaften der Objekte und bilden keine real existierenden Objekte ab. Eine Ontologie, welche die Art der Objekte näher beschreibt wird nicht verwendet.

In (Dusan & Flanagan, A System for Multimodal Dialogue and Language Acquisition, 2003) wird ein System vorgestellt, das in der Lage ist, Objekte zu manipulieren und zu lernen. Dafür wurde ein multimodales System entworfen, das Sprache, Zeigegesten, Zeichnungen, Text, Mausklicks und aufgenommene Bilder verarbeitet, um eine Benutzerintention zu erkennen und auszuführen. Der realisierte Spracherkenner ist in der Lage, Kommandos und ein vordefiniertes Vokabular zu erkennen. Zudem können Konzepte von Objekten, wie die Farbe oder Form, um Ausprägungen erweitert werden. In Abbildung 3-5 ist eine Situation in einer Simulation dargestellt, in der mehrere Objekte, wie Autos und Hubschrauber auf einer militärischen Karte angeordnet sind.

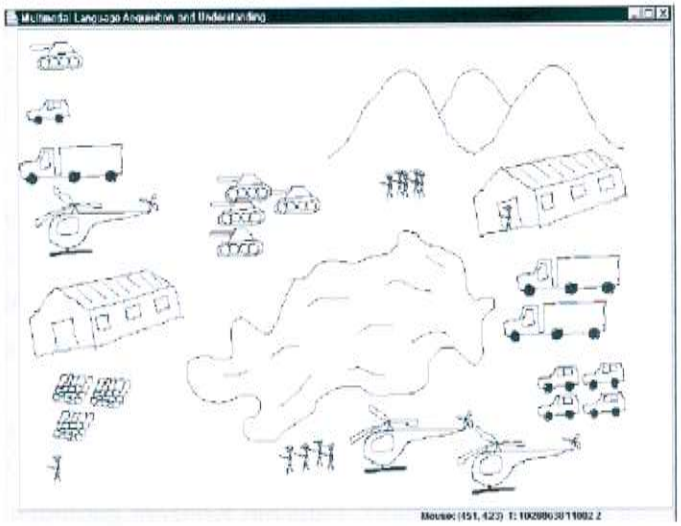

Abbildung 3-5: Beispiel aus einer Simulation mit militärischen Karten

3.3.3. Lernen neuer Wörter im Dialog

In (Schulz, 2005) wird das Auftreten von unbekannten Wörtern in natürlichsprachlichen Mensch-Maschine Dialogen untersucht. Das Hauptaugenmerk sind Veränderungen am Spracherkenner und den verwendeten Modellen für die Grammatik und die Ontologie. Dabei werden die Ergebnisse aus der Arbeit von Thomas Schaaf, die in Kapitel 3.1.1 vorgestellt wurde, verwendet und so erweitert, dass online neue Wörter in den Wortschatz des Spracherkenners aufgenommen werden können. Schwerpunkte der Arbeit sind die Anpassungen der Grammatik und der Repräsentationen der Objekte in dieser Grammatik, damit der verwendete Dialog, der ein unbekanntes Objekt lernt, möglichst kurz gehalten werden kann.

Der Spracherkenner Janus wird an das Auftreten von OOVs angepasst. In das akustische Modell werden unbekannte Wörter (Out Of Vocabulary OOV) eingefügt und trainiert. Um zu gewährleisten, dass OOVs nur erkannt werden, wenn tatsächlich kein anderes Wort passt, werden Strafterme verwendet, welche die Wahrscheinlichkeit eines OOVs vermindern.

In der Ontologie werden Konzepte und Unterkonzepte modelliert, die alle bisher bekannten Objekte abdecken. Ein Konzept beschreibt die Eigenschaften eines Objektes, indem es zum Beispiel alle Objekte umfasst, die essbar sind. Dieses Konzept kann weiter unterteilt werden, in kalte und warme Lebensmittel. So wird in der Ontologie immer genauer beschrieben, um was für ein Objekt es sich handelt. Es werden Untersuchungen durchgeführt, wie man die Domäne von Objekten in der Küche sinnvoll unterteilen kann.

In der Grammatik muss das Auftreten von unbekannten Wörtern ebenfalls modelliert werden. In den Unterkonzepten werden OOVs hinzugefügt, die es ermöglichen, aus der Position, in der das neue Wort auftritt direkt die semantische Information auszulesen. Zudem können OOVs als Oberbegriffe erscheinen.

Ein Dialog zur Klärung der genauen Schreibweise eines OOVs wird vorgestellt. Dieser Dialog ist in Abbildung 3-6 dargestellt. Er beginnt mit der Detektion eines OOVs. Falls es fälschlicherweise detektiert wurde, geht der Dialog in den Fehlerzustand (F) über und wird nicht weiter verfolgt. Falls das OOV korrekt erkannt wurde (B), beginnt der OOV Dialog mit der Information, dass nun geklärt wird, was gesagt wurde. Dabei werden Informationen, die bereits aus der Grammatik abgeleitet wurden, bestätigt. So lange noch nicht genügend Informationen gesammelt wurden (U) fragt das System weiter nach, welchem Konzept das Objekt angehört. Falls ein Konzept gefunden wurde, dieses aber Unterkonzepte enthält (G1) werden diese vorgelesen und die Einordnung wird weiter verbessert, bis es keine weitere Verzweigung mehr gibt (G2). Nun wird der Benutzer aufgefordert, den Namen in ein Textfeld einzugeben und der Dialog endet erfolgreich. Um den Namen in der aktuellen Stelle in die Konzepthierarchie einzutragen, kann der Benutzer jederzeit in den Zustand N übergehen und den Dialog dann beenden (Endzustand X). Der Dialog kann jederzeit abgebrochen werden (abort).

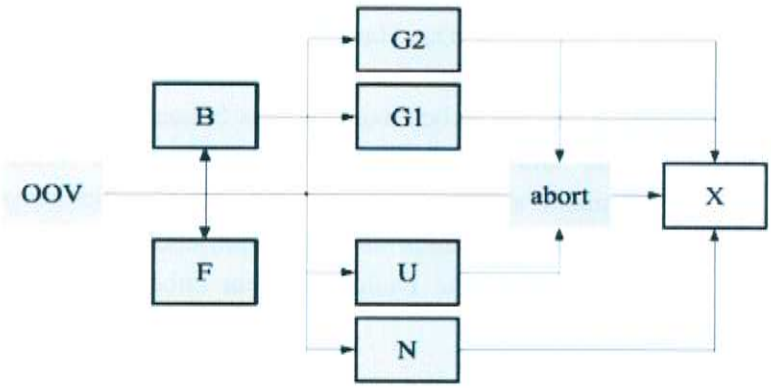

Abbildung 3-6: Ablaufdiagramm des OOV Dialogs

Durch die Vielzahl an Änderungen werden der natürliche Dialog mit einem System und die Erweiterung des Vokabulars mit Angabe der semantischen Bedeutung des Wortes ermöglicht. Das Wort wird in die Ontologie eingetragen und an den notwenigen Stellen im Spracherkenner gespeichert. Die semantischen Klassen, in welche die Objekte einsortiert werden können, sind direkt an die Ontologie geknüpft und spezifizieren die Objektklasse weiter. Ein Objekt wird dabei in eine eindeutige Objektklasse eingeordnet.

3.3.4. Interaktive Objektmodellierung

Regine Becher hat in (Becher, Boesnach, Steinhaus, & Dillmann, 2006) und (Becher, Steinhaus, Zöllner, & Dillmann, 2006) ein Modell für Objekte vorgestellt, das in der Lage ist das Hintergrundwissen des Roboters detailliert abzubilden. Das System kann neue Objekte lernen, indem es von einem Benutzer und von einer Vielzahl an Sensoren Informationen über den Gegenstand sammelt und in ein Objektmodell füllt.

Das verwendete Objektmodell bildet für jedes Objektkonzept ab, welche Eigenschaften und welche Attribute dieses haben kann. Dabei fassen Eigenschaften mehrere Attribute zusammen. Jedes Attribut und jede Eigenschaft kann Standardwerte enthalten und wird beim Lernen von Objekten genauer bestimmt. Eigenschaften sind zum Beispiel „ist füllbar" oder „ist transportierbar", wobei diese verschiedene Attribute vereinen, die zum Beispiel den Füllstand, die maximale Transportbeschleunigung oder den Neigungswinkel, der abhängig vom Füllstand sein kann, beschreiben. Ein Attribut kann für mehrere Eigenschaften verwendet werden. Die Attribute sind detaillierter als die möglichen Eigenschaften und enthalten auch Informationen über die Farbe und Form des Objektes. Eigenschaften modellieren ebenfalls Aktionen, die angeben, was man mit einem Objekt machen kann. So wird auch modelliert, dass ein Gegenstand gefüllt werden kann. Die Aktionen sind direkt an Attribute geknüpft, um deren Wert zu ändern, wenn eine Aktion ausgeführt wird. So steigt zum Beispiel der Füllstand, wenn man das Objekt füllt. Abbildung 3-7 zeigt das Objektmodell. Objektklassen werden hierarchisch verwaltet und erben die möglichen Attribute und Eigenschaften von deren Elternklassen.

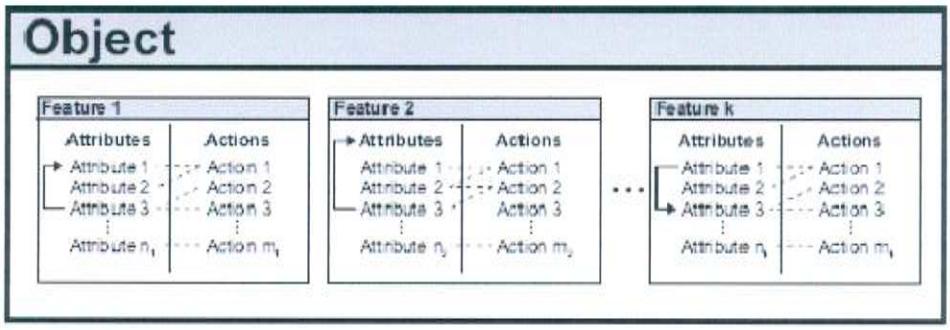

Abbildung 3-7: Struktur des Objektmodells

Mit Hilfe eines Touchscreens und sprachlichen Kommandos kann der Benutzer ein Objekt beschreiben. Zudem kann er das Objekt bewegen und drehen. Alles wird mit Hilfe eines 3D Laser Scanners, einer hochauflösenden Farbkamera, einem Drehteller und einem Datenhandschuh untersucht. Die Sensoren versuchen automatisch die Attribute zu füllen. Da die Attribute vorher festgelegt wurden, können Methoden entwickelt werden, die Werte dieser Attribute zu ermitteln. In (Becher, Kasper, Steinhaus, & Dillmann, 2007) werden zwei Beispiele zum automatischen Ermitteln gegeben. Darin wird beschrieben, wie das System stabile Positionen ermittelt, also wie Gegenstände zum Liegen kommen und wie das System Einschränkungen der Bewegung nachprüft. Dabei werden maximale Geschwindigkeit, Beschleunigung und Rotation untersucht, damit das Objekt während der Verwendung nicht kaputt geht. Der Benutzer kontrolliert diese Hypothesen und kann sie ändern oder korrigieren. Beim Lernen eines neuen Objektes, gibt der Benutzer an, welche Eigenschaften und Attribute das Objekt haben kann und in welcher Hierarchieebene das Objekt einsortiert werden kann. Dafür wählt er aus der Menge aller möglichen Eigenschaften die passenden aus. Dann versucht das System, die notwendigen Informationen mit Hilfe der Sensoren zu ermitteln.

Das beschriebene System zeigt ausführliche Möglichkeiten, wie Objekte sehr genau beschrieben werden können. Dabei setzt es auf einen umfangreichen Einsatz von verschiedenen Sensoren. Die Interaktion mit dem Benutzer ist nötig, um Fehler zu vermeiden. Das System kann keinen Dialog führen, sondern dient zum Erstellen des Hintergrundwissens eines Roboters.

3.4. Zusammenfassung

Ein wichtiger Punkt ist, dass das gesamte System zum Lernen von Objekten online die neuen Objekte lernt, so dass diese sofort für die Weiterverarbeitung vorhanden sind. Der Dialog, in dem der Roboter das Objekt einlernt, sollte nicht zu lang sein und der Benutzer sollte nicht in seiner Wortwahl eingeschränkt werden. Trotzdem soll das System robust die Informationen sammeln und auch bei fehlerhaften oder lückenhaften Informationen weiter erfolgreich herausfinden, was für ein Gegenstand vor ihm liegt. Dafür ist es notwendig, dass alle drei Teilbereiche des Lernens von Objekten einzeln gut arbeiten und erfolgreich zu einem Gesamtsystem integriert werden. Tabelle 3-1 zeigt einen Überblick über die realisierten Teilaspekte vom Objektlernen in den verschiedenen vorgestellten Systemen.

Autor (Kapitel)	Neue Namen lernen	Visuelle Merkmale lernen	Semantik lernen
Thomas Schaaf (3.1.1)	Ja	Nein	Nein
Deb Roy (3.1.2)	Ja	Ja	Nein
Wersing und Kirstein (3.2.1)	Nein	Ja	Nein
Frank Lömker (3.2.2)	Nein	Ja	Nein
Jamie Carbonell (3.3.1)	Nein	Nein	Ja
Dusan und Flanagan (3.3.2)	Nein	Nein	Ja
Boris Schulz (3.3.3)	Nein	Nein	Ja
Regine Becher (3.3.4)	Nein	Ja	Ja

Tabelle 3-1: Überblick über Funktionen in verwandten Arbeiten

Alle Arbeiten zusammen bilden eine Basis, die man verwenden kann, um reale Objekte zu lernen. Keine der vorliegenden Arbeiten befähigt einen Roboter aber dazu, dass er tatsächlich sieht und versteht, was für ein Objekt vor ihm steht, so dass er es dann auch für die Erfüllung der Aufgaben, die ihm ein Benutzer stellt, verwenden kann. Dafür müssen verschiedene Arbeiten kombiniert und angepasst werden. Das Lernen von Objekten und deren Bedeutung umfasst alle hier vorgestellten Aspekte, die zusammen in einem integrierten System kombiniert werden müssen. Damit ein Dialog mit dem Roboter geführt werden kann, wird untersucht, welche Anforderungen von dem Objektmodell und der Ontologie erfüllt werden müssen.

Beim Lernen der Bedeutung ist es notwendig, auf die Sichtweisen von Benutzern einzugehen und möglichst einfach für den Benutzer lernt, was für eine Art von Objekt das ist. Der Lernalgorithmus muss mit dem verwendeten Objektmodell zusammen entwickelt werden. Um einen Ansatz zu finden, wurden bisherige Lernalgorithmen überprüft, die Bedeutungen von Objekten untersuchen. Der Ansatz von Jamie Carbonell aus Kapitel 3.3.1 ist nicht anwendbar, da in Sätzen wie „Gib mir die rote Tasse" nicht genügend Information trägt, um Informationen über darin verwendete Wörter zu sammeln. In Kapitel 3.3.2 wird das System von Dusan und Flanagan vorgestellt, das zwar semantische Eigenschaften, wie Farben und Formen lernt, aber dabei ist es nicht möglich, dass neue Wörter gelernt werden oder weitere Informationen über die Art von Objekten beschrieben werden. Die Einteilung in semantische Klassen bedeutet, dass semantische Attribute, wie Farbe und Form, untersucht werden, aber es wird keine Einteilung in Objektklassen vorgenommen. Die Ansätze von Boris Schulz aus Kapitel 3.3.3 stellen einen Ausgangspunkt dar, der in der vorliegenden Arbeit weiter verfeinert wurde. Eigenschaften und Klassen werden bei Schulz zusammen in einer Ontologie behandelt und zum Lernen neuer Objekte werden neue Unterklassen in der Ontologie gelernt. Die Arbeit von Regine Becher, die in Kapitel 3.3.4 vorgestellt wird, stellt kein Dialogsystem dar, sondern befasst sich mit dem Modellieren und Lernen von Objekten und deren Eigenschaften, um ein detailliertes und umfangreiches Hintergrundwissen aufzubauen. Dabei werden viele Sensoren eingesetzt, die automatisch Attribute der Objekte erfassen. Interaktionen mit dem Benutzer werden verwendet, um Hypothesen zu bestätigen oder zu korrigieren. Das System verwendet nicht nur sprachliche Dialoge, sondern auch andere Modalitäten, wie Tastatur und Maus.

4. Lernen von Objekten durch Dialoge

In Kapitel 1.1 wurde beschrieben, was ein System, das Objekte mit Hilfe eines Dialogs lernen soll, können muss. In diesem Kapitel wird genauer beschrieben, wie diese Funktionalitäten im Einzelnen realisiert wurden und wie das Gesamtsystem aufgebaut und in den mobilen Roboter ARMAR-III integriert ist.

Erst wird in Kapitel 4.1 das Problem des Objekte Lernens untersucht, um darzustellen, was das System können muss. Darin werden die Anforderungen erklärt, die das Dialogsystem erfüllen muss, um erfolgreich im Dialog Objekte zu lernen.

Dann wird in Kapitel 4.2 erklärt, wie der Aufbau der Objektbeschreibungen, die der Roboter für die Objekte lernt, entstanden ist. Dafür wurde ein Vortest durchgeführt, um den Aufbau des Objektmodells zu evaluieren.

In Kapitel 4.3 wird das Szenario, in dem der vorgestellte Dialog ablaufen soll vorgestellt. Dabei werden auch der verwendete Roboter ARMAR-III, die Umgebung des Roboters und die Rolle des Benutzers erklärt.

Kapitel 4.4 beschreibt den Aufbau des Gesamtsystems, das durch ein Dialogsystem realisiert wird, das um eine Objekterkennung und eine Objektmodellierung erweitert wurde.

In Kapitel 4.5 wird beschrieben, wie der Dialog konzipiert und realisiert wurde und wie die Modellierung der Objekte und die Ontologie erstellt und umgesetzt wurde. Der genaue Dialogablauf wird erläutert und das genaue Vorgehen des Roboters wird geschildert.

4.1. Problemanalyse

In dieser Arbeit wird ein System entworfen, das in der Lage ist, visuelle, akustische und semantische Informationen zu sammeln und zu kombinieren. Ein Roboter soll in einem Dialog mit verschiedenen Personen seine Umwelt untersuchen und kennenlernen. Für die Benutzer steht im Vordergrund, dass der Roboter ihnen in alltäglichen Situationen hilft und Aufgaben erledigen kann. Der Mensch kann in einem natürlichsprachlichen Dialog den Roboter dazu auffordern, verschiedene Arbeiten zu erledigen. Dabei kann er beliebige Bezeichnungen für Objekte verwenden und das System lernt nach und nach mehrere mögliche Arten, dieses spezielle Objekt zu beschreiben.

Damit ein Mensch möglichst frei mit dem System interagieren kann, ist es nötig, dass dieses jederzeit auf verschiedene Äußerungen des Benutzers reagieren kann. In einem natürlichsprachlichen Dialogsystem heißt das, dass die Grammatik des Spracherkenners eine große Vielfalt an Sätzen verarbeiten können muss. Dadurch wird es schwer, dass die Erkennung immer korrekt die aufgenommene Sprache in eine Hypothese dessen, was gesagt wurde, umwandelt. Durch eine große Grammatik entstehen viele Möglichkeiten von ähnlichen Sätzen, aber das System soll trotzdem robust erkennen, was der Benutzer will.

Die Objekte sollen von der Objekterkennung identifiziert werden, wenn diese bereits bekannt sind. Anderenfalls soll der Objekterkenner ein unbekanntes Objekt anzeigen, damit dieses gelernt werden kann. Die Art und Weise, wie Objekte referenziert werden, soll variabel bleiben, da Benutzer nicht unbedingt wissen, unter welchem Namen der Roboter ein Objekt gelernt hat. Das Lernen von Objekten soll online geschehen, ohne dass ein Administrator eingreifen muss. Auftretende Fehler von der Objekterkennung, dem Spracherkenner oder beim Lernen von Objekten sollen im Dialog aufgelöst und behandelt werden können.

Erfolgreich Objekte Lernen bedeutet, dass ein Roboter ein real existierendes Objekt, das er sehen kann, erfolgreich in ein Objektmodell eintragen kann, so dass er später in der Lage ist, dieses Objekt mit Hilfe einer Beschreibung durch einen menschlichen Benutzer zu finden und zu verwenden. Das Lernen umfasst Speichern von visuellen Informationen, eine Objektklassifikation mit Hilfe einer Ontologie und eine Beschreibung des Objektes.

4.2. Beschreibungen und Klassen von realen Objekten

Der Roboter verwendet eine Objektklasse und eine Objektbezeichnung, um Gegenstände zu beschreiben. Die Idee, Objekte auf diese Art zu beschreiben, wurde aus den Ergebnissen eines Vortests abgeleitet. Dabei wurden Benutzer dazu aufgefordert, über reale Objekte zu sprechen. Bei diesem Test wurden 10 verschiedene Testpersonen zu jeweils 18 Objekten befragt um zu ermitteln, wie das Objektmodell für die Verwendung in einem Dialog aufgebaut werden muss. Die Auswahl an Objekten ist in Abbildung 4-1 zu sehen. Dabei viel auf, dass die Benutzer häufig Eigenschaften und Objekttypen verwendeten, um bestimmte Objekte in dieser Auswahl zu identifizieren. In dem Befehl „Gib mir die blaue Tasse" verwendet man zum einen die Art oder Klasse „Tasse" des Objektes, um es gegen ein blaues Buch abzugrenzen, das auf dem Tisch liegt und zum anderen die Eigenschaft „blau", die als Bezeichnung angesehen werden kann, um klarzustellen, dass man eine blaue und keine schwarze Tasse bekommt. Diese Eigenschaft stellt eine semantische Kategorie dar, welche die Objekte in Farben unterteilt.

Abbildung 4-1: Objekte des Vortests

Jeder Benutzer wurde vor dem Test mit einer kurzen Einführung auf seine Aufgabe vorbereitet. Dabei wurde ihm/ihr erklärt, dass es darum geht, dass ein Roboter neue Objekte lernen möchte. Dafür stellt dieser ein paar Fragen, um herauszufinden, was für ein Objekt er vor sich hat. Alle Objekte wurden auf einem Tisch aufgebaut und das Experiment durchgeführt. Dabei hatten die Probanden von Anfang an alle Objekte auf einmal im Blick und wurden einzeln dazu befragt.

Über jedes Objekt wurden den Benutzern in einem ersten Schritt drei Fragen gestellt: „Wie heißt dieses Objekt?“, „Welche Art von Objekt ist das?“ und „Was kann man mit diesem Objekt machen?“. Dies sollte das Lernen von Objekten durch einen Roboter simulieren, der diese Informationen ansammelt. Nachdem alle Objekte „gelernt“ wurden, mussten die Teilnehmer ein Objekt beschreiben, indem sie den Satz „Gib mir …“ vervollständigten.

Die Ergebnisse wurden alle zusammen auf das Auftreten von verschiedenen Arten, wie man das Objekt referenzieren kann, untersucht. Für jedes Objekt wurde vom Autor als Versuchsleiter festgelegt, welche Eigenschaften das Objekt beschreiben und welcher Klasse das Objekt in dieser Auswahl von Objekten zugeordnet werden kann, um die Antworten zu untersuchen. Bei der Zusammenstellung der Objekte wurde darauf geachtet, dass verschiedene Konstellationen von Objektbeziehungen im Testset vorhanden sind. Es gab zum Beispiel drei große Löffel und einen kleinen. Außerdem gab es drei Bücher, wobei ein Roman und zwei Reiseführer ausgestellt waren. Hierbei gab es die Klasse „Buch“ und die Unterklassen „Roman“ und „Reiseführer“. Eine genaue Auflistung der möglichen Antworten ist in Tabelle 4-1 gegeben. Die Angaben der Probanden wurden klassifiziert um herauszufinden, wie Personen Objekte beschreiben und referenzieren. Die Unterklasse von einem Objekt ist zwar noch eine Art der Klassifizierung, aber ist bereits so speziell, dass es auch als Bezeichnung dieser Klasse angesehen werden kann. Eine Klasse macht aus, dass damit andere Objekte abgegrenzt werden und das Objekt, in dieser Sammlung von Gegenständen, damit so klassifiziert wird, dass nur noch gleiche oder ähnliche Objekte mit in dieser Klasse sind. Eine Oberklasse ist die nächst generellere Klasse und eine Unterklasse die speziellere Klasse zum gewählten Typ. Eine Oberklasse von „Roman“ wäre zum Beispiel „Buch“. Es fällt auf, dass die verschiedenen Eigenschaften, die ein Objekt beschreiben nicht bei allen Objekten angegeben werden können. So macht es zum Beispiel keinen Sinn den Titel eines Löffels anzugeben, wenn dieser nicht aus einer speziellen Designerlinie stammt oder die Größe von einem Buch, wenn es nicht außergewöhnlich groß ist. Es gibt zwar auch Bücher die einen halben Meter groß sind, oder nur wenige Zentimeter klein sind, aber die normalen Bücher in diesem Szenario haben keine speziell angebbare Größe.

Objekt	Klasse	Oberklasse	Unterklasse	Titel	Farbe	Zustand	Größe
1	Roman	Buch	Fantasy	Silmarillion	golden		
2	Reise-führer	Buch	Lonely Planet	Estland	lila		
3	Tee			Kamille	gelb		
4	Tasse	Geschirr	Kaffee-tasse	Schlosstasse	schwarz	leer	
5	Tasse	Geschirr	Kaffee-tasse	Schlosstasse	schwarz	voll	
6	Tasse	Geschirr	Kaffee-tasse	Skype	blau	leer	
7	Tasse	Geschirr	Kaffee-tasse	Studenten-werk	weiß	leer	
8	Schlag-sahne	Lebens-mittel		Milbona	blau		
9	Schlag-sahne	Lebens-mittel		Milbona	blau		
10	Senf	Lebens-mittel	süßer Senf	Hendlmaier	rot		
11	Handy	Telefon		Sony Ericsson	silbern		
12	Löffel	Besteck	Esslöffel		silbern		groß
13	Löffel	Besteck	Esslöffel		silbern		groß
14	Löffel	Besteck	Esslöffel		silbern		groß
15	Löffel	Besteck	Kaffee-löffel		silbern		klein
16	Buch	Lektüre	Reise-führer	Island	blau		
17	Kaffee	Getränk	Pulver-kaffee	Nescafe	rot		
18	DVD	Film	Science Fiction	Cypher	weiß		

Tabelle 4-1: Mögliche Antworten für alle Objekte

Bei der Frage, welche Beschreibung das Objekt haben könnte, verwendeten manche Benutzer mehrere Wörter. Eine übliche Äußerung war zum Beispiel „schwarze Tasse". Dies bildet nicht nur die Eigenschaft Farbe ab, sondern bestimmt auch gleich die Art des Objektes, nämlich „Tasse". Dadurch entstanden bei allen 177 Namensgebungen insgesamt 204 einzelne Informationen. Die genaue Auflistung, welche Mittel die Probanden wählten, um das Objekt näher zu bestimmen, ist in Tabelle 4-2 angegeben. Darin ist angegeben, wie oft die einzelnen Merkmale und die verschiedenen Hierarchiestufen der Objektklassifizierung bei den verschiedenen Fragen verwendet wurden. Im Folgenden werden die einzelnen Ergebnisse der Fragen ausgewertet.

Benutzeräußerung	Wie heißt das Objekt?	Was für eine Art von Objekt?	Gib mir…
Autor (z.B. Tolkien)	4	0	3
Transfer (z.B. Kaffee)	2	4	7
Persönlich (z.B. Lieblingstasse)	1	0	1
Titel (z.B. Studentenwerk)	37	0	36
Farbe (z.B. schwarz)	5	0	19
Zustand (z.B. voll)	4	4	11
Größe (z.B. klein)	9	0	10
Klasse (z.B. Tasse)	76	51	106
Unterklasse (z.B. Kaffeetasse)	53	2	46
Oberklasse (z.B. Geschirr)	13	120	19
Gesamtanzahl	204	181	258
Anzahl Versuche	177	177	176

Tabelle 4-2: Gesamtergebnisse des Vortests zu Objektreferenzen

4.2.1. Objektbeschreibungen

Die erste Frage über jedes Objekt war die nach der Bezeichnung die das Objekt bekommen soll. Ohne genaue Erklärung, was diese Bezeichnung für den Roboter bedeutet, wurde hier oft die eigentlich noch nicht gefragte Objektklassifikation in die Klasse (37%) oder sogar deren Oberklasse (6%) vorgenommen. Ein so gelerntes Objekt wäre nicht ausreichend genau in die Objekthierarchie einsortiert, damit es in späteren Dialogen erfolgreich gefunden werden kann. Hier ist eine genauere Instruktion des Benutzers notwendig, damit klar wird, welche Information gesucht wird. Die Verteilung der Antworten auf die Frage nach der Beschreibung ist in Abbildung 4-2 dargestellt.

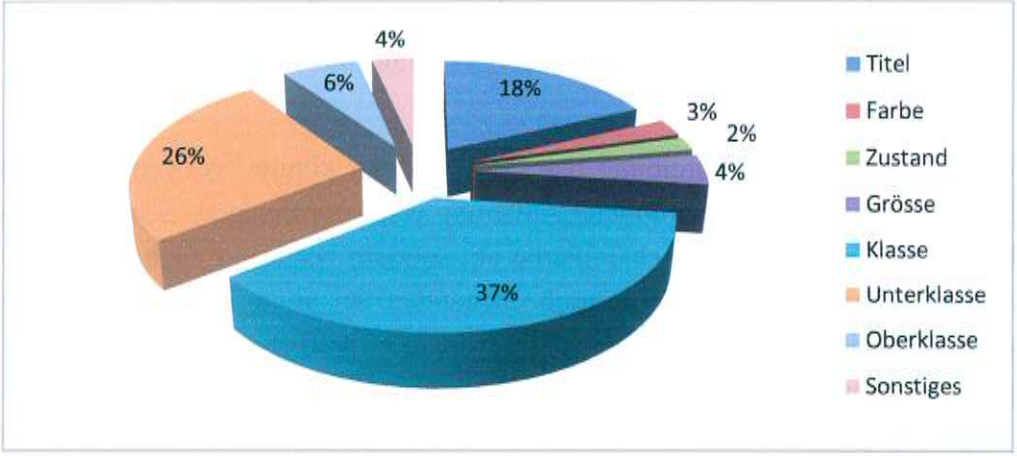

Abbildung 4-2: Angaben bei der Frage „Wie heißt dieses Objekt?“

Die Frage funktionierte gut, wenn der Benutzer einen Titel angeben konnte. Ansonsten wurde eine Klassifikation angegeben. Das heißt, dass diese Frage sehr ähnlich verstanden wurde, wie „Was für eine Art von Objekt ist das?“, die aber noch bessere Antworten zur Klassifikation als Antworten bekam. Um eine Eigenschaft als Antwort zu bekommen, muss diese Frage umgestellt werden.

4.2.2. Objektklassen

Die nächste Frage hatte das Ziel, eine Klassifikation des Objektes vorzunehmen. Dabei ist nicht dringend notwendig, dass mit der initialen Äußerung die Einteilung bereits vollständig und eindeutig ist. Für das Beispiel eines Buches, das den Titel „The Silmarillion“ trägt, wäre eine Einteilung als „Buch“ oder „Lektüre“ oder „Roman“ sinnvoll. Die Verteilung der gegebenen Angaben zur Klassifikation sind in Abbildung 4-3 dargestellt.

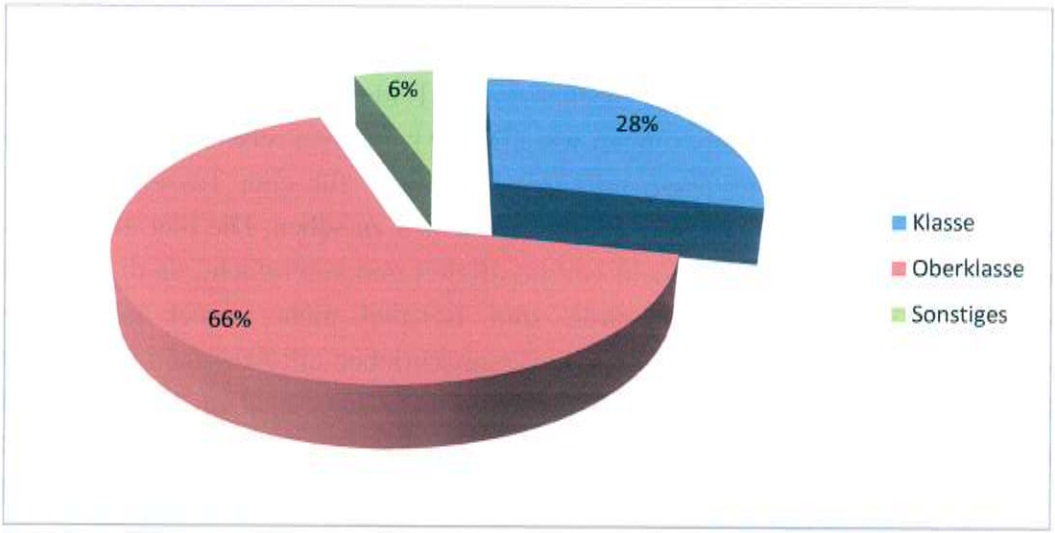

Abbildung 4-3: Angaben bei der Frage „Was für eine Art Objekt ist das?“

Die Ergebnisse dieser Frage sind erfolgreich, da die Objekte durchwegs erfolgreich klassifiziert wurden. Dabei wurden aber verschiedene Abstraktionsebenen verwendet, so dass es nötig ist, das Objektmodell so aufzubauen, dass der Roboter verstehen kann, dass ein Roman eine Unterklasse von einem Buch ist. Viele Benutzer verwendeten Klassen, die weiter spezifiziert werden können, um die Objekte in dieser Auswahl besser zu beschreiben. Da für das gleiche Objekt verschiedene Ebenen der Abstraktion verwendet wurden, ist es nötig, dass das Objektmodell um eine Ontologie erweitert wird, in der abgebildet wird, was für eine Art von Objekt das ist.

4.2.3. Objektfunktionalität

Während dieses Experiments wurde außerdem noch die Frage gestellt, was man mit diesem Objekt machen kann. Daraus sollte eine automatische Klassifikation in Objekte abgeleitet werden, die den Benutzer schneller an das Ziel führt, dem Roboter das Objekt zu lehren. Leider waren die gegebenen Antworten überhaupt nicht in eine Form zu bringen, damit man robust Informationen gewinnen konnte. Mögliche Antworten, die Informationen enthalten hätten, die für den Roboter verständlich wären, wären zum Beispiel „lesen", „essen" oder „kochen" gewesen. Diese Antworten wurden auch in vielen Fällen gegeben. Aber in oft waren Antworten wie „löffeln" für Löffel, „trinken" für Tassen, aus denen man zwar trinken kann, aber die man nicht trinken kann und „zubereiten" für Tee nicht genau genug, dass daraus Aussagen über das Objekt gemacht werden konnten. Es fällt auf, dass manche Benutzer lange zögerten, bis sie eine Antwort darauf geben konnten, was man mit einem Glas Senf machen kann. Je länger die Benutzer überlegten, desto komplizierter wurden die Antworten. Die Funktionalität von einem materiellen Objekt anzugeben kann schwer sein, wenn man damit zu erklären versucht, was man generell damit machen kann. Dadurch werden die Erklärungen länger und es entstehen Formulierungen wie „Damit kann man reisen im Baltikum" für einen Reiseführer oder „die Tasse hilft beim Trinken" für eine Tasse. Die genaue Auflistung der gegebenen Antworten ist in Tabelle 4-3 zu sehen. Die hier aufgeführten Antworten sind aber bereits von Hand vorklassifiziert und vereinfacht, da die einzelnen Antworten variabler waren. So wurde zum Beispiel nicht immer nur „Lesen" angegeben, sondern diese Funktionalität wurde umschrieben mit Aussagen wie „Ich lese das Buch", „Man kann es lesen" und „Lesen". Auch die Angaben „Helfen" stellen Sammlungen von Antworten dar, die aus Aussagen wie „Das ist ein Hilfsmittel", „Das hilft beim Trinken" und „Das ist eine Esshilfe" zusammengefasst wurden. Aus solchen Äußerungen Aussagen über das Objekt abzuleiten ist schwer.

Objekt	Genannte Kategorien				
Buch	Lesen: 28	Reisen: 2			
Tee	Trinken: 7	Öffnen: 1	Verzehren: 1	Zubereiten: 1	
Handy	Telefonieren: 8	Kommunizieren: 2			
Tasse	Trinken: 26	Füllen: 12	Helfen: 2		
Löffel	Essen: 21	Umrühren: 5	Helfen: 7	Zu sich nehmen: 4	Löffeln: 3
Lebens-mittel	Essen: 13	Kochen: 8	Konsumieren: 3	Verfeinern, Backen, Würzen: je 2	
Kaffee	Trinken: 5	Kaffee: 2	Aufkochen, Aufsetzen, Füllen: je 1		
DVD	Anschauen: 5	Abspielen: 2			

Tabelle 4-3: Benutzerangaben zur Funktionalität von Objekten

Sehr allgemeine Fragen über Objekte sind für Benutzer schwer zu beantworten. Deshalb wird auf Fragen nach allgemeinen Konzepten und Funktionalitäten verzichtet, um dem Benutzer einen leicht zu verstehenden Dialog zu ermöglichen. Um die Informationen über mögliche Funktionalitäten von Objekten in dieser Arbeit zu modellieren, werden Verwendungszwecke in der Ontologie modelliert, so dass direkt nach möglichen Funktionalitäten gefragt werden kann, um das gesuchte Objekt weiter einzugrenzen. Dabei werden die möglichen Funktionalitäten an Objektklassen vererbt.

4.2.4. Objektreferenzen

Nachdem alle Objekte durch die vorherigen Fragen beschrieben wurden, mussten die Benutzer noch einmal jedes der Objekte mit ihren eigenen Worten referenzieren, damit der fiktive Roboter ihnen das Objekt reichen kann. Die Ergebnisse der Auswertung der Referenzen, welche die Probanden verwendeten um ein Objekt zu beschreiben, sind in Abbildung 4-4 dargestellt.

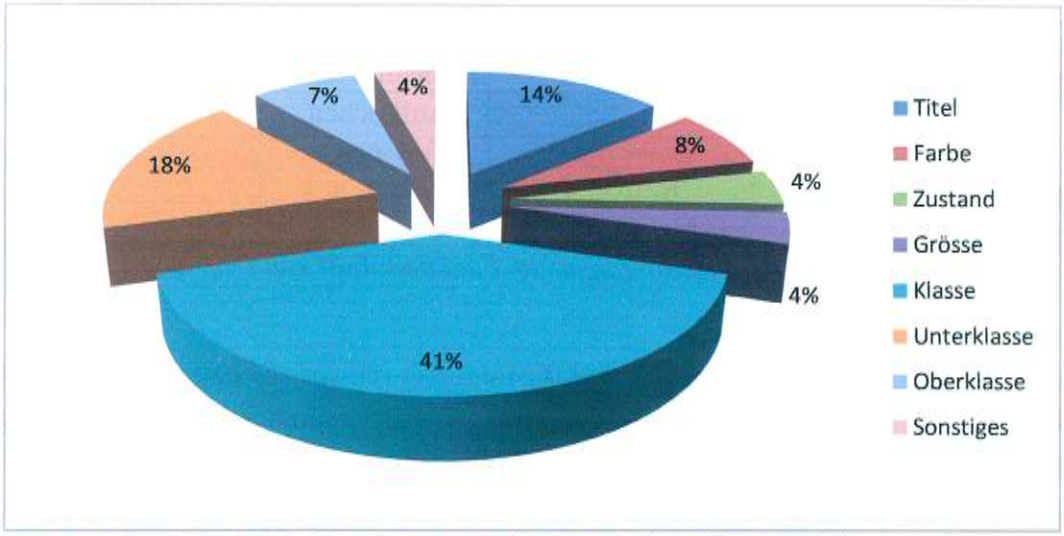

Abbildung 4-4: Verteilung von Angaben bei der Frage „Gib mir…"

Hierbei verwendeten die Benutzer oft nicht nur die Klasse oder ein Merkmal, um den Gegenstand zu identifizieren, sondern sie verwendeten Kombinationen aus verschiedenen Informationen. Interessanterweise hielten sich manche Benutzer hierbei nicht an die vorher gelernten Klassen oder Namen, sondern verwendeten andere Merkmale, die der Roboter so noch nicht gekannt hätte. Um das eben gelernte Objekt zu finden, hätte der Roboter nochmal einen Dialog starten müssen, um herauszufinden, dass der gewünschte „Esslöffel" der eben gelernte „große Löffel" ist.

Des Weiteren viel auf, dass bei der freien Formulierung viele Benutzer mehrere Bezeichner verwendeten. Wie in Tabelle 4-2 aufgelistet, wurden in etwa der Hälfte aller Fälle Kombinationen aus einer Eigenschaft und einer Einordnung in Klassen verwendet. Zum Teil wurden sogar mehrere Eigenschaften angegeben, wie zum Beispiel „Gib mir die volle, schwarze Tasse" oder „Gib mir den lila Reiseführer über Estland". Hierbei war die Kombination von Objekteigenschaften und Objektklassen für alle Benutzer eine gängige und normale Art, die Objekte zu beschreiben. Deshalb muss das Objektmodell erweiterbar sein und mehrere Klassen und Eigenschaften von einem Objekt modellieren können.

4.2.5. Zusammenfassung

Die durchgeführten Tests haben gezeigt, dass Menschen Kombinationen aus Eigenschaften und Typen verwenden, um Objekte zu beschreiben. Dabei tauchen bei verschiedenen Benutzern und in verschiedenen Situationen unterschiedliche Eigenschaften und Klassifikationen auf. Deshalb muss das Objektmodell verschiedene Attribute für ein Objekt speichern und verwalten können.

Es werden Eigenschaften wie Farbe und Form, aber auch spezielle Attribute, wie der Titel oder Zustand verwendet. Die Frage „Wie heißt dieses Objekt?" wurde von den Testpersonen sehr unterschiedlich verstanden. Zum Lernen von Objekten muss aber trotzdem eine Art der Fragestellung gefunden werden, die klar genug formuliert, dass der Benutzer eine Eigenschaft angeben sollte. Damit in den späteren Tests für die Evaluation des Gesamtsystems Antworten über Eigenschaften gegeben werden, werden genauere Informationen für die Probanden gegeben. Die Frage „Wie heißt dieses Objekt?" bedingt für viele die Art des Objekts, da der Name, auf den die Eigenschaften abstrahiert werden, nur bei Objekten wie Büchern oder Waren von bestimmten Firmen zu erkennen ist. Deshalb wird die Frage umformuliert zu „Welche Eigenschaft hat dieses Objekt?" und vor der Interaktion mit dem Roboter wird genauer erklärt, wie der Roboter Objekte erkennt und unterscheidet. Die Eigenschaften, die zur Differenzierung von Objekten der gleichen Klasse verwendet werden und somit keine direkten Unterklassen darstellen können ansonsten frei formuliert werden.

Die Klassifikation von Objekten in Oberklassen, Klassen und Unterklassen muss in einer hierarchischen Ontologie realisiert werden, da Benutzer verschiedene Abstraktionsebenen verwenden, um einen Gegenstand zu beschreiben.

Die Funktionalität eines Objektes ist für Benutzer schwer durch rein sprachliche Aussagen zu beschreiben. Deshalb werden mögliche Verwendungszwecke in der Ontologie modelliert und nur zum Interpretieren von Objekten verwendet. Beim Lernen können durch Fragen nach Funktionalität die möglichen Objektklassen eingegrenzt werden.

Diese Erkenntnisse wurden verwendet, um ein Objektmodell und eine Ontologie, die in Kapitel 4.5.3 beschrieben werden, zu entwickeln, die in einem Dialog über Objekte verwendet werden können, um Objekte zu referenzieren und zu lernen. Eng damit verbunden sind Entscheidungen, die den Algorithmus zum Lernen von Objekten, der in Kapitel 4.5.4 erklärt wird, beeinflussen.

4.3. Applikationsumgebung

Um das gesetzte Ziel, das Erlernen von Objekten und deren Bedeutung in einem Dialog mit einem Roboter zu untersuchen, müssen die Benutzer mit echten Objekten und einem echten Roboter interagieren. Es geht darum, dass untersucht wird, welche Einflüsse diese Umstände auf die Menschen haben, wie sie die Objekte wahrnehmen und den Dialog aufbauen. Deshalb wurde der Dialog in die Präsentations- und Forschungsplattform ARMAR-III, die im Sonderforschungsbereich 588 an der Universität Karlsruhe in Zusammenarbeit mit einer Vielzahl von anderen Forschungseinrichtungen entwickelt wird (Deutsche Forschungsgemeinschaft, 2001), integriert. Der Dialog ist nach dem Vorbild des Dialogs in (Holzapfel & Prommer, 2006) entwickelt, in dem der Roboter mit Objekten interagieren soll und herausfinden muss, welches Objekt ein menschlicher Benutzer gerade beschreibt. Dabei gibt es verschiedene Objekttypen und verschiedenfarbige Objekte, die der Roboter in seinem Umfeld wahrnehmen kann.

ARMAR-III ist ein neu entwickelter humanoider Roboter, der in einem Küchenszenario lernt, mit seiner Umwelt und mit Menschen zu interagieren. Der Roboter besteht aus einer fahrbaren Plattform mit einem holonomen Antrieb, der in der Lage ist, jederzeit in jede Richtung zu fahren und mit drei Laserscannern die Umgebung abtastet. Darauf sind drei Computer und zwei Akkus integriert, welche die ganze Verarbeitung auf dem Rechner und die Stromversorgung garantieren. Darauf sitzt ein Torso mit drei Freiheitsgraden, mit zwei weiteren Computern und 10 digitalen Signalprozessoren, welche die Sensordatenverarbeitung behandeln. Zwei Leichtbauarme mit jeweils sieben Freiheitsgraden ermöglichen es, dass der Roboter mit Objekten in seiner Umwelt interagiert. Jede Roboterhand hat acht weitere Freiheitsgrade an den fünf Fingern. Der Roboterkopf hat ebenfalls sieben Freiheitsgrade und hat zwei Farbkameras pro Auge und sechs Mikrophone. Dadurch ist der Roboter in der Lage, dass er viele Details in seiner Umgebung wahrnehmen und verarbeiten kann. Mehr über ARMAR-III gibt es in (Asfour, et al., 2006). Abbildung 4-5 zeigt den Roboter in der Küche an der Universität Karlsruhe.

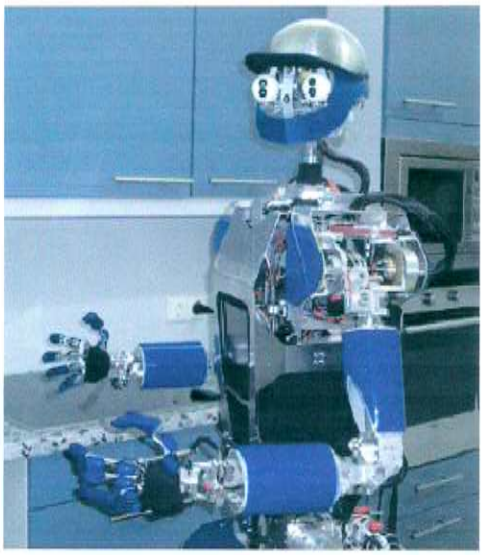

Abbildung 4-5: ARMAR-III

Zur Steuerung des Roboters soll ausschließlich Sprache verwendet werden, wobei der Benutzer nicht durch Vorgaben, betreffend das Vokabular oder dem Satzbau, eingeschränkt werden soll. Der Benutzer hat die Möglichkeit, frei und ohne lange Einarbeitungszeit mit dem Roboter zu sprechen, wie mit einem anderen Menschen.

Die Objekte, mit denen der Roboter im Laufe der Dialoge zu tun haben kann, sind aus dem Umfeld der Küche und persönlichen Gegenständen ausgewählt. Die genaue Art, wie der Roboter einzelne Objekte eingelernt hat und welche er bereits kennt wird den Benutzern nicht direkt offenbart, da der Benutzer in einer realen Umgebung auch nicht wissen kann, was der Roboter bereits kennt. Diese Objekte können an verschiedenen Stellen in der Küche auftreten, die der Roboter namentlich kennt. Eine Gestenerkennung um relative Positionen von Objekten zu verarbeiten wird nicht integriert.

4.4. Setup des Gesamtsystems

Das Gesamtsystem zum Lernen neuer Objekte ist ein Dialogsystem, das um die Objekterkennung durch die Kameras des Roboters erweitert wurde. Abbildung 4-6 zeigt den Systemaufbau des Dialogsystems.

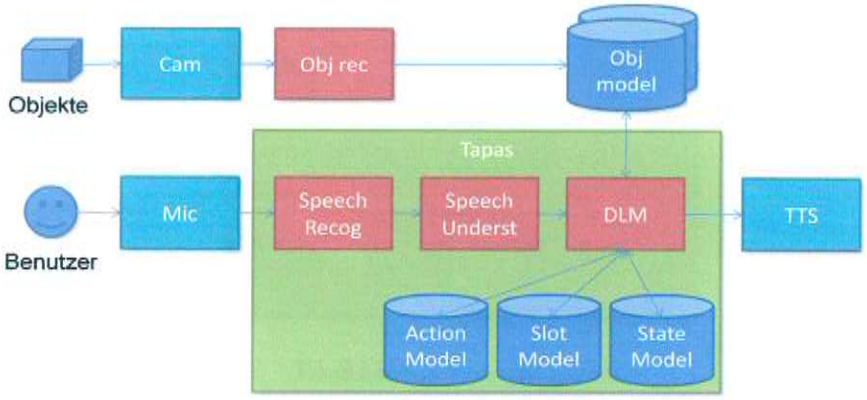

Abbildung 4-6: Systemaufbau

Der Benutzer wird mit Hilfe eines Mikrophons aufgenommen. Dafür können Nahbesprechungsmikrophone oder Fernbesprechungsmikrophone verwendet werden. Ein Fernbesprechungsmikrophon nimmt nicht nur den Benutzer selbst, sondern auch verschiedene Hintergrundgeräusche, wie die Lüfter der Klimaanlage oder von Computern, Stimmen von anderen Menschen und andere Bürogeräusche auf. Ein Nahbesprechungsmikrophon filtert diese Hintergrundgeräusche heraus und ist zwar aufwendiger zu bedienen, weil der Benutzer dieses Mikrophon erst aufsetzen muss, aber die Erkennung ist meistens deutlich besser. In dieser Arbeit werden daher Headsets verwendet.

Der verwendete Spracherkenner wird in Kapitel 2.1 beschrieben. Die Ausgabe des Roboters wird mit Hilfe des Text-To-Speech Moduls gemacht. Dieses Modul liest einen Text, den der Dialogmanager generiert, von einer Computerstimme vor, damit der Benutzer einen echten Dialog führen kann und nicht an eine herkömmliche Ausgabe mit Hilfe von Computerbildschirmen gebunden ist.

Der Objekterkenner liefert Hypothesen, welche Objekte gerade im Bild der Roboterkamera präsent sind. Das integrierte System wird in (Azad, Asfour, & Dillmann, 2007) vorgestellt. Dabei verwendet es die in Kapitel 2.2 vorgestellten SIFT Features, um texturierte Objekte zu erkennen. Diese können auch online gelernt werden, da diese Objekterkennung in der Lage ist, neue Objekte zu detektieren, indem sie gefundene Features in dem Bild untersucht und auf kein bekanntes Objekt zuweisen kann. Dann generiert der Objekterkenner eine Hypothese für ein neues Objekt und übergibt diese Information an den Dialogmanager. Einfarbige dreidimensionale Objekte können nicht online neu gelernt werden, weil hierfür die Form und Größe des Objektes

mit einem aufwendigen 3D-Scanner erfasst werden muss, bevor der Roboter diese Objekte erkennen kann.

Im Dialogmanager (DLM) wird die gewonnene Information des Benutzers verarbeitet und der nächste Schritt des Dialogsystems berechnet. Eine spezielle Strategie wird entwickelt, mit der interaktiv das Objektmodell um neue Objekte erweitern werden kann, die Informationen des Benutzers mit bereits bekannten Objekten vergleichen werden und der Dialogzustand verarbeitet wird. Die Funktionalität des Gesamtdialogs ist in Dialogmodule aufgeteilt, die für die Bearbeitung von einzelnen Zwischenzielen zuständig sind. Die genaue Aufteilung und die Auswahl der Dialogmodule wird in Kapitel 4.5.4 erläutert.

4.5. Gesamtdialog

In diesem Kapitel wird beschrieben, was nötig ist, die in Kapitel 1.1 gestellten Anforderungen zu erfüllen und wie das in einer Applikation umgesetzt wurde.

4.5.1. Grammatik und Diskursinformationen

Natürliche Sprache kann vielfältig auftreten. Einerseits ist es wünschenswert, dass der Benutzer nicht durch Regeln der verwendeten Grammatik eingeschränkt wird, aber andererseits ist es viel schwerer freie Sprache korrekt zu verstehen. Viele kurze, ähnliche Wörter, die in ähnlichem Kontext auftreten können, werden leicht verwechselt. Für die erfolgreiche Weiterverarbeitung durch den Roboter ist es jedoch meistens nicht wichtig ob der Benutzer „This is a red cup" oder „That's a red cup" gesagt hat. Vor der Verarbeitung werden semantische Informationen aus der Aussage extrahiert. Die Bedeutung der beiden Beispielsätze oben ist gleich und deshalb werden solche Fehler toleriert.

Die verwendete Grammatik hat viele verschiedene Varianten, die möglich sind, das gleiche zu sagen. Aus dem Verstandenen werden aber nur folgende Informationen für den Dialog weiter verwendet: Intention des Benutzers, Eigenschaft, Art, Aufenthaltsort und Zielort des Objektes Am Beispielsatz „Carry the red cup from the table to the sink" soll dieses Vorgehen beschrieben werden.

- Die Intention des Benutzers („carry")
    - Der Benutzer kann den Roboter dazu auffordern, verschiedene Aktionen auszuführen. Diese werden mit Hilfe von Schlüsselwörtern erkannt. Durch „carry" wird klar, dass der Roboter etwas tragen soll. Andere Möglichkeiten wären „bring" oder „take". Als Standardaktion, wenn der Benutzer keine Aktion benannt hat, erzählt der Roboter nur, was für Objekte er findet.
- Eigenschaft des Objektes („red")
    - Die Eigenschaft eines Objektes kann zum Beispiel die Farbe oder der Titel sein. Bereits bekannte Eigenschaften werden in einer Datenbank gespeichert, damit der Spracherkenner diese erkennen kann und an den Dialog weiterleiten kann. Unbekannte Eigenschaften werden als OOV (Out of Vocabulary) angezeigt. Sollte dieser Fall eintreten, wird ein Unterdialog ausgeführt, in dem der Name gelernt wird.
- Art des Objektes („cup")
    - Die Art eines Objektes wird in der Ontologie beschrieben und ebenfalls in einer Datenbank für den Spracherkenner gespeichert. Diese Objektklassen sind auch erweiterbar und können in einem Dialog erlernt werden.

- Aufenthaltsort und Zielort des Objektes („table“ und „sink“)
    - Diese Angaben treten in Verbindung mit Schlüsselwörtern wie „from“ oder „to“ auf. Mögliche Orte von Objekten sind im Vorhinein festgelegt, da der Roboter diese Angaben mit Punkten in seiner Umgebung verknüpfen muss.
- Ja und Nein
    - Für Nachfragen von dem Roboter an den Benutzer stehen dem Benutzer eine Reihe an verschiedenen Antworten zur Verfügung, um diese Hypothesen anzunehmen oder abzulehnen.

Diese Informationen werden aus dem Diskurs in sogenannte Slots eines Slotmodels geschrieben, damit diese Information über eine Benutzeräußerung hinaus bestehen bleiben. Sie stellen das aktuelle Wissen des Roboters über das gewünschte Objekt dar. Zudem gibt es Slots, in denen Informationen über zu lernende Objekte gespeichert werden.

4.5.2. Die Objekterkennung

Der Objekterkenner hat eine Reihe von Objekten mit Hilfe ihrer SIFT Features gespeichert und liefert diese an das Dialogsystem. Das Dialogsystem hat für jedes bekannte Objekt einen Eintrag in einer Objektdatenbasis, in der mögliche Eigenschaften und Klassen dieses speziellen Objektes aufgeführt sind. Diese Objektdatenbasis kann im Laufe des Dialogs erweitert werden, um sich an verschiedene Benutzer anzupassen. Dieses Vorgehen wird in Kapitel 4.5.4 beschrieben. Tabelle 4-4 zeigt einen Ausschnitt aus dieser Datenbasis.

ID	Type	Name
1	cup	red
2	dvd	friends
3	dvd	batman
4	plate	blue
5	book	iceland travel
6	ipod	black

Tabelle 4-4: Objektdatenbasis

Objekte gleichen Typs werden mit Hilfe von verschiedenen Eigenschaften unterschieden. Die genaue Vorgehensweise, wie die gefundenen mit dem beschriebenen Objekt verglichen werden, wird in Kapitel 4.5.4 beschrieben.

4.5.3. Objektmodell und Ontologie

Wie Kapitel 4.2 zeigt, ist es sinnvoll, Objekte durch Eigenschaften und Klassen zu beschreiben. Da die Frage nach der Objekteigenschaft jedoch nicht richtig beantwortet wurde, wurde für die Tests des Gesamtsystems die Einführung der Benutzer und die Art der Fragestellung leicht geändert, damit diese besser erklären, was nötig ist, um Objekte erfolgreich zu lernen. Es wird darauf hingewiesen, dass jedes Objekt eine Klasse hat. Zudem hat jedes Objekt eine Beschreibung, oder einen Namen, der dieses spezielle Objekt dieser Klasse näher beschreibt. Das kann im Fall einer DVD der Titel, im Fall eines Buches der Titel oder der Name des Autors sein. Aber auch andere Eigenschaften, wie die Farbe, Größe oder Form können dazu verwendet werden, das Objekt zu benennen. Ein „kleiner Löffel", eine „rote Tasse" oder ein „Harry Potter Buch" können somit gelernt werden. Für den anschließenden Test des Gesamtsystems ist eine kleine Änderung in der Versuchsbeschreibung notwendig, in der kurz erläutert wird, was die Objektbezeichnungen und Objektklassen für den Roboter bedeuten. Dieses Wissen verbessert die Kommunikation zwischen dem Benutzer und dem Roboter ohne die Art und Weise, wie der Benutzer mit dem Roboter interagiert einzuschränken.

Das mögliche Wissen des Roboters über Objekte wird in einem Objektmodell gespeichert. Dafür verwendet der Roboter eine Ontologie, die einen initialen Teil der auftretenden Objekte modelliert. Diese Ontologie ist um weitere Konzepte erweiterbar und dient dazu, die Objekte untereinander zu unterscheiden. In der Ontologie werden zum einen die Arten von Objekten hierarchisch klassifiziert. Das heißt, dass allgemeine Konzepte weiter verfeinert werden, je weiter man in der Ontologie ansteigt. In Abbildung 4-7, die das Konzept der verwendeten Ontologie darstellt, werden Objektklassen durch die roten Knoten abgebildet. Ein Objekt, das in der Küche auftritt, kann in diesem Fall zum Beispiel Geschirr, ein Getränk oder etwas zu Essen sein. In einer weiteren Verfeinerung werden die Getränke in heiße und kalte Getränke unterteilt. Es kann auftreten, dass eine Klasse von mehreren Klassen erbt und diese Konzepte zusammen vereinigt. Die reinen Objektklassen sind Konzepte, in welche die Objekte nach ihrer Art unterteilt werden können. Wenn ein neues Objekt zu einer bestimmten Klasse gehört, aber in keine der vorhandenen Unterklassen eingeordnet werden kann, kann eine neue Objektklasse erstellt werden.

In einem zweiten Schritt werden Funktionalitäten modelliert, um darzustellen, was man mit einem Objekt machen kann. In Abbildung 4-7 stellen die blauen Knoten Funktionalitäten dar. Diese Verwendungszwecke können von verschiedenen Knoten in der Ontologie geerbt werden, um zu beschreiben, wofür man das Objekt verwenden kann. Damit werden semantische Klassen abgebildet, da an diese Funktionalität verschiedene Aufgaben geknüpft werden können, die der Roboter erfüllen kann. So ist es zum Beispiel möglich, dass der Benutzer sagt „Ich bin durstig" oder „Gib mir etwas zu trinken" und der Roboter sucht nur Objekte, die „trinkbar" sind. Zum anderen kann der Roboter beim Lernen von Objekten diese Funktionsweisen benutzen, um die

Objektklasse des unbekannten Gegenstandes herauszufinden, indem er fragt „Kann man dieses Objekt trinken?“

Der Roboter lernt Objekte, indem er den Benutzer so lange über dieses Objekt befragt, bis er eine eindeutige Objektklasse gefunden hat. Zudem kann der Benutzer eine Eigenschaft dieses speziellen Gegenstandes angeben. Das kann zum Beispiel der Titel, die Farbe oder der Geschmack sein. In Abbildung 4-7 stellen die lila Knoten Eigenschaften dar. So ist der Roboter in der Lage, einen „Orangen Saft“, eine „rote Tasse“ oder eine „Star Wars DVD“ zu lernen. Damit können Objekte nach Ihrer Objektklasse unterschieden werden und gleiche Objekte über spezielle Eigenschaften dieses Gegenstandes. Der genaue Ablauf dieses Klärungsdialogs wird in Kapitel 4.5.4 beschrieben.

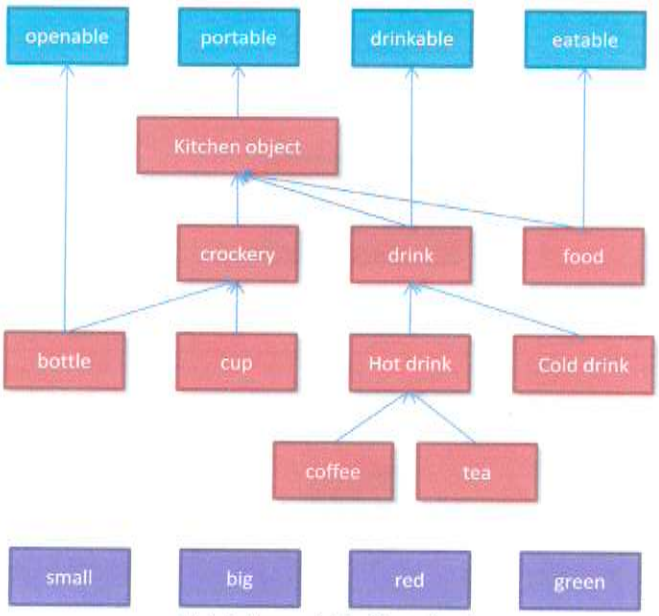

Abbildung 4-7: Ontologie

Um sicherzustellen, dass sich das Objektmodell an den Benutzer während dem Ablauf anpasst, ist es außerdem möglich, dass der Roboter im Laufe mehrerer Dialoge lernt, dass ein spezielles Objekt auch andere Bezeichnungen haben kann. So ist es zum Beispiel möglich, dass zwei Benutzer das gleiche Objekt als „Orangen Saft“ und als „Hohes C Saft“ bezeichnen. Da der Roboter den neuen Namen noch nicht kennt, wird er durch Nachfragen an den Benutzer den alternativen Namen lernen, sobald er herausgefunden hat, welches Objekt gemeint ist. Auch mehrere verschiedene Objektklassen können gelernt werden, wenn ein Benutzer das Objekt als „Saft“ und der andere als „Flasche“ bezeichnet. Der genaue Ablauf dieses Subdialogs wird in Kapitel 4.5.4 beschrieben.

Die erstellte Ontologie umfasst Objekte, die in der Küche auftreten können und neun Funktionalitäten. Die abgebildeten Verwendungsmöglichkeiten sind: cook, drink, eat,

fill, open, play, carry, switch on und watch. In der verwendeten Ontologie ist es möglich, dass der Benutzer die Ontologie um neue Objekte erweitert, aber es ist nicht möglich, dass die Ontologie komplett interaktiv aufgebaut wird.

4.5.4. Dialogablauf

In erster Linie hat der Dialog immer das Ziel, herauszufinden, welches Objekt der Benutzer gerade will. Dieser minimale Dialog ohne Zwischenfälle ist in Abbildung 4-8 dargestellt.

Abbildung 4-8: Dialogablauf ohne Zwischenfälle

- Intention erkennen
  Aus der Äußerung des Benutzers wird seine Intention extrahiert. Dies geschieht anhand von Schlüsselwörtern wie „carry" oder „bring". Wenn keines dieser Schlüsselwörter auftaucht, wird der Roboter die gefundenen Objekte nur benennen aber nicht damit interagieren, da der gesamte Dialog darauf ausgerichtet ist, dass der Benutzer etwas mit den Objekten machen oder neue lernen möchte.

- Objekt identifizieren
  Als erstes und immer, wenn ihn der Benutzer dazu auffordert, versucht der Roboter Objekte an verschiedenen Stellen der Küche zu erkennen. Dafür bewegt er sich zu der, von dem Benutzer genannten Stelle und startet die Objekterkennung.
  In einem zweiten Schritt wird das gewünschte Objekt identifiziert, indem die gefundenen Objekte und die verstandenen Aussagen über dieses Objekt verglichen werden, bis ein eindeutiges Objekt gefunden wurde. Hierbei ist es möglich, dass der Benutzer eine Oberklasse des Objektes verwendet, indem er zum Beispiel sagt: „give me the dish" und damit meint er die Tasse, die ebenfalls ein Geschirr ist. Außerdem sind Synonyme in der Ontologie gespeichert, die angeben, dass „cup" und „mug" die gleiche Objektklasse darstellen. Um mehrere Eigenschaften für ein Objekt zu verstehen, werden in der Objektdatenbasis verschiedene Eigenschaften gespeichert. Somit ist es möglich, dass ein Benutzer eine Tasse als „the big cup" oder „the red cup" bezeichnet. Durch Nachfragen des Roboters werden die möglichen Objekte soweit eingegrenzt bis eines übrig bleibt.
  Ein Objekt wird als mögliche eingestuft, wenn alle folgenden Punkte zutreffen:
  - Wenn das Objekt an der von dem Benutzer beschriebenen Stelle in der Küche ist
    oder der Benutzer keine spezielle Stelle angegeben hat
  - Wenn die Beschreibung des Objektes der Beschreibung des Benutzers entspricht
    oder der Benutzer keine Beschreibung verwendet hat
    oder der Benutzer ein Synonym zu der Beschreibung verwendet hat
    oder der Benutzer eine in der Objektdatenbasis gespeicherte alternative Beschreibung verwendet hat
  - Wenn die Klasse des Objektes der Klasse des Benutzers entspricht
    oder der Benutzer keine Klasse verwendet hat
    oder der Benutzer ein Synonym zu der Klasse verwendet hat
    oder der Benutzer eine in der Objektdatenbasis gespeicherte alternative Klasse verwendet hat
    oder der Benutzer eine Oberklasse der gespeicherten Klasse verwendet hat
  - Wenn das Objekt nicht direkt abgelehnt wurde

  Bei gleichen Objekten, wie zum Beispiel zwei identischen roten Tassen wird nach einer speziellen Tasse gefragt, ob der Benutzer diese will. Falls kein Objekt zu der Beschreibung des Benutzers passt, wird direkt nach einem der gefundenen Objekte gefragt. Es könnte möglich sein, dass ein Benutzer eine neue Eigenschaft des Objektes beschreibt. Nach der Klärung, ob es sich doch um das Objekt handelt, wird das Objektmodell dieses Gegenstandes erweitert.

Dieser Dialog wird so lange ausgeführt, bis alle notwendigen Informationen über Eigenschaft, Klasse und Position von dem Objekt gesammelt wurden. Der Benutzer kann diese direkt in seiner Aufgabenstellung angeben. Falls einige Informationen fehlen, fragt das System direkt nach diesen Angaben.

- Ende
  Wenn die Erkennung eindeutig ist, schließt der Roboter den Dialog ab und erfüllt den Wunsch des Benutzers. Das identifizierte Objekt wird je nach Intention behandelt oder es wird ausgegeben, dass es nicht möglich ist, den Wunsch auszuführen.

Es kann auch sein, dass während des Dialogs mit dem Benutzer ein unbekanntes Wort auftaucht. Sollte dies der Fall sein, wird der Dialog zum Finden eines Objektes unterbrochen und ein neuer Subdialog wird ausgeführt, in dem erst geklärt wird, wie dieses Wort heißt. Abbildung 4-9 zeigt diesen Ablauf.

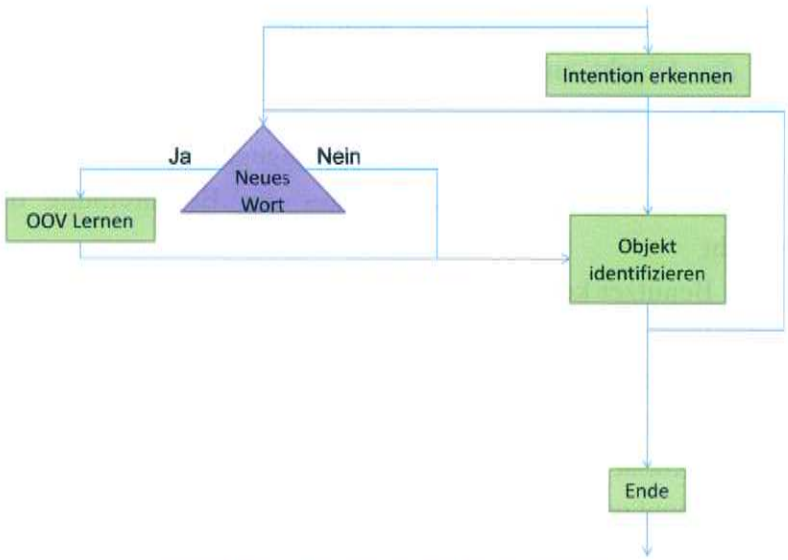

Abbildung 4-9: Dialogablauf mit OOV

- Neues Wort
  Hier wird entschieden, ob ein OOV aufgetreten ist oder normal mit dem Dialog weitergemacht werden kann.
- OOV Lernen
  Der Dialog zum Wörter Lernen wird in zwei Schritten ausgeführt, bis der Benutzer die gewonnene Hypothese des Roboters annimmt. Dabei wird erst darauf hingewiesen, welches Wort nicht richtig erkannt wurde. Das kann an der Stelle der Objekteigenschaft oder der Objektklasse auftreten. Der Benutzer kann bis zu drei Mal dieses Wort wiederholen und der Spracherkenner versucht dieses zu verstehen. Die mehrmalige Wiederholung erhöht bei dieser kurzen Spracheinheit, die nur aus

„It is gray" oder „gray" selbst besteht, die Chancen, dass das richtige Wort verstanden wird. Der Roboter fragt den Benutzer nach einem verstandenen Wort, falls er überhaupt etwas Sinnvolles verstanden hat.

Wenn diese Erkennung nicht erfolgreich angenommen wurde, kann der Benutzer das Wort buchstabieren. Da der verwendete Buchstabiererkenner nur ein Wort verstehen kann, ist die Namensgebung auf ein einzelnes Wort begrenzt. Aus der erstellten Hypothese der Buchstabiersequenz generiert der Roboter eine Phonemfolge, die er dem Benutzer vorliest, um herauszufinden, ob das Wort richtig erkannt wurde. Dies geschieht bis das Wort richtig gelernt wurde oder abgebrochen wird. Das Wort wird mit der erzeugten Phonemfolge in das Wörterbuch gespeichert und in der Grammatik als Möglichkeit eingefügt. So ist es zur Laufzeit möglich, dieses Wort direkt zu verwenden.

Das unbekannte Wort wird als Alternative in der Objektdatenbasis gespeichert, wenn das zugehörige Objekt gefunden wurde.

Wenn der Roboter die Objekte erkennt, kann es auch dazu kommen, dass er ein unbekanntes Objekt vor sich sieht. Dann wird er den Benutzer darauf hinweisen und das Objekt gelernt. Abbildung 4-10 zeigt den realisierten Dialogablauf, der so das Gesamtsystem beschreibt.

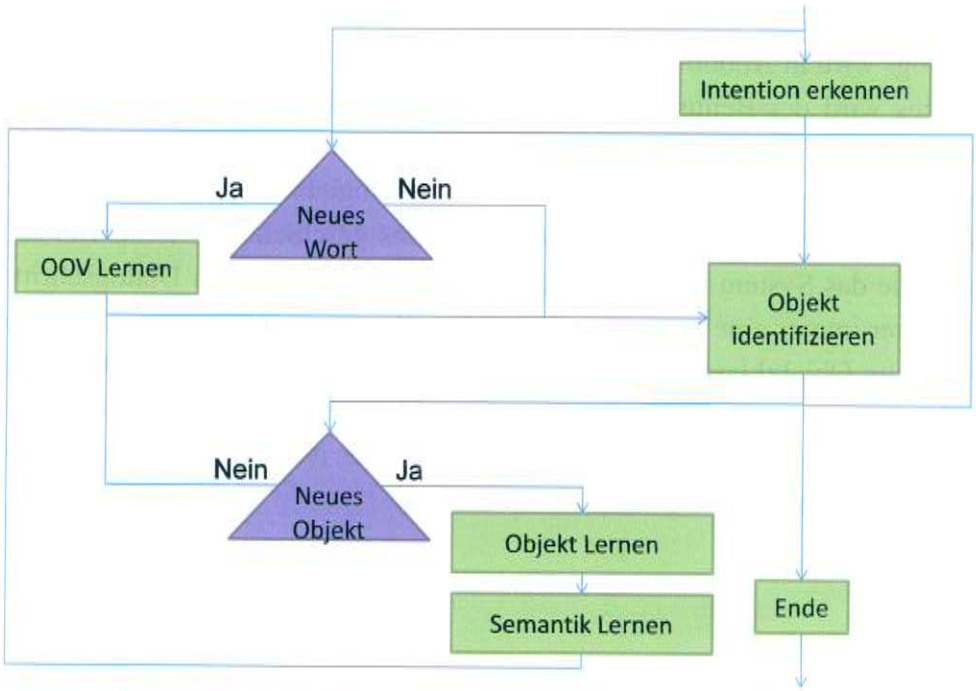

Abbildung 4-10: Dialogablauf mit OOV und neuen Objekten

- Neues Objekt
  Falls der Objekterkenner liefert, dass ein unbekanntes Objekt gefunden wurde, startet ein Subdialog zum Lernen von Objekten, bevor mit dem Finden des gewünschten Objektes weitergemacht wird. Die visuellen Daten, also die in Kapitel 4.5.2 beschriebenen SIFT Features, werden gespeichert und im Objektmodell mit einem neuen Objekt registriert.
- Objekte Lernen
  Dieser Teilbereich läuft genauso ab wie der OOV Lernen Subdialog, der weiter oben erklärt ist. Da es sich jedoch um ein unbekanntes Objekt handelt, startet der Dialog mit der Einführung, dass eine Objekteigenschaft gelernt werden soll. Darin erklärt der Roboter mit Hilfe von Beispielen, dass Eigenschaften Farben, Größen, Titel oder der Autor sein können. Dann wird der Benutzer aufgefordert, eine Eigenschaft anzugeben. Der Roboter versucht, diese zu verstehen und lernt dabei gegebenenfalls ein neues Wort. Der gelernte Name wird für das neue Objekt übernommen.
- Semantik Lernen
  In diesem Subdialog fragt der Roboter nach der Objektklasse. Erst wird kurz erklärt, was eine Objektklasse ist, indem der Roboter anhand von Beispielen angibt, dass das Objekttypen wie Tasse, DVDs oder Flaschen sein können. Der Ablauf dieses Subdialogs wird in Abbildung 4-11 dargestellt.
  Als erstes wird der Benutzer dazu aufgefordert, eine Objektklasse anzugeben. Im weiteren Verlauf des Dialogs versucht der Roboter, diese Objektklasse so weit zu verfeinern, dass er sie möglich tief in dem Ontologiebaum einsortieren kann, also bis die gewählte Objektklasse eindeutig ist. Dies geschieht mit Hilfe von Ja-Nein-Fragen, die das System über das Objekt generiert. Als nächsten Dialogschritt wählt der Roboter immer eine Frage aus, die er dem Benutzer stellen kann, um die aktuell ausgewählte Objektklasse weiter zu verfeinern. Dabei kann er entweder nach einer Funktionalität fragen oder nach einer Unterklasse. Der Benutzer bestätigt diese Hypothesen oder lehnt sie ab, indem er mit Ja oder Nein darauf antwortet, bis das System angibt, die Klasse gelernt zu haben.

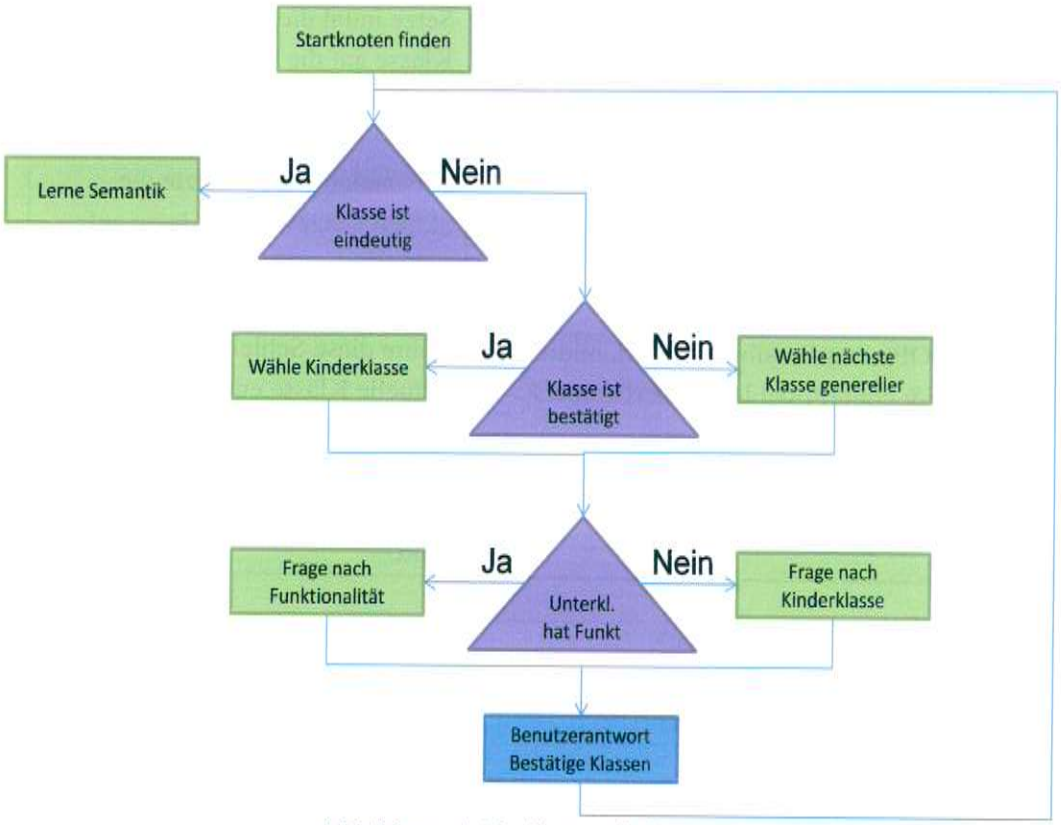

Abbildung 4-11: Semantik Lernen

Um sicherzustellen, dass der Roboter den Startknoten richtig gewählt hat und nicht eine falsch verstandene Hypothese weiter verfolgt wird, fragt das System nach, ob der Startknoten so richtig erkannt wurde. Jeder Knoten in der Ontologie kann möglich oder bestätigt gesetzt werden. Am Anfang sind alle Knoten möglich und keiner ist bestätigt.

Jeder Schleifendurchlauf dieses Algorithmus stellt eine Frage des Roboters und eine Antwort des Benutzers dar. Der Benutzer kann außerdem jederzeit durch die Angabe einer Klasse direkt an diese Stelle in dem Ontologiebaum springen und so den Dialog abkürzen. Der genaue Ablauf ist in Tabelle 4-5 dargestellt.

1	actualClass = user input	Setze inital die ausgewählte Klasse auf die verstandene Klasse des Benutzers
2	IF (!askUserForActualClass()) actualClass = „object“	Falls diese Klasse falsch verstanden wurde, setzte die aktuell ausgewählte Klasse auf „object“
3	actualClass = confirmed	Bestätige die aktuelle Klasse
4	LOOP(countPossibleNumberChildren > 0 && countPossibleNumberBrothers > 0)	Führe diese Schleife aus, bis die aktuelle Klasse entweder keine Kindsknoten mehr hat oder keine anderen Knoten auf der gleichen Ebene wie dieser Knoten möglich sind.
	{	
5	IF(actualClassHasUsage) askUserForUsage()	Falls es eine Funktionalität gibt, frage danach
6	ELSE askForNextClass()	Falls es keine Funktionalität gibt, frage nach einer speziellen Klasse
	}	Ende der Schleife
7	return actualClass	Liefere die gelernte Klasse zurück

Tabelle 4-5: Semantik Lernen Algorithmus

In Tabelle 4-6 wird der Ablauf eines Beispieldialogs zum Lernen der Objektklasse dargestellt. Der Benutzer startet den Dialog mit der Aussage, dass es ein Objekt aus der Küche ist. Vor ihm steht ein Saft. Die rot markierten Teile, sind die tatsächlich gesprochenen Sätze im Dialog. In Abbildung 4-12 ist der hier angewandte Teil der Ontologie abgezeichnet. Dieser Ausschnitt stellt einen Auszug aus der verwendeten Ontologie dar.

Step					
1	User: „This is an object from the kitchen“ actualClass = „kitchen object“				
2	Robot: „Is your item a kitchen object?“ User: „Yes“				
3	kitchen object = confirmed				
4	kitchen object has 4 children: food, drink, dish, cutlery kitchen object has 2 brothers: media object and office tool	kitchen object has 3 children: drink, dish, cutlery kitchen object has 2 brothers: media object and office tool	Drink has 3 children: hot drink, juice, soft drink Drink has no brothers	Drink has 2 children: juice, soft drink Drink has no brothers	Juice has no children Juice has no brothers Go To 7
5	kitchen object has a possible usage: eat Robot: „Can I eat this object?“ User: „No“ Go To 4	kitchen object has a possible usage: drink Robot: „Can I drink this object?“ User: „Yes“ actualClass = „drink“ Go To 4	Drink has no possible usage	Drink has no possible usage	
6			Ask for child: hot drink Robot: „Is this a hot drink?“ User: „No“ Go To 4	Ask for child: juice Robot: „Is this a juice?“ User: „Yes“ actualClass = „juice“ Go To 4	
7					Return Juice Robot: „I have learned, that this is a juice“

Tabelle 4-6: Beispieldialog

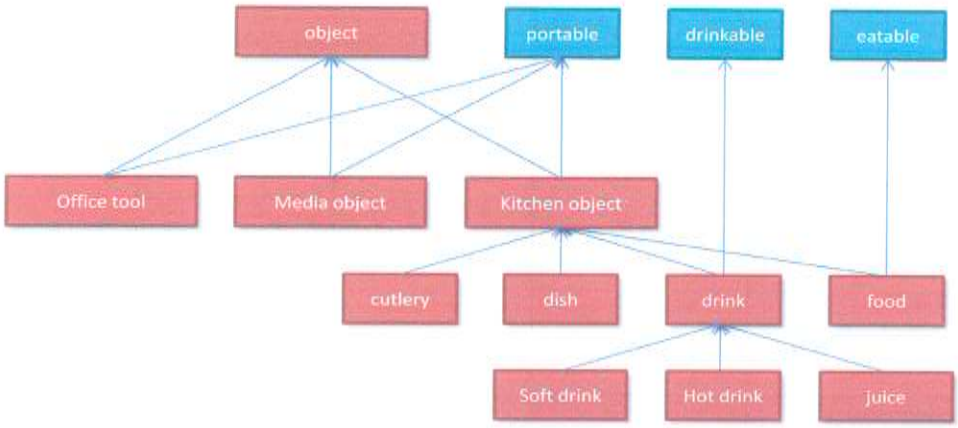

Abbildung 4-12: Ontologieausschnitt

Falls es an der aktuellen Stelle in der Ontologie möglich ist, dass der Roboter nach einem Verwendungszweck des Objektes fragt, wird dies ausgeführt. Das lockert den Dialog auf, weil nicht immer nach Objekttypen gefragt wird und es können mehrere Objektklassen zusammen behandelt werden. So sind zum Beispiel eine Flasche, ein Gefäß und eine Schüssel füllbar, aber ein Teller und eine Platte nicht. Alle stehen aber auf der gleichen Ebene und stellen ein Unterkonzept von „Geschirr" dar. So kann die Anzahl der möglichen Klassen schneller eingeengt werden. Der Ablauf dieser Nachfrage ist in Tabelle 4-7 aufgelistet.

askUserForUsage()	Fragen nach Funktionalitäten einer Klasse
{	Da auch die Funktionalitäten Knoten in der Ontologie sind, werden diese ebenfalls auf möglich und bestätigt gesetzt.
IF (actualClassIsConfirmed)	Falls die aktuelle Klasse bestätigt wurde
{	
askNextUsageOfClass	Frage nach einer Funktionalität von Kindern dieser Klasse
}	
ELSE	Falls die Klasse nicht bestätigt wurde
{	
askNextUsageOfParent	Frage nach einer Funktionalität, von einem anderen Kind der Elternklasse
actualClass = parent	Und setze die aktuelle Klasse auf die Elternklasse
}	
IF (userAnswerIsYes)	Verarbeite eine Bestätigung der Funktionalität
{	
actualClass = possible	Die Klasse, die diese Funktionalität hat ist weiter möglich
usageClass = confirmed	Die Klasse, die diese Funktionalität hat wird bestätigt
actualUsage = confirmed	Die Funktionalität ist bestätigt
}	
ELSE	Verarbeite eine Ablehnung der Funktionalität
{	
actualClass = notPossible	Die Klasse, die diese Funktionalität hat ist nicht mehr möglich
allChildrenOfActualClass = notPossible	Alle Klassen, die von der Klasse mit dieser Funktionalität erben, sind nicht mehr möglich
actualUsage = notConfirmed	Die Funktionalität ist abgelehnt
}	
}	

Tabelle 4-7: Algorithmus: Fragen nach Funktionalität

Falls kein Verwendungszweck angegeben ist, wird der Benutzer direkt nach einer Unterklasse gefragt, um immer speziellere Aussagen über das Objekt zu bekommen. Die Generierung und Behandlung dieser Frage ist in Tabelle 4-8 angegeben.

askForNextClass()	Fragen nach der nächsten Klasse
{	
IF (actualClassIsConfirmed)	Falls die aktuelle Klasse bestätigt wurde
{	
askNextSubclassOfClass	Frage nach einer Unterklasse dieser Klasse
}	
ELSE	Falls die Klasse nicht bestätigt wurde
{	
askNextSubclassOfParent	Frage nach einer anderen Unterklasse der Elternklasse
actualClass = parent	Und setzte die aktuelle Klasse auf die Elternklasse
}	
IF (userAnswerIsYes)	Verarbeite eine Bestätigung der Frage
{	
actualClass = possible	Die Klasse, die diese Funktionalität hat ist weiter möglich
}	
ELSE	Verarbeite eine Ablehnung der Funktionalität
{	
actualClass = notPossible	Diese Unterklasse, ist nicht mehr möglich
allChildrenOfActualClass = notPossible	Alle Klassen, die von dieser Unterklasse erben, sind nicht mehr möglich
}	
}	

Tabelle 4-8: Algorithmus: Fragen nach Klassen

Diese Sammlung an Dialogmodulen stellt zusammen die Funktionalität zum Lernen neuer Objekte dar. Die Auswahl, welches Dialogmodul gerade arbeiten darf, geschieht durch die Dialogstrategie, in der jedes Modul aufgefordert wird, eine Bewertung des aktuellen Zustandes vorzunehmen. Wenn ein Modul den aktuellen Zustand verarbeiten kann, weil zum Beispiel ein unbekanntes Wort aufgetaucht ist und das Modul zum Lernen von OOVs ausgeführt werden möchte, liefert es eine hohe Wertung zurück. In der Dialogstrategie wird ausgewählt, welches Modul in der nächsten Interaktion ausgeübt wird. Die realisierten Dialogmodule sind leicht durch alternative Lösungen für dieses Teilproblem austauschbar und erfüllen eine abgeschlossene Aufgabe in dem Gesamtdialog. Somit sind sie auch für andere Applikationen nutzbar. Wie in Abbildung 4-10 dargestellt, gibt es Module zum Dialogsetup (Intention erkennen), Objekt grounding (Objekt identifizieren), Wörter lernen (OOV lernen und Objekt lernen), Bedeutung von neuen Objekten lernen (Semantik lernen) und dem Ende des Dialogs (Ende).

5. Experimente und Analysen

Um den Dialog zu evaluieren, wurden Tests mit Benutzern durchgeführt. Hierfür haben sechs Studenten aus verschiedenen Fachrichtungen mit dem Roboter jeweils mehrere Dialoge geführt. Als Einführung und Erklärung wurden Handzettel verteilt, damit alle Benutzer ein einheitliches Vorwissen hatten. Dieser Text ist in Tabelle 5-1 aufgeführt.

Sehr geehrter Testteilnehmer,

vielen Dank, dass Du dir Zeit genommen hast für meine Benutzerstudie im Rahmen meiner Diplomarbeit. Ich werde mit etwa 10 Personen Tests durchführen, um mein Dialogsystem zu evaluieren. Jeder Teilnehmer bekommt die gleichen Fragen und Objekte, damit ich vergleichbare und aussagekräftige Ergebnisse erhalten kann, die meinen Dialog bewerten.

In diesem Szenario, geht es darum, dass Armar, der Roboter vor dir, mit Objekten interagiert. Er kann Objekte erkennen, diese transportieren oder dir reichen. Die Gegenstände stehen in der Küche verteilt und der Roboter wird auf deine Anweisungen hin, Aufgaben ausführen. Dabei wird er versuchen, herauszufinden, welches Objekt gerade gemeint ist und was er damit machen soll. Leider wird er nicht in der Lage sein, die Objekte tatsächlich in die Hand zu nehmen und dir zu bringen. Die meisten Objekte verliert er bisher noch beim Greifen oder unterwegs. Er wird dir also zum Abschluss sagen, was er machen wird, ohne das tatsächlich auszuführen.

Armar hat schon einige Objekte gelernt. Diese wird er wieder erkennen, wenn er sie in der Küche findet. Manche Gegenstände kennt er noch nicht. Wenn er ein unbekanntes Objekt findet, wird er dich fragen, was für ein Objekt das ist. Später wird er diese dann auch wieder erkennen. Für den Roboter hat jedes Objekt einen Typ, wie zum Beispiel Tasse oder Buch und eine Eigenschaft, wie zum Beispiel eine Farbe oder spezielle Bezeichnung, um es von anderen Objekten mit dem gleichen Typ zu unterscheiden. Der Roboter unterscheidet also eine rote Tasse von einer blauen Tasse, ein großes Buch von einem kleinen und generell, Bücher von Tassen. Die Eigenschaften, also auch die Farbe der Objekte werden nicht automatisch erkannt, sondern müssen einmal am Anfang für jedes Objekt gelernt werden. Deshalb fragt er dich auch nach einer Eigenschaft. Dabei kannst du alles Beliebige verwenden, das dir gerade passend erscheint, dieses Objekt von anderen seiner Art zu unterscheiden.

Du kannst mit dem Roboter fast reden, wie mit einem Menschen. Bitte beachte aber, dass der Roboter ein wenig braucht, bis er das, was du gesagt hast, verstanden hat und darauf reagieren

kann. Warte deshalb, bevor du ihm mehrere Anweisungen gibst. Bitte verwende auch nur einen Satz, um dich an den Roboter zu wenden, da er nach jeder kurzen Pause in deiner Äußerung direkt damit beginnt, diese zu verarbeiten und dir dann nicht mehr zuhört. Wenn der Roboter ein Wort noch nicht kennt, wird er dich bitten, dieses zu wiederholen. Wenn er dich bittet, das Wort zu buchstabieren, versuche so flüssig und schnell wie möglich zu buchstabieren. Du verwendest ein Headset, damit dich der Roboter besser verstehen kann. Armar versteht gerade nur Englisch.

Deine Aufgabe ist es, dir ein Objekt auszusuchen. Stell dir vor, dass du gerade am Tisch sitzt und nicht selbst aufstehen willst, um den Gegenstand zu holen. Dann kannst du den Roboter auffordern, dass er dieses Objekt nimmt, dir bringt oder einfach nur sagt, was das für ein Objekt ist. Dazu unterscheidet der Roboter zwischen verschiedenen Orten in der Küche. Du kannst zum Beispiel „on the table" oder „at the sink" verwenden. Der Roboter erkennt keine Gesten, also versteht er leider nicht, wenn du ihm sagst „it is there" und auf das Objekt zeigst. Um eine möglichst genaue Evaluation meines Dialogs machen zu können, habe ich andere Teile des Dialogs nicht realisiert. So ist es zum Beispiel nicht möglich, dass du dich mit dem Roboter über das Wetter oder Politik unterhalten kannst. Der Roboter soll nur dazu dienen, dir Objekte zu reichen und wird Objekte lernen, wenn er diese nicht kennt.

Für die Auswertung nehme ich alle Dialoge auf, damit ich herausfinden kann, wie gut das System auf deine Wünsche und Anweisungen reagiert hat.

Vielen Dank und viel Spaß,

Daniel

Tabelle 5-1: Erklärung für Probanden

Nach dieser kurzen Einführung wurden Objekte im Raum verteilt und der Benutzer sollte sich davon eines aussuchen, das er von dem Roboter haben möchte. Als ersten Dialog hatten alle Probanden nur bereits bekannte Objekte zur Auswahl. Dadurch wurde sichergestellt, dass sie sich erst an den Dialog und die Interaktion mit dem Roboter gewöhnen können, bevor dieser etwas von ihnen lernen möchte.

Zwischen einzelnen Benutzern wurden nicht alle neu gelernten Objekte aus der Datenbasis entfernt, um mehr Daten für die Auswertung zum Wiederfinden von gelernten Objekten zu erhalten. Da nur eine begrenzte Auswahl an Objekten zur Verfügung stand, mussten aber immer wieder gelernte Objekte gelöscht werden.

Nachdem bereits ein Vortest durchgeführt worden war, wurde in den Experimenten mit den Versuchspersonen das gesamte Dialogsystem untersucht. Dadurch, dass die Benutzer mit realen Objekten konfrontiert wurden und sie nicht wussten, welche dem Roboter bekannt sind und welche dieser lernen muss, kommen die Ergebnisse denen eines Einsatzes in einem Haushalt nahe. Dabei wurde ihnen zwar erklärt, was der Roboter kann und welche Möglichkeiten sie in dem Dialog haben, aber es wurden keine Einschränkungen bezüglich der Wortwahl oder dem Satzbau gemacht.

5.1. Evaluation

Ausgewertet wurden 50 Dialoge mit sechs unerfahrenen Benutzern. Die Dialoge starteten, indem der Benutzer den Roboter dazu aufforderte, ihm etwas zu bringen oder die Objekte an einer Stelle im Raum zu erkennen. Daraufhin wurde die visuelle Objekterkennung gestartet und der Roboter versuchte, die Aufgabe des Benutzers zu erfüllen. Wenn nötig, lernte das System neue Objekte.

5.1.1. Visuelle Objekterkennung

Insgesamt wurde der Roboter 52 mal dazu aufgefordert, Objekte zu erkennen. 40 mal war ein unbekanntes Objekt zu erkennen und 12 mal ein bekanntes. In 81% wurden die Objekte auf Anhieb korrekt erkannt und der Dialog konnte direkt fortgesetzt werden. In weiteren 13% lieferte der Dialog eine Fehlermeldung, die den Benutzer dazu veranlasste, das Objekt ein wenig zu drehen oder den korrekten Platz der gewünschten Objekte anzugeben. Dann versuchte der Roboter erneut das Objekt zu erkennen und war dabei meist erfolgreich. In 6% scheiterte die Objekterkennung, wobei einmal die Fehlermeldung des Roboters nicht verstanden wurde und zweimal ein bereits gelerntes Objekt als unbekanntes Objekt erkannt wurde und das Objekt erneut gelernt wurde. Tabelle 5-2 zeigt die Ergebnisse der Objekterkennung.

Richtig erkannt	94%	49	Sofort richtig erkannt	81%	42
			Nach Wiederholung richtig erkannt	13%	7
Falsch erkannt	6%	3	Unbekanntes Objekt zu schlecht um es zu lernen	2%	1
			Bekanntes Objekt als unbekanntes Objekt erkannt	4%	2

Tabelle 5-2: Ergebnisse der visuellen Objekterkennung

Die Erkennung von Objekten durch visuelle Merkmale war insgesamt sehr erfolgreich weil die Erkennung wenig Fehler machte und der Benutzer im Fehlerfall die Erkennung neu starten konnte. Durch eine klare Information des Benutzers mit Fehlermeldungen, konnten die Benutzer entweder die Erkennung direkt noch einmal ausführen lassen, oder erst die Objekte neu anordnen und dann erkennen lassen.

5.1.2. Objekt reichen

Die Aufgabe, dass der Roboter dem Benutzer etwas reichen sollte, wurde 18 mal gestellt. Das System war in 72% in der Lage, diesen Auftrag erfolgreich auszuführen. In fünf Fällen konnte der Dialog nicht erfolgreich beendet werden. Dabei hat der Spracherkenner in drei Fällen falsche Aussagen des Benutzers verstanden. Zwei mal

verstand der Spracherkenner den Benutzer bei der Spezifikation des gewünschten Objektes falsch und suchte deshalb nach dem falschen Objekt. Einmal verstand der Spracherkenner den Benutzer nicht, als dieser ein unter anderem Namen bekanntes Objekt als das gewünschte Objekt bestätigen wollte. Einmal konnte der Lerndialog zum Lernen eines unbekannten Objektes nicht beendet werden, so dass das System das gewollte Objekt auch nicht reichen konnte. Einmal konnte das gewünschte Objekt nicht vom visuellen Objekterkenner gefunden werden. Tabelle 5-3 zeigt die Ergebnisse des Systems beim Reichen von Objekten.

Erfolgreich ausgeführt	72%	13			
Fehler	28%	5	Fehler des Spracherkenners	17%	3
			Lernen nicht abgeschlossen	6%	1
			Fehler der Objekterkennung	6%	1

Tabelle 5-3: Ergebnisse zum Objekte reichen

Die Erkennung des gewünschten Objektes konnte meistens erfolgreich durchgeführt werden. Die auftretenden Fehler des Dialogsystems entstanden, weil der Spracherkenner eine Aussage des Benutzers falsch verstanden hat. Durch Ausgaben, was der Roboter jeweils verstanden hat, könnten diese Fehler behoben werden. Dadurch würden aber die Dialoge länger, was nicht gewünscht ist.

5.1.3. Lernen von Objekteigenschaften

Bei den 52 Erkennungen traten 40 unbekannte Objekte auf. Das System versuchte in diesen Fällen erst, eine Eigenschaft für jedes Objekt zu lernen. Um eine Objekteigenschaft zu lehren, konnten die Benutzer eine beliebige Eigenschaft wählen und eine Ausprägung davon angeben. Dies war in 83% der Fälle erfolgreich. Der Benutzer hatte die Möglichkeit, drei mal eine Eigenschaft anzugeben, die das System direkt versuchte zu verstehen. Wenn dies nicht erfolgreich war, wurde der Benutzer aufgefordert, den Namen zu buchstabieren. Tabelle 5-4 zeigt die Ergebnisse beim Lernen einer Objekteigenschaft.

In 55% der unbekannten Objekte konnte eine Eigenschaft direkt durch Nachfragen gelernt werden. Durchschnittlich wurde eine Eigenschaft 1,36 mal genannt, bevor diese erfolgreich gelernt wurde. In 8% trat während diesen Nachfragen ein Fehler auf, sodass eine fehlerhafte Eigenschaft gelernt wurde. Einmal wurde statt dem gewünschten „lemon" Geschmack die Art „lemonade" erkannt und im weiteren Verlauf „tea" als Name gelernt, weil der Proband „this is a lemon tea" wiederholt hat und das System nachfragte „Is this a tea?". Das wurde angenommen. Zwei Mal konnte der Spracherkenner die Äußerungen des Benutzers nicht korrekt verarbeiten, da dieser zu schnell neue Anweisungen gab, bevor der Roboter selbst antworten konnte. Diese Fälle

wurden abgebrochen und als Fehler bewertet. Die nicht erfolgreichen Dialoge hatte eine Durchschnittslänge von 2,67 Benutzerangaben.

In 28% wurden die Bezeichnungen erfolgreich durch Buchstabierung des Wortes gelernt. Dafür waren im Schnitt 2,55 Buchstabierungen durch den Benutzer notwendig. In 10 % war es nicht möglich, das Wort korrekt durch die Buchstabierung zu erkennen. Diese Versuche wurden nach 2,5 Durchgängen abgebrochen und als Fehler gewertet. Hierbei ist zu beachten, dass der Gesamtdialog zusätzlich eine Länge von drei Benutzerangaben hat, weil die automatische Erkennung des Wortes vor der Buchstabierung nicht erfolgreich war.

Erfolgreich gelernt	83%	33	Gelernt vor Buchstabieren	55%	22
			Anzahl der Nachfragen	Gesamt: 30	Durchschnitt: 1,36
			Gelernt durch Buchstabieren	28%	11
			Anzahl der Buchstabierungen	Gesamt: 28	Durchschnitt: 2,55
Fehler	18%	7	Fehler vor Buchstabieren	8%	3
			Anzahl der Nachfragen	Gesamt: 8	Durchschnitt: 2,67
			Fehler beim Buchstabieren	10%	4
			Anzahl der Buchstabierungen	Gesamt: 10	Durchschnitt: 2,5

Tabelle 5-4: Ergebnisse zum Lernen einer Objekteigenschaft

Das Vorgehen, erst zu versuchen, das gewünschte Wort direkt zu verstehen, war erfolgreich, da es bei bereits bekannten Worten schnell zum Ziel kam. Wenn das Wort nicht bekannt war, konnte das Wort mit Hilfe der Buchstabierung gelernt werden. Die erkannten, buchstabierten Sequenzen wurden am Besten verstanden, wenn der Benutzer sehr schnell und ohne Pausen buchstabiert. Dies ist für deutschsprachige Personen nicht einfach. Deshalb waren ein paar Wiederholungen nötig, bis das gewünschte Wort richtig verstanden wurde.

Während den ganzen Dialogen kam es nicht vor, dass das System neue Bezeichnungen für bereits bekannte Objekte lernen musste, da sich die Benutzer an die Namen, die der Roboter verwendet hatte, angepasst haben. Diese Möglichkeit wäre zwar in dem Dialog vorgesehen gewesen, wurde aber nie ausgeführt.

5.1.4. Lernen von Objektklassen

34 Mal versuchte das System, nach dem Lernen einer Eigenschaft, eine semantische Klasse für ein unbekanntes Objekt zu lernen. In 82 % der Fälle konnte dieser Dialog erfolgreich abgeschlossen werden. Sechs mal wurde der Dialog aufgrund von Fehlern

abgebrochen. Einmal stürzte der Dialog beim Bestätigen aufgrund von falsch verstandenen Eingaben des Spracherkenners ab, obwohl die richtige Klasse gefunden worden wäre. Zweimal lernte das System ebenfalls aufgrund von falsch verstandenen Aussagen beim Bestätigen eine falsche Klasse. Dreimal wurde bei der ersten Benutzeranfrage eine falsche Klasse, anstatt eines unbekannten Wortes verstanden und der Dialog war nicht in der Lage, diesen Fehler zu korrigieren. Eigentlich hätte der Dialog beim generellsten Objekt anfangen sollen, um herauszufinden, welche Klasse dieser Gegenstand hat. Der auftretende Fehler, dass der Roboter nach einer falschen initialen Klasse fragte und diese abgelehnt wurde, wurde aber nicht behandelt. Tabelle 5-5 zeigt die Ergebnisse in der Übersicht.

Erfolgreich gelernt	82%	28	Länge des Dialogs	Gesamt: 102	Durchschnitt: 3,6
Fehler	18%	6	Fehler des Spracherkenners	9%	3
			Fehler im Dialog	9%	3
			Länge des Dialogs	Gesamt: 18	Durchschnitt: 3

Tabelle 5-5: Ergebnisse zum Lernen der Objektklasse

Der vorgestellte Ansatz realisiert das Lernen von Objektklassen durch eine gemischte Lösung: Nach einer initialen Benutzeranfrage wird die Ontologie nach der richtigen Klassifizierung durchsucht, bis der Benutzer dieses Objekt erfolgreich einsortiert hat. Diese Dialogstrategien können einzeln oder kombiniert verwendet werden. Das vorgestellte System realisiert eine kombinierte Lösung, bei der erst durch eine Benutzeranfrage versucht wird, die Objektklasse direkt zu lernen (One Shot). Falls die Klassifikation nicht eindeutig ist, wird von dieser Klasse aus weiter spezifiziert (Combined). Falls keine Erstklassifikation möglich ist, startet das System bei der allgemeinsten Klasse (Browse). In Kapitel 4.5.4 wird dieser Ansatz beschrieben. In Tabelle 5-6 werden die drei Teile verglichen:

- One Shot: Das System lernt direkt die Klasse, der als erstes von dem Benutzer genannt wird, wenn diese in der Ontologie so eingeordnet werden kann, dass keine weitere Verfeinerung der Klassifikation nötig ist. Dieser Dialog ist zwei Benutzerinteraktionen lang, da erst die Klasse angegeben wird und das System nachfragt, ob diese richtig verstanden wurde. Dieses Vorgehen wurde in 47% ausgeführt und war in 81% der Fälle erfolgreich. Die gelernten Klassen waren in der Ontologie direkt an einem speziellsten Knoten einsortiert und konnten so sofort beendet werden.
- Browse: Falls die initiale Klassifikation nicht durchgeführt werden konnte, wird ein reiner Browse Dialog durchgeführt. Das System lernt die Klasse, indem es bei der generellsten Klasse beginnt und durch Fragen an den Benutzer die Klasse weiter

spezifiziert. Wird ausgeführt, wenn das System bei der initialen Frage keine Klassifikation versteht. Der durchschnittliche Dialog auf diese Weise ist 10,5 Interaktionen lang. In 6% wurde diese Strategie angewendet und alle Dialoge waren erfolgreich. Dieser Dialog war sehr erfolgreich, weil Benutzer nur einzelne Hypothesen bestätigen oder ablehnen mussten, aber da der Dialog sehr lang im Vergleich mit den anderen Strategien war, wurde dieser nur ausgeführt, wenn eine Vorklassifikation nicht möglich war. In Verbindung mit anderen Strategien ist dieser Ansatz sinnvoll, alleine wäre es aber nicht benutzerfreundlich, weil der Benutzer nicht immer durch eine schrittweise Klassifikation navigieren sollte, wenn die Objektklasse bekannt und direkt angegeben werden kann.

- Combined: In 47% aller Dialoge war die Erstklassifikation nicht eindeutig. Deshalb wurde ein kombinierter Dialog ausgeführt. Das System startet an der Stelle in der Ontologie, die der Benutzer in einer initialen Klassifikation angibt und sucht von dort aus nach der passenden Klasse. Dieser Dialog dauert im Durchschnitt 4,2 Interaktionen. Von den ausgeführten kombinierten Dialogen, die diesen Ansatz verfolgten waren 81% der Dialoge erfolgreich.

One Shot	47%	16	Richtig gelernt	81%	13
			Länge des Dialogs	Gesamt: 26	Durchschnitt: 2
			Fehler	19%	3
			Länge des Dialogs	Gesamt: 6	Durchschnitt: 2
Browse	6%	2	Richtig gelernt	100%	2
			Länge des Dialogs	Gesamt: 21	Durchschnitt: 10,5
			Fehler	0%	0
			Länge des Dialogs	Gesamt: 0	Durchschnitt: 0
Combined	47%	16	Richtig gelernt	81%	13
			Länge des Dialogs	Gesamt: 55	Durchschnitt: 4,2
			Fehler	19%	3
			Länge des Dialogs	Gesamt: 12	Durchschnitt: 4

Tabelle 5-6: Realisierte Lernstrategie

Wenn der Dialog nur das One-Shot Lernen realisiert hätte, wäre zwar der Dialog deutlich kürzer, da alle Objekte mit zwei Interaktionen gelernt worden wären, aber die Fehlerrate und die Genauigkeit der Klassifikation wären schlechter. Für erfolgreich gelernte Klassen benötigt die kombinierte Strategie 3,6 Interaktionen. Die Aussagen des Benutzers wurden untersucht und bewertet, ob die Erstklassifikation richtig gewesen wäre. Dabei traten durch Fehlhypothesen des Spracherkenners oder fehlendes Hintergrundwissen mehr Fehler auf. Die beiden Fälle, in denen keine initiale Klassifikation vorgenommen werden konnte, könnten aber von dieser Strategie nicht verarbeitet werden, deshalb wurden die Fälle als Fehler gewertet. Ein kombinierter Ansatz hätte die Klassen erfolgreich durch den Browse Dialog gelernt. In drei weiteren Fällen wurde bei der initialen Erkennung eine falsche Klasse erkannt, in denen der

kombinierte Dialog aber trotzdem zu einem korrekten Ende geführt hat. In einer reinen One Shot Strategie wären diese als Fehler aufgetreten. So würde die Fehlerrate von 18% aus Tabelle 5-5 auf 32% wie in Tabelle 5-7 dargestellt steigen.

One Shot	100%	34	Richtig gelernt	68%	23
			Länge des Dialogs	Gesamt: 52	Durchschnitt: 2
			Fehler	32%	11
			Länge des Dialogs	Gesamt: 16	Durchschnitt: 2

Tabelle 5-7: Nur One-Shot Lernstrategie

Beim Klassifizieren konnte der vorgestellte Algorithmus mit einem kurzen Dialog in 82% erfolgreich abschließen. Obwohl dieser Algorithmus im Schnitt um 1,6 Interaktionen länger als ein One-Shot Learning ist, sind die gelieferten Ergebnisse besser. Zum einen konnten 14% weniger Fehler gemessen werden und zum anderen waren die gelernten Klassen genauer. So wurden zum Beispiel Säfte und Tees nicht einfach als Getränk einsortiert, sondern weiter spezifiziert. Gelernte DVDs wurden nicht nur als Filme gelernt, sondern genauer beschrieben. Was genau diese bessere Einsortierung bringt, wurde nicht getestet, da hierfür weitere aufwendige Benutzertests in realen Umgebungen notwendig gewesen wären. Für eine längerfristige Nutzung des Roboters ist es aber leicht ersichtlich, dass eine Einordnung besser ist, wenn sie spezieller gemacht wird, da jede generelle Klasse auch verstanden wird, um dieses Objekt zu finden.

5.1.5. Nutzerbefragung

Nach den Tests wurden die Benutzer zu einzelnen Aspekten des Dialogsystems befragt. Dabei sollten sie auf einer Skala von 1 bis 10 (10 ist die Bestnote, 1 die schlechteste Note) bewerten, wie gut das System gearbeitet hat.

Objekterkennung	Wie gut hat der Roboter die vorhandenen Objekte erkannt?	Schnitt: 7,6
Spracherkenner	Wie viel von dem, was du gesagt hast, hat der Roboter verstanden?	Schnitt: 7,2
Dialogablauf	Wie gut hat der Roboter auf deine Forderungen reagiert?	Schnitt: 7,6
Neue Objekte	Wie gut hat der Roboter neue Objekte gelernt?	Schnitt: 8,0
Neue Wörter	Wie gut hat der Roboter neue Wörter gelernt?	Schnitt: 6,6
Neue Bedeutung	Wie gut hat der Roboter die Art von neuen Objekten gelernt?	Schnitt: 8,8
Gesamtsystem	Wie gut war das System im Ganzen?	Schnitt: 7,8

Tabelle 5-8: Bewertung der Probanden

Die Ergebnisse der Benutzerumfrage zeigen, dass das Gesamtsystem bei den Benutzern positiv bewertet wurde, wobei die Objekterkennung und das Lernen von Klassen die besten Noten bekamen. Hierbei wurde der kombinierte Ansatz zum Lernen von Objektklassen bewertet.

5.2. Ergebnisse

Zur Modellierung von realen Gegenständen sind Eigenschaften und hierarchische Objektklassen nötig. Wie die Ergebnisse des Vortests, der in Kapitel 4.2 beschrieben wird, zeigen, verwenden Menschen Klassifikationen von Gegenständen, um diese zu referenzieren. Dabei können verschiedene Eigenschaften dieser Gegenstände mit angegeben werden. Deshalb werden Objekte durch variable Kombinationen aus Eigenschaften und Objektklassen modelliert. Da die Klassen auf verschiedenen Abstraktionsebenen verwendet werden, ist es nötig, die Arten und Funktionalitäten in einer Hierarchie zu modellieren, damit der Roboter verschiedene Konzepte von Gegenständen unterscheiden kann. Dadurch können die Beschreibungen von den Benutzern robuster verstanden werden.

Ein kombinierter Ansatz zum Lernen von Objektklassen ist besser als reines One Shot Learning. Wie die Ergebnisse aus Kapitel 5.1.4 zeigen, kostet der hier vorgestellte Algorithmus kaum mehr Interaktionen mit dem Benutzer, dieser liefert aber bessere Ergebnisse betreffend Fehlerrate und Genauigkeit gegenüber einem einfacheren Ansatz. Das heißt, dass die nötige Objekthierarchie, die durch eine Ontologie realisiert wurde, erfolgreich und schnell mit Hilfe des vorgestellten Algorithmus durchforstet werden kann, um Objekte zu lernen.

In der Arbeit von Boris Schulz (Schulz, 2005) wurden auch neue Objekte gelernt und in eine Ontologie einsortiert. Diese Arbeit wird in Kapitel 3.3.3 vorgestellt. Dabei mussten neue Wörter gelernt werden und das Objekt wurde in semantische Klassen von Getränken einsortiert. Schulz behandelte Eigenschaften von Objekten als weitere Unterklassen in der Ontologie. Dadurch konnte er auch im Dialog nicht zwischen Eigenschaften und Klassen unterscheiden. Der hier vorgestellte Ansatz verfügt über einen genaueren Dialog und kann Fragen über die Objekte an den Benutzer detaillierter stellen. Die beiden Ansätze unterscheiden sich durch die Art der Fragestellung beim Lernen der Objektklasse. Während Schulz alle Möglichkeiten auf einmal aufzählt und keine Funktionalitäten abbildet, wählt der hier vorgestellte Algorithmus einzelne Knoten aus und fragt direkt danach oder nach einer möglichen Funktionalität. Der Algorithmus von Schulz hatte eine Erfolgsquote von 75%, wobei auch zum Lernen einer weiteren Eigenschaft ein langer Dialog durchlaufen werden musste. Die verwendeten Objekte waren aber nicht real und die semantischen Klassen waren begrenzt auf Getränke. Von den gesamten 40 Dialogen zum Lernen von Objekten konnten bei dem hier vorgestellten Algorithmus in dieser Arbeit 27 erfolgreich abgeschlossen werden. Diese Ergebnisse kommen zustande, wenn das System Eigenschaften und semantische Klassen zusammen bewertet werden. Von 40 Dialogen konnten 33 erfolgreich Eigenschaften lernen. 34 Dialoge, also 33 richtige und ein Dialog mit einer fehlerhaften Eigenschaft, wurde fortgesetzt, wobei 28 erfolgreich Klassen lernten. Das ergibt eine Performanz von 68% die sich mit den Ergebnissen von Boris Schulz messen lassen kann, zumal es sich um reale Objekte handelte und die

Ontologie einen größeren Objektbereich abdeckte. Um neue Eigenschaften für ein bekanntes Objekt zu lernen müsste Schulz eine weitere Unterklasse in der Ontologie lernen, was einen erneuten Lerndialog bedingt. Im hier vorgestellten Ansatz, der diese Aussagen trennt, wäre ein reiner Lerndialog für die Eigenschaft ausreichend. Dieser ist kürzer durchführbar.

Die Arbeiten von Dusan und Flanagan, die in Kapitel 3.3.2 beschrieben werden, haben semantische Informationen nur in Form von Farben und Formen. Das sind keine Aussagen über die Art der Objekte. Es gibt keine Ontologie, die Objekttypen abbildet, sodass es nicht möglich ist, Aussagen über die Sorte von Gegenstand zu machen. Deshalb kann auch kein Vergleich mit dem hier vorgestellten Ansatz zum Lernen von Objekttypen angegeben werden.

Das Gesamtsystem wird besser, wenn man Dialogfähigkeiten ausnutzt, um Erkennungsfehler zu kompensieren. Dabei werden zum Beispiel Statusinformationen verwendet, um Erkennungshypothesen implizit zu bestätigen und im Fehlerfall durch Benutzerunterstützung zu korrigieren. Zum Beispiel geschehen beim Erkennen von Objekten auf dem Tisch Fehler, wie Kapitel 5.1.1 zeigt. Da der Roboter deutlich meldet, was er erkannt hat oder warum er nichts erkennen konnte, kann der Benutzer einfach darauf reagieren, den Fehler beheben und den Roboter die Erkennung wiederholen lassen. Beim Lernen von Eigenschaften und Objektklassen, werden einzelne Hypothesen des Spracherkenners mit Hilfe von Konfirmationsfragen bestätigt, damit der Roboter nur richtige Attribute lernt, obwohl der Spracherkenner zum Teil fehlerhafte Hypothesen liefert.

Die Erkennung, Objektmodellierung und der Dialog wurden zusammen in einem kombinierten System integriert, das in der Lage ist, den Benutzer zu verstehen, Objekte zu lernen und damit zu interagieren.

5.3. Diskussion

Zusätzlich zu der reinen Auswertung der Qualität des Dialogs wurden mehrere intensive Diskussionen über das Objektmodell und die Ontologie geführt. Die Sichtweise von Objekten ist bei Menschen nicht einheitlich. So wird das gleiche Objekt einmal als „Schwarztee" und einmal als „rote Schachtel" bezeichnet werden. Der hier realisierte Ansatz verfolgt eine Einteilung der Objekte in Klassen, die den Verwendungszweck der Objekte beschreibt. So wird eine Packung Saft als Getränk einsortiert, das zu der Klasse der Säfte gehört. Eine sinnvolle Eigenschaft könnte die Farbe des Kartons, der Herstellername oder der Geschmack sein. Eine Packung mit Teebeuteln stellt ein Getränk dar, das Tee ist. Auch hier sind Farbe, Hersteller oder Geschmack mögliche Eigenschaften. Diese Klassifikation wurde gewählt, weil der Roboter als Zielsetzung gestellt bekommt, reale Objekte in einer natürlichen Umgebung zu erkennen und diese zu manipulieren. Der Benutzer wird nach Ansicht des Autoren eher die Aussage treffen: „Gib mir mal den Tee" als „Gib mir mal die rote Schachtel". Damit der Roboter die Objekte in semantische Klassen einordnen kann, an die Verwendungszwecke wie „Trinken" oder „Essen" geknüpft werden können, werden nicht einfach alle Objekte als „rote, blaue und grüne Boxen" gelernt. Obwohl es möglich wäre, die Ontologie um Verpackungen zu erweitern und alternative Klassen in dem Objektmodell zu lernen, ist der Ansatz, „Boxen" zu lernen, nicht als erste Klassifikation sinnvoll. Die hier vorgestellte Klassifikation soll dem Roboter ermöglichen, Getränke von Essen zu unterscheiden, dass er dem Benutzer das richtige bringen kann, wenn dieser durstig ist. Wenn alles nur als „Box" klassifiziert wird, ist das nicht möglich. Für einen Benutzer ist diese Vorgehensweise aber nicht immer einfach nachzuvollziehen, so dass manchmal trotzdem Verpackungen genannt werden. In welcher Situation wie vorgegangen wird, ist für jedem Menschen anders und so kann es sein, dass ein Mensch zum Beispiel einen „Orangensaft" aber auch eine „volle Flasche" lernt, je nachdem, ob sein Fokus auf dem Inhalt oder der Verpackung liegt. Auch die Situation ändert die Wahrnehmung von Objekten. So kann zum Beispiel beim Frühstückstisch, mit der vollen Kaffetasse „Gib mir mal den Tetra pack" bedeuten, dass der Benutzer die Milch möchte, während am Mittag mit einem leeren Saftglas „Gib mir mal den Tetra Pack" bedeuten kann, dass der Benutzer den Orangensaft möchte. Eine präzisere Aussage wäre sicher, wenn der Benutzer direkt das Getränk jeweils angegeben hätte, aber in der Situation war es nicht weiter nötig.

Beim Sammeln der Informationen über ein Objekt müssen die Fragen genau und präzise gestellt werden, damit die gewünschte Antwort gegeben wird. Die beiden verwendeten Fragen: „I am sorry, but i do not know this object. What is a property of this object? Please do not use the type of the object to describe it now. I will ask you later what kind of object this is. Please tell me a specific title, the size or the color of the object." und „Now I know, that this item is a [red] object. What kind of object is it? Please tell me the class of object, that describes this object. An object class can be a cup or a book, for

example." wurden nicht immer richtig verstanden und lösten Diskussionen aus, was die korrekte Antwort darauf ist. Je nach Fachrichtung, Denkweise und Vorwissen der Benutzer können die Antworten verschieden ausfallen. Begriffe wie „Eigenschaft" und „Art" werden unterschiedlich interpretiert und dadurch fällt es den Benutzern zum Teil schwer, eine Antwort zu finden. Wenn erst nach der Objektklasse gefragt würde, könnten manche Benutzer leichter eine Eigenschaft benennen.

Die Erwartungshaltung von den Testpersonen an den Roboter und den Dialog waren trotz der intensiven Erklärung, immer noch hoch. Das fällt auf, wenn man einzelne Antworten untersucht, welche die Benutzer gegeben haben. Die Art der Abstraktion von einem tatsächlich existierenden, realen Objekt, das vor dem Roboter auf dem Tisch liegt zu der Beschreibung, die der Benutzer dem Roboter beibringen möchte, erfordert in manchen Fällen mehr Hintergrundwissen, als der hier vorgestellte Ansatz abbildet. Einmal wurde zum Beispiel ein Orangensaft neu gelernt. Der Roboter hätte diesen als Saft mit dem Geschmack „orange", als Getränk mit dem Namen des Herstellers, oder als teures oder leckeres Getränk lernen können. Der Roboter hätte daraufhin von sich aus versucht, die Art des Objektes näher einzuschränken, bis ein Saft eingelernt werden kann. Mit der gegebenen Erklärung, die dem Objekt die Eigenschaft „flüssig" und die Objektklasse „Orange" zuweist, konnte das System so nicht direkt umgehen. Für Menschen ist es offensichtlich, dass eine flüssige Orange ein Orangensaft ist, aber ob dieses Ergebnis für den Roboter als ein erfolgreiches Lernen betrachtet werden kann, bleibt fraglich. Für die Ergebnisse dieser Arbeit wurde dieser Fall als Fehler bewertet. Ein anderes Beispiel trat auf, als ein „english breakfast tea" als „red box" gelernt wurde, weil der Tee in einer roten Schachtel aufbewahrt wurde. Trotzdem bezeichneten die beiden Testpersonen ihre Objekte später, im Gespräch als „Orangensaft" und als „Schwarztee". Der Roboter hätte also in einem zweiten Schritt noch lernen müssen, dass der nun referenzierte „Schwarztee" die „rote Schachtel" ist. Diese übergenaue Beschreibung und die falschen Erwartungen kommen aus einer Unsicherheit, was Roboter tatsächlich können. Eine Vielzahl an Büchern und Filmen hat das Roboterbild so stark geprägt, dass Menschen nicht wissen, was ein Roboter tatsächlich können kann und wie man mit einem interagiert. Um diese Unsicherheiten zu überwinden kann es zum Beispiel sinnvoll sein, eine Klassifikation von Objekten über eine graphische Lösung machen zu lassen, in der der Benutzer die Idee der Entwickler und seine eigenen Möglichkeiten im Überblick hat. Eine zweite Möglichkeit wäre, dass während der Entwicklung bereits viel Arbeit in eine allgemeinere Ontologie gesteckt wird, oder dass sich die Besitzer von Heimrobotern jeweils Updates von den neuesten Produkten für die Wissensbasis ihres Roboters aus dem Internet herunterladen können. Eine einfache Herangehensweise, wie „Rede mit dem Roboter wie mit einem anderen Menschen" erklärt die Arbeitsweise und Funktionalität nicht genau genug.

6. Zusammenfassung und Ausblick

Die vorliegende Arbeit beschäftigt sich mit dem Lernen von Objekten und deren Bedeutung. Ein humanoider Roboter wird dazu befähigt, reale Gegenstände zu erkennen und führt einen Dialog mit dem Benutzer, um bekannte Objekte zu servieren oder unbekannte neu zu lernen. Das Lernen von Objekten spaltet sich auf in Aspekte des Lernens von visuellen Informationen, neuen Wörtern für Spracherkenner und der Bedeutung von Objekten. Zu jedem Teilbereich werden aktuelle Forschungsprojekte vorgestellt. In der vorliegenden Arbeit wird ein System beschrieben, das alle drei Teilbereiche in einem Gesamtsystem integriert. Damit lernt der Roboter visuelle und semantische Informationen über neue Objekte.

Es wird ein Test durchgeführt, bei dem ermittelt wird, wie Menschen Objekte wahrnehmen und beschreiben, um ein Objektmodell und eine Ontologie zu erstellen, die in dem Dialog verwendet werden, um Objekte zu verwalten. Bekannte Objekte werden in einem dynamischen Objektmodell verwaltet und von einer Ontologie interpretiert. Dabei können in dem Objektmodell verschiedene Eigenschaften angegeben werden und die Ontologie umfasst mehrere Ebenen von Objektklassifikationen, die hierarchisch Gegenstände immer spezieller beschreibt, wobei Klassen von ihren Elternklassen alle Eigenschaften erben. Mehrfachvererbung ist möglich. Dabei werden auch Funktionalitäten von Objekten modelliert, indem Objektklassen eine Funktionalität erben. Der Ansatz, Eigenschaften und Objektklassen zu lernen wird als möglicher Ansatz validiert.

Um das Dialogsystem und die Einzelkomponenten zu bewerten, werden Tests durchgeführt, bei denen nicht speziell trainierte Benutzer mit dem System interagieren und Objekte verwenden. Dabei werden neue Objekte gelernt. Ein Algorithmus zum Lernen von semantischen Klassen wird erklärt, der die Ontologie nach einer initialen Benutzerklassifikation nach dem passenden Typ durchsucht, indem Klassen weiter spezifiziert werden. Obwohl dieser Algorithmus mehr an Interaktionen benötigt, ist er besser als ein reiner One-Shot-Learning Algorithmus, weil er weniger Fehler bei der Klassifikation macht und die gelernten Klassen spezieller, also genauer klassifiziert werden. Das gesamte Dialogsystem ist in der Lage, in einem natürlichen Dialog mit unerfahrenen Nutzern neue Objekte zu lernen und erfolgreich wieder zu finden. Fehler,

wie falsch erkannte Objekte oder falsch erkannte Eigenschaften oder Klassen beim Lernen können im Dialog erfolgreich aufgelöst werden, wodurch die Gesamtperformanz des Systems gesteigert wird.

6.1. Ausblick

Das vorgestellte System ist in der Lage, Objekte zu verstehen und diese zu lernen. Bevor der Roboter die Objekte auch greifen und manipulieren kann, ist es nötig, weitere Informationen über das Objekt zu sammeln. Dafür ist ein verstärkter Einsatz an Sensoren notwendig, um zum Beispiel die Umrisse genauer wahrzunehmen, damit der Roboter das Objekt auch wirklich greifen kann.

Zum anderen ist es sinnvoll, die bisher global wirkenden Eigenschaften von Objekten zu beschränken. So kann es vorkommen, dass sich Eigenschaften über die Zeit oder in verschiedenen Situationen ändern. Eine Tasse kann zur Kaffeetasse oder Teetasse werden, wenn diese mit Kaffee oder Tee gefüllt ist, aber nachdem sie wieder geleert wurde, verliert sie diese Eigenschaft wieder. Auch der Füllstand und der sonstige Zustand ändern sich in verschiedenen Situationen. In dem bisherigen System können mehrere Eigenschaften wie „voll", „Kaffeetasse", „Teetasse" und „schmutzig" angegeben und parallel gespeichert werden, aber diese bleiben global gültig. Diese Modellierung erfordert zu dem bisherigen Objektmodell, das keinen temporären Verlauf beinhaltet, ein weiteres Zustandsmodell, in dem festgehalten wird, wie sich das Objekt verändert. Darin kann zum Beispiel der Füllstand, die Temperatur oder der Frischegrad modelliert werden. Durch das zeitliche Begrenzen könnte eine Eigenschaft später wieder deaktiviert werden. Auch verschiedene Benutzer können verschiedene Eigenschaften angeben, die nur für sie persönlich gültig sind. So kann zum Beispiel eine „Lieblingstasse" gelernt werden. Es würde keinen Sinn machen, diese Tasse nun als „Lieblingstasse" in die Ontologie einzutragen, da diese Referenz eine persönliche Beziehung dieses speziellen Probanden zu diesem Objekt darstellt. Da der Roboter aber für möglichst viele Benutzer allgemein die Objekte lernen soll, wäre eine Erweiterung des Objektmodells um ein Benutzermodell erforderlich, bei dem solche Beschreibungen abgebildet werden können. Deshalb ist es sinnvoll, verschiedene Modelle, wie Situations- und Benutzermodell mit dem Objektmodell zu verknüpfen. Während den Tests kam es zu folgenden Situationen, in denen eine Erweiterung sinnvoll ist.

Eine genauere Modellierung der möglichen Eigenschaften ist für Menschen in einem Dialog schwer. Wie schon die Frage „Was kann man mit dem Objekt machen?" in Kapitel 4.2 gezeigt hat, fällt es Benutzern schwer, die Frage nach Art und Zweck eines Objektes zu beschreiben. Durch das Angeben von bestimmten Eigenschaften für jede Objektklasse könnten aber leichter Informationen über ein Objekt gesammelt werden. So könnte der Roboter zum Beispiel direkt nach dem Autor und dem Titel, oder nach der ISBN Nummer fragen, wenn der Benutzer ein Buch referenziert. Die möglichen Eigenschaften einer neuen Klasse in einem Dialog von dem Benutzer neu zu lernen ist aber nicht sinnvoll, da die Beschreibung schwer zu vermitteln ist. Hierfür wäre ein Ansatz, wie ihn Regine Becher, vorgestellt in Kapitel 3.3.4, verfolgt, sinnvoller, als dies in einem Dialog mit dem Benutzer zu klären. Ein Benutzer kann mit Hilfe von

herkömmlichen Modalitäten mögliche Attribute und Eigenschaften für die neue Klasse auswählen, ohne diese frei formulieren zu müssen.

Die Einordnung von Objekten in Klassen kann in unterschiedlichen Abstraktionsebenen geschehen. Intuitiv ist eine speziellere Einordnung besser, weil dieses spezielle Konzept eine Unterklasse aller Konzepte darstellt, von denen es erbt und diese damit einschließt. Durch weitere Tests kann untersucht werden, wie viel eine genauere Einordnung durch den vorgestellten Algorithmus bringt. Dadurch kann die optimale Tiefe der verwendeten Ontologie bestimmt werden.

7. Literaturverzeichnis

Arsenic, A. (2004). Developmental learning on a humanoid robot. *Proceedings 2004 IEEE International Joint Conference on Neural Networks* (S. 3167 - 3172). Budapest, Hungary: IEEE Operations Center.

Asfour, T., Regenstein, K., Azad, P., Schröder, J., Bierbaum, A., Vahrenkamp, N., et al. (2006). ARMAR-III: An Integrated Humanoid Platform for Sensory-Motor Control. *Proceedings IEEE-RAS International Conference on Humanoid Robots.* Genova, Italy: IEEE Service Center.

Azad, P. (01. 12 2007). *Integration Vision Toolkit.* Von http://ivt.sourceforge.net. abgerufen

Azad, P., Asfour, T., & Dillmann, R. (2007). Stereo-based 6D Object Localization for Grasping with Humanoid Robot Systems. *Proceedings International Conference on Intelligent Robots and Systems (IROS).* San Diego, USA.

Becher, R., Boesnach, I., Steinhaus, P., & Dillmann, R. (2006). From Subject to Objects and back - Combining Human Motions and Object Properties to Unterstand User Actions. *Proceedings 2nd International Workshop on Human-Centered Robotic Systems.* München: Universitätsverlag.

Becher, R., Kasper, A., Steinhaus, P., & Dillmann, R. (2007). Developing and Analyzing Intuitive Modes for Interactive Object Modeling. *Proceedings Ninth International Conference on Multimodal Interfaces (ICMI 07)* (S. 74-81). Nagoya, Japan: Association for Computer Machinery.

Becher, R., Steinhaus, P., Zöllner, R., & Dillmann, R. (2006). Design and Implementation of an Interactive Object Modelling System. *Proceedings International Symposium on Robotics (ISR 2006)* (S. 22-27). Düsseldorf: VDI-Verlag.

Bekel, H., Bax, I., Heidemann, G., & Ritter, H. (2004). Adaptive computer vision: Online learning for object recognition. *Proceedings Deutsche Arbeitsgemeinschaft für Mustererkennung (DAGM)*, (S. 447–454). Tuebingen.

Carbonell, J. G. (1979). Towards a Self-Extending Parser. *Meeting of the Association for Computational Linguistics (ACL).*

Carpenter, B. (1992). *The Logic of Typed Feature Structures.* Cambridge, England: Cambridge University Press.

Deutsche Forschungsgemeinschaft. (Juni 2001). *SFB 588 Humanoide Roboter.* Von http://www.sfb588.uni-karlsruhe.de/. abgerufen

Dusan, S., & Flanagan, J. (2003). A System for Multimodal Dialogue and Language Acquisition. *Proceedings 2nd Romanian Academy Conference on Speech Technology and Human-Computer Dialogue.* Bucharest, Romania: Romanian Academy.

Dusan, S., & Flanagan, J. (2002). Adaptive Dialog Based upon Multimodal Language Acquisition. *Proceedings Fourth IEEE International Conference on Multimodal Interfaces (ICMI 02).*

Dusan, S., & Flanagan, J. (2002). An Adaptive Dialogue System Using Multimodal Language Acquisition. *Proceedings International CLASS Workshop: Natural, Intelligent and Effective Interaction in Multimodal Dialogue Systems.*

Finke, M. (1997). Recognition Of Conversational Telephone Speech Using The Janus Speech Engine. *International Conference on Acoustics, Speech, and Signal Processing (ICASSP '97).* Munich, Deutschland.

Garcia, L.-M., Oliveira, A. A., Grupen, R. A., Wheeler, D. S., & Fagg, A. H. (2000). Tracing Patterns and Attention: Humanoid Robot Cognition. *IEEE Intelligent Systems* , S. 70-77.

Gruber, T. R. (1993). A Translation Approach to Portable Ontology Specifications. In *Knowledge Acquisition 5(2)* (S. 199-220). Academic Press.

Holzapfel, H. (2005). Towards Development of Multilingual Spoken Dialog Systems. *Proceedings 2nd Language and Technology Conference (L&T'05).* Poznan.

Holzapfel, H., & Prommer, T. (2006). Rapid Simulation-Driven Reinforcement Learning of Multimodal Dialog Strategies for Human-Robot Interaction. In T. Prommer (Hrsg.), *Proceedings IEEE/RSJ International Conference on Intelligent Robots and Systems.* Beijing, China: IEEE Press.

Honda Robots. (01. 12 2007). *Honda Robots.* Von http://www.honda-robots.com abgerufen

Kirstein, S., Wersing, H., & Körner, E. (2005). Rapid Online Learning of Objects in a Biologically Motivated Recognition Architecture. *Deutsche Arbeitsgemeinschaft für Mustererkennung (DAGM).* DAGM-Symposium.

Lömker, F. (2004). Lernen von Objektbenennungen mit visuellen Prozessen. *Dissertation* . Bielefeld: Universität Bielefeld.

Lowe, D. G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision, 60, 2* , 91-110.

Ogden, C. K., & Richards, I. A. (1923). *The meaning of meaning.* London: Kegan Paul, Trench, Trubner & Co.

Roy, D. (1999). Learning words from sights and sounds: A computational model. *Dissertation* . Massachusetts Institute of Technology.

Roy, D., & Mavridis, N. (2006). Grounded Situation Models for Robots: Where words and percepts meet. *Proceedings IEEE/RSJ International Conference on Intelligent Robots and Systems.* Beijing, China: IEEE Press.

Roy, D., Mavridis, N., Levit, M., & Gorniak, P. (2006). The Human Speechome Project. *Proceedings 28th Annual Meeting of the Cognitive Science Society.* Vancouver.

Schaaf, T. (2004). Erkennen und Lernen neuer Wörter. *Dissertation* . Karlsruhe: Universität Karlsruhe (TH).

Schulz, B. (2005). Lernen neuer Wörter im Dialog. *Diplomarbeit* . Karlsruhe: Universität Karlsruhe (TH).

Soltau, H. (2001). A One-Pass Decoder Based on Polymorphic Linguistic Context Assignment. *Proceedings Automation Speech Recognition and Understanding Workshop, ASRU-01.* Trento, Italy: Madonna di Campiglio.

Talamazzini, E.-G. S. (2001). *Automatische Spracherkennung. Grundlagen, statistische Modelle und effiziente Algorithmen.* Vieweg.

Waibel, A. (1990). *Readings in Speech Recognition.* Morgan Kaufmann.

Wersing, H., Kirstein, S., Goetting, M., Brandl, H., Dunn, M., Mikhailova, I., et al. (2007). Online Learning of Objects and Faces in an Integrated Biologically Motivated Architecture. *Proceedings 5th International Conference on Computer Vision Systems (ICVS).* Bielefeld.

Wersing, H., Kirstein, S., Götting, M., Brandl, H., Dunn, M., Mikhailova, I., et al. (2006). A Biologically Motivated System for Unconstrained Online Learning of Visual Objects. *Proceedings International Conference Artificial Neural Networks (ICANN 2006)*, (S. 508-517). Athens, Greece.