

Kapitel 3

Die Struktur multimedialer Präsentationen

Als erster Schritt im Hinblick auf die automatische Erzeugung von Multimedia-Präsentationen wird im folgenden untersucht, welchen Strukturierungsprinzipien solche Präsentationen unterliegen.

3.1 Die Generierung multimedialer Präsentationen als zielgerichtete Tätigkeit

Kerngedanke der vorliegenden Arbeit ist die Verallgemeinerung einer sprechakttheoretischen Sichtweise auf Multimedia-Präsentationen. Nicht nur sprachliche Äußerungen, sondern allgemein die Präsentation von Multimedia-Dokumenten soll als kommunikative Handlung zur Erreichung von Zielen aufgefaßt werden.

3.1.1 Kommunikative Ziele und Präsentationshandlungen

Theorien über menschliches Handeln fußen auf der Annahme, daß Menschen nicht ohne Grund Handlungen ausführen. Die treibende Kraft hinter einer Handlung ist vielmehr der Wunsch, durch sie ein bestimmtes Ziel zu erreichen. Übertragen auf die Erstellung von Präsentationen folgt hieraus, daß ein Autor zunächst einmal ein oder mehrere Ziele haben muß, die ihn dann dazu veranlassen, beispielsweise Text zu formulieren, eine Graphik zu zeichnen oder eine Photographie in eine Präsentation einzubinden.

Wird eine sprechakttheoretische Sichtweise eingenommen, so ist zu berücksichtigen, daß in der Sprechakttheorie vor allem die Erzeugung gesprochener Sprache behandelt wurde. Nach [Searle, 80] geht beim Sprechen das Erzeugen von Lauten mit dem Vollbringen eines propositionalen, eines illokutiven und möglicherweise eines perlokutiven Aktes einher (vgl. Abschnitt 2.3.1.1). Bei einer multimedialen

Präsentation sind die mit einer kommunikativen Handlung assoziierten Akte jedoch häufig zeitlich versetzt und werden unter Umständen nicht einmal von ein und demselben Autor durchgeführt. D.h. eine kommunikative Handlung kann auch dadurch vollzogen werden, daß vorhandene Präsentationsteile in anderen Zusammenhängen neu verwendet werden. Insgesamt sind drei Fälle zu unterscheiden:

- Herstellung und Verwendung einer Präsentation sind als Einheit aufzufassen. Diese Situation liegt beispielsweise dann vor, wenn Graphik vor den Augen eines Adressaten angefertigt und kommentiert wird.
- Herstellung und Verwendung einer Präsentation werden von derselben Person bewerkstelligt, jedoch liegen Herstellung und Verwendung zeitlich auseinander. Dies ist beispielsweise dann der Fall, wenn jemand Material für einen späteren Vortrag vorbereitet.
- Herstellung und Verwendung einer Präsentation erfolgen durch unterschiedliche Autoren. Diese Situation tritt beispielsweise dann auf, wenn auf vorgefertigte Präsentationsteile, z.B. aus einer Multimedia-Datenbank, zurückgegriffen wird.

In den letzten beiden Fällen können sich die Ziele, die der Herstellung einer Präsentation zugrundeliegen, erheblich von den Zielen unterscheiden, die durch die Verwendung der Präsentation erreicht werden sollen. Beispielsweise kann ein Bild, das angefertigt wurde, um jemandem zu zeigen, wie eine bestimmte Espressomaschine aussieht, in einem anderen Zusammenhang dazu verwendet werden, um jemandem von dieser Maschine abzuraten. Schließlich ist anzunehmen, daß auch durch Zufall entstandene Bilder, bei deren Erzeugung keine Absicht vorlag, unter Umständen dazu eingesetzt werden können, um ganz konkrete Ziele zu erreichen.

“Protogenes soll sich vergeblich bemüht haben, den Schaum vor dem Munde eines keuchenden Hundes darzustellen; ärgerlich warf er den Schwamm nach dem Bilde - da schuf der Schwamm den gewünschten Effekt.” (vgl. [Kris & Kurz, 34], S. 71)

3.1.2 Die Unterscheidung zwischen Haupt- und Nebenhandlungen

In vielen Fällen sind zur Realisierung eines Ziels mehrere Handlungen auszuführen. Im folgenden wird zwischen zwei Arten von kommunikativen Handlungen unterschieden. Handlungen, die die unmittelbaren Ziele des Präsentators zum Ausdruck bringen, werden als *Haupthandlungen* (HH) bezeichnet. Handlungen, die zur Unterstützung der Haupthandlung dienen, heißen hingegen *Nebenhandlungen* (NH).

Im wesentlichen erfüllen sie drei Funktionen: Sie verstärken die Wirkung der Haupthandlungen, sorgen dafür, daß notwendige Bedingungen für die erfolgreiche Ausführung der Haupthandlungen erfüllt sind und stellen sicher, daß die Wirkung der Haupthandlungen erreicht wurde. In den meisten Fällen wird eine einzige Haupthandlung durch ein oder mehrere Nebenhandlungen unterstützt. Es kann jedoch auch vorkommen, daß zur Erreichung eines Ziels mehrere Haupthandlungen notwendig sind oder die Haupthandlung wegfällt.

Haupt- und Nebenhandlungen entsprechen im wesentlichen Searles [Searle, 80] *global* und *subsidiary acts*, van Dijks [van Dijk, 80] *main* und *subordinate speech acts*, Brandts [Brandt et al., 83] *dominierenden* und *subsidiären Handlungen* sowie Manns und Thompsons [Mann & Thompson, 87b] *nuclei* und *satellites*.

3.1.3 Strukturierungsprinzipien

Die in den vorangegangenen Abschnitten angestellten Überlegungen legen nahe, multimediale Präsentationen wie reine Texte unter drei Aspekten zu strukturieren.

3.1.3.1 Rhetorische Struktur

Läßt man zu, daß sich Handlungen wiederum aus elementarerer Handlungen zusammensetzen können, entsteht eine hierarchische Handlungsstruktur. Diese Struktur wird in der Textlinguistik auch als *rhetorische Struktur* bezeichnet. Bei der Wurzel der rhetorischen Struktur handelt es sich meist um komplexe kommunikative Handlungen, wie z.B. die Einführung eines Objekts, die Motivation einer Handlung oder die Beschreibung einer Handlungssequenz. Die Blätter sind elementare Sprechakte (vgl. [Searle, 80]) oder Bildakte (vgl. [Kjørup, 78]). Beispiele für Sprechakte sind *Behaupten*, *Auffordern* oder *Bitten*, während Handlungen wie *Annotieren*, *Einrahmen* oder *Hervorheben* zu den Bildakten zählen.

Eine kommunikative Handlung wird durch ihre kommunikative Funktion, den Präsentator, den Adressaten, ihren Inhalt und das verwendete Medium bestimmt. Die kommunikative Funktion spiegelt wider, wie eine Handlung zu verstehen ist. Beispiele für kommunikative Funktionen sind: *Einführen*, *Motivieren* oder *Beschreiben*. Um einen direkten Vergleich mit den in der englischsprachigen Fachliteratur referierten Funktionen zu erleichtern, werden die Bezeichner kommunikativer Funktionen im folgenden an die entsprechenden englischsprachigen Bezeichner angelehnt. Bei dem Inhalt einer kommunikativen Handlung kann es sich um Objekte oder Relationen handeln. Beispielsweise führt die Ausführung der kommunikativen Handlung

(Describe P A (Sequence act-sequence-1) Graphics)

zum Aufbau einer Bildfolge, in der mehrere aufeinanderfolgende Ereignisse dargestellt sind.

Bei der Analyse von Präsentationen zeigt sich, daß mit bestimmten Dokumententypen charakteristische Handlungsmuster assoziiert werden können. Beispielsweise setzen sich Bedienungsanleitungen oft aus einer Produktbeschreibung und einer Instruktionsfolge zusammen. Bei der Produktbeschreibung handelt es sich üblicherweise um eine Auflistung der einzelnen Teile eines Produkts (meist in Verbindung mit einer Abbildung des Produkts). Die Instruktionen bestehen in den meisten Fällen aus einer Aufforderung an den Benutzer, eine bestimmte Handlung auszuführen. Darüberhinaus wird dem Benutzer Information geliefert, die ihn dazu befähigen soll, der Aufforderung nachzukommen.

Obwohl Sprache das am weitesten verbreitete Kommunikationsmittel darstellt, besteht kein Zweifel, daß sich viele kommunikative Handlungen auch mit Bildern vollziehen lassen. Beispielsweise wird in den meisten Bedienungsanleitungen ein Produkt dadurch eingeführt, daß ein Bild von ihm gezeigt wird und seine Teile mit den entsprechenden Begriffen im Bild beschriftet werden. Auch Information darüber, wie ein Gerät zu bedienen ist, wird häufig durch Bilder wiedergeben. Eine genauere Untersuchung darüber, für welche kommunikativen Handlungen sich Bilder eignen, wird in Kapitel 4 durchgeführt.

3.1.3.2 Intentionale Struktur

Durch die mit den einzelnen Handlungen assoziierten Ziele ergibt sich neben einer hierarchischen Handlungsstruktur auch eine hierarchische Struktur von Zielen. Dabei tragen die Handlungen, die zur Erreichung eines untergeordneten Ziels dienen, zur Erreichung übergeordneter Ziele bei. Die hierarchische Zielstruktur wird auch *intentionale Struktur* genannt (vgl. [Grosz & Sidner, 86]).

Kommunikative Ziele beziehen sich auf den mentalen Zustand eines Adressaten. Auch wenn durch eine kommunikative Handlung letztendlich eine Veränderung in der Welt, z.B. das Öffnen eines Fensters, herbeigeführt wird, so kann diese Änderung nur dadurch zustandekommen, daß zunächst einmal auf den mentalen Zustand des Adressaten eingewirkt wird. Beispiele für kommunikative Ziele sind:

- der Benutzer soll wissen, aus welchen Teilen sich ein bestimmtes Gerät zusammensetzt.
- der Benutzer soll in der Lage sein, eine bestimmte Aktion auszuführen
- der Benutzer soll die Durchführung einer bestimmten Handlung wünschen
- der Benutzer soll wissen, daß nach Ausführung einer Aktion ein bestimmter Zustand eingetreten ist.

Um solche Ziele zu repräsentieren, wird eine ähnliche Notation wie in dem von Hovy entwickelten RST-Planer verwendet (vgl. 2.3.2.3):

- (*Bel User p*) steht für: *User* glaubt, daß *p* erfüllt ist.
- (*Done User action*) bedeutet, daß *User* die Aktion *action* ausgeführt hat.
- (*Happened event*) steht für: das Ereignis *event* fand statt.
- Durch (*Goal User p*) wird ausgedrückt: *User* hat *p* zum Ziel.
- (*Cando User action*) heißt: *User* ist in der Lage, *action* auszuführen.
- (*After action p*) steht für: Nach Ausführung von *action* gilt *p*.
- (*Before action p*) bedeutet: Vor Ausführung von *action* gilt *p*.
- (*BMB System User x*) ist eine Abkürzung für die unendliche Konjunktion: (*And (Bel System x) (Bel System (Bel User x)) (Bel System (Bel User (Bel System x))) ...*).

Anzumerken ist, daß die intentionale Struktur einer Präsentation nicht in dem Sinne vollständig sein muß, daß alle untergeordneten Ziele ein übergeordnetes Ziel implizieren. Vielmehr liegt die Annahme zugrunde, daß die zur Realisierung untergeordneter Ziele vermittelten Inhalte in der jeweiligen Präsentationssituation zur Realisierung des übergeordneten Ziels ausreichen. Beispielsweise kann mit den einzelnen Bildern einer Bildfolge das Ziel assoziiert werden, daß der Benutzer weiß, daß die in den Bildern gezeigten Aktionen auszuführen sind. So soll zwar mit der Bildfolge als Ganzes erreicht werden, daß der Benutzer die einzelnen Schritte in der richtigen Reihenfolge ausführt. Dieses Ziel ergibt sich jedoch nicht zwangsläufig aus den mit den Einzelbildern assoziierten Zielen, da diese zwar Information über die einzelnen Aktionen, aber nicht über deren Beziehungen zueinander vermitteln.

Desweiteren sei darauf hingewiesen, daß zwischen Handlungen und Zielen nicht notwendigerweise eine 1:1-Beziehung besteht. So kann das Ziel "der Benutzer soll wissen, wo sich ein bestimmtes Geräteteil befindet" dadurch erreicht werden, daß man ein Hintergrundobjekt darstellt oder zusätzliche Merkmale des zu lokalisierenden Teils beschreibt. Umgekehrt können durch ein und dieselbe kommunikative Handlung (z.B. ein Hintergrundobjekt darstellen) unterschiedliche Ziele erreicht werden (z.B. der Benutzer soll wissen, wo sich ein bestimmtes Geräteteil befindet, oder der Benutzer soll wissen, in welcher Situation ein bestimmtes Teil zu verwenden ist).

3.1.3.3 Attentionale Struktur

Die *attentionale Struktur* wird durch Repräsentationen der Entitäten (Objekte und Relationen) bestimmt, auf die der Benutzer während der Präsentation seine Aufmerksamkeit richtet (vgl. [Grosz & Sidner, 86]). Sie setzt sich aus einer Menge von hierarchisch angeordneten Fokusräumen zusammen, wobei jeder der Fokusräume zu einem kommunikativen Ziel korrespondiert.

Ein wesentlicher Unterschied zwischen Texten und Bildern besteht darin, daß ein Leser weitgehend den organisatorischen Aufbau eines Textes akzeptieren muß und diesen in der Regel sequentiell verarbeitet. Im Gegensatz dazu präsentiert sich ein Bild dem Benutzer simultan und überläßt diesem weitgehend die Verarbeitung, wenn man von einigen Konventionalisierungen (etwa die Bevorzugung von links/rechts sowie oben/unten) einmal absieht. Dennoch läßt sich auch Bildern eine attentionale Struktur zuordnen. Beispielsweise deuten Blickbewegungsuntersuchungen (vgl. u.a. [Loftus & Bell, 75]) darauf hin, daß Bilder in einer Weise wahrgenommen werden, die dem hierarchischen Aufbau des Informationsgehaltes entspricht.

Eine Besonderheit von graphischen Darstellungen besteht darin, daß die abgebildeten Objekte nur indirekt über ihren graphischen Repräsentanten in den Fokus gelangen. Wird beispielsweise ein Schalter einer Espressomaschine abgebildet, so werden sowohl das Abbild des Schalters als auch der Schalter selbst fokussiert.¹

Die Verwendung von Fokusräumen erweist sich nicht nur bei der Analyse und Generierung von Texten, sondern auch bei der Verarbeitung von Multimedia-Präsentationen als nützlich. Ein Beispiel hierfür ist die Generierung von Anaphern und multimedialen Referenzen. So ist bei der Generierung von Anaphern zu beachten, daß ein Benutzer als Antezedenten bevorzugt Objekte aus Fokusräumen in Betracht zieht, deren Abstand zum aktuellen Fokusraum in der Hierarchie möglichst gering ist. Bemerkt sei, daß nicht alle Objekte innerhalb eines bestimmten Fokusraumes den gleichen Fokussierungsgrad aufweisen. Wie bei reinem Text kommen in erster Linie besonders stark fokussierte Objekte eines Fokusraumes als Referenten für Anaphern in Frage.

3.1.4 Durchführung von Strukturanalysen

In folgenden werden einige Probleme bei der Durchführung einer Strukturanalyse skizziert und ihre Konsequenzen für den Aufbau von Präsentationen aufgezeigt.

3.1.4.1 Aufteilung in Basiseinheiten

Um eine Strukturanalyse durchführen zu können, muß eine Präsentation zunächst in Basiseinheiten aufgeteilt werden. Dies sind Einheiten, deren zugrundeliegende Handlungen/Ziele nicht mehr weiter zerlegbar sind.

In Arbeiten zur Textlinguistik handelt es sich hierbei meistens um vollständige (Teil-)sätze. Allerdings vertritt Rösner [Rösner & Stede, 92] die Auffassung, daß auch kleinere Einheiten wie z.B. Präpositionalphrasen als Basiseinheiten zu betrachten sind. Betrachtet man beispielsweise den nachfolgenden deutschen Satz und seine Übersetzung ins Englische, so wird man feststellen, daß ein und dieselbe Information

¹Ein ähnlicher Fall liegt übrigens vor, wenn der Fokus aufgrund von begrifflichem Wissen ausgeweitet wird. So werden bei Äußerungen der Form "Der Wagen parkt falsch. Seine Vorderräder versperren den Weg." über den Wagen auch die Teile des Wagens in den Fokus gebracht.

im Englischen durch einen Nebensatz, im Deutschen jedoch durch eine Präpositionalphrase wiedergegeben wird. Eine Strukturanalyse sollte jedoch in beiden Fällen zum gleichen Ergebnis führen.

“Check the coolant level in the reserve tank *when the engine is at normal operating temperature.*”

“Den Kühlmittelstand im Reservetank *bei normaler Betriebstemperatur* des Motors kontrollieren.”

Um Basiseinheiten zu bestimmen, schlägt Rösner zwei Tests vor: den *Paraphrasentest*, bei dem Teile eines Textes durch einfachere Teile ersetzt werden, und den *Übersetzungstest*, bei dem Teile eines Textes in andere Sprachen übersetzt werden.

Bei der Strukturanalyse einer multimedialen Präsentation stellt sich nun die Frage, welche Teile der Präsentation als elementare Einheiten anzusehen sind. Die bloße Aufteilung in Bilder und Textblöcke, wie sie beispielsweise in der Text-Bild-Forschung vorgenommen wird, ist sicherlich ungeeignet. Einerseits ermöglicht sie nur eine grobe Analyse der Präsentationsstruktur, da nur die mit ganzen Bildern assoziierten Handlungen und Ziele betrachtet werden. Andererseits kann sich eine Basiseinheit, z.B. bei multimedialen Referenzausdrücken, auch auf mehrere Medien erstrecken.

Anhand eines einfachen Beispiels kann man sich klarmachen, daß auch einzelne Bildteile als Basiseinheiten fungieren können. Beispielsweise wird bei der Gebrauchsanleitung für die Ciatronic Espressomaschine KA 2055 das Hochheben des Klappdeckels und das Abschrauben des Boilerdeckels durch drei Pfeile in einem Bild dargestellt. Diese drei Pfeile könnten jedoch auch auf zwei oder sogar drei Bilder aufgeteilt werden. Für die Verwendung von zwei Bildern hat man sich beispielsweise bei der Gebrauchsanleitung für die Philips Espressomaschine HD 5649 entschieden (vgl. Abbildung 3.1).

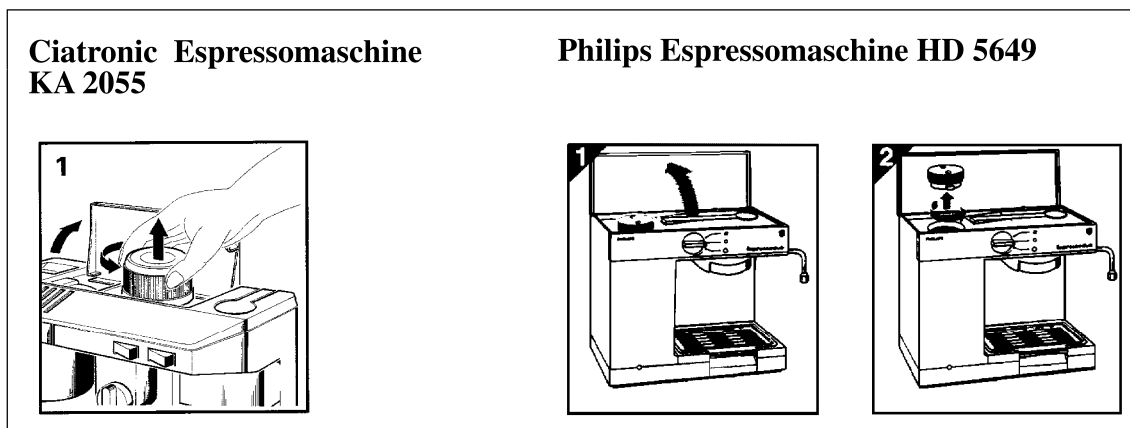


Abbildung 3.1: Unterschiedliche Aufteilung von Information auf mehrere Bilder

In allen drei Fällen würde eine Sequenz von drei Aktionen dargestellt. Läßt man als Basiseinheiten nur ganze Bilder zu, so würde Information bei einer Strukturanalyse in den ersten beiden Fällen verlorengehen.

Basiseinheiten eines Bildes können jedoch nicht nur durch “Paravisualisierung”, sondern auch durch Ersetzung von Bildteilen durch Text gefunden werden. So ließe sich die durch die drei Pfeile wiedergegebene Information auch durch drei verbale Aufforderungen wiedergeben: den Klappdeckel nach oben heben, den Boilerdeckel nach links drehen und gleichzeitig nach oben abheben.

Insgesamt läßt sich feststellen, daß nicht nur Texte, sondern auch Bilder eine dekompositionelle Struktur aufweisen. D.h. nicht nur mit Bildern als Ganzes, sondern auch mit einzelnen Bildteilen lassen sich bestimmte Ziele und kommunikative Handlungen assoziieren.

Beim Aufbau einer Präsentation ist folglich zunächst einmal zu klären, in welche Basiseinheiten ein komplexes kommunikatives Ziel oder eine komplexe kommunikative Handlung zerlegt werden kann. Da diese Basiseinheiten nicht unbedingt vollständigen Sätzen oder Bildern entsprechen, ist nach dem Dekompositionsprozeß ein Aggregationsprozeß durchzuführen, durch den festgelegt wird, welche Basiseinheiten im selben Satz bzw. im selben Bild unterzubringen sind.

3.1.4.2 Strukturmarkierungen

Ein weiteres Problem bei der Durchführung von Strukturanalysen besteht darin, daß diese oft nicht eindeutig sind. Beispielsweise geht aus einer Bildfolge nicht immer hervor, ob hiermit lediglich die zeitliche Aufeinanderfolge zweier Ereignisse beschrieben wird oder ob es primär um die Vermittlung einer Ursache-Wirkung-Beziehung geht.

Eine Hilfe bei der Analyse der rhetorischen Struktur stellen Markierungen an der Oberfläche dar. Dies gilt nicht nur für reine Texte, sondern auch für Bilder und Text-Bild-Kombinationen. So deuten nebeneinander stehende Bilder auf die Handlungen “Vergleiche” (Contrast) oder “Beschreibe-Sequenz” (Describe-Sequence) hin. Elaborationen werden häufig durch Nebenbilder markiert. Ein hoher Realitätsgrad (z.B. bei Photos in Zeitungsberichten) kann auf die Handlung “Belege-Fakt” (Provide-Evidence) hinweisen. Bemerkt sei, daß Bilder es nur indirekt gestatten, die kommunikative Funktion anzuzeigen. So kann man zwar mit Bildern vor etwas warnen; es ist jedoch nicht möglich, graphisch auszudrücken, daß eine Äußerung als Warnung aufzufassen ist. Mit sprachlichen Mitteln bereitet dies hingegen keine Probleme wie Äußerungen der Form “Dies ist eine Warnung.” zeigen.

Die intentionale Struktur läßt sich in den meisten Fällen aus der rhetorischen Struktur ableiten. Wird beispielsweise durch eine kommunikative Handlung Information über einen bestimmten Sachverhalt vermittelt, so soll damit erreicht werden, daß der Adressat diesen Sachverhalt glaubt. In einigen Fällen wird auch explizit auf das Ziel eines Präsentationsteils hingewiesen. Man denke an Äußerungen der Form:

“Ziel dieses Vortrags ist es, ...”.

Auch die attentionale Struktur wird häufig an der Oberfläche markiert. Beispielsweise deuten Bilder mit neuem Weltausschnitt in den meisten Fällen auf eine Fokusverschiebung hin. Der Übergang zu einem untergeordneten Fokusraum wird durch Nebenbilder oder durch weitere Bilder mit eingeschränktem Weltausschnitt zum Ausdruck gebracht. Darüberhinaus wird auch häufig von sprachlichen Mitteln Gebrauch gemacht, um wie in “Wenden wir uns nun der rechten Bildhälfte zu!” zu untergeordneten Fokusräumen überzugehen oder um wie in “Siehe auch Abbildung x!” zu früheren Fokusräumen zurückzukehren. Wie bereits festgestellt können innerhalb eines Fokusraumes Objekte unterschiedlich stark fokussiert sein. Dies äußert sich bei Texten beispielsweise durch die Stellung im Satz, bei Graphiken durch die Positionierung, Einfärbungen und Markierungen.

Techniken zur Markierung der Struktur sind umgekehrt beim Aufbau einer Präsentation gezielt einzusetzen, um dem Benutzer die Interpretation zu erleichtern. Darüberhinaus ist auch bei der Auswahl eines Mediums zu berücksichtigen, inwieweit dieses dazu in der Lage ist, bestimmte kommunikative Handlungen auszuführen oder Fokusverschiebungen zu markieren.

3.1.4.3 Mehrfachfunktionen

Die Tatsache, daß mit einem Präsentationsteil unter Umständen mehrere Ziele/Handlungen assoziiert werden können und daß dies auch vom Präsentator beabsichtigt ist, führt zu einem weiteren Problem bei der Strukturanalyse. Beispielsweise kann die Darstellung eines Geräts zum einen als Teil einer Beschriftung fungieren, die den Benutzer über den Gerätetyp informieren soll, zum anderen kann sie als Hintergrund für ein Geräteteil dem Benutzer die Identifizierung dieses Teils ermöglichen. Folglich handelt es sich bei der intentionalen/rhetorischen Struktur nicht immer um einen Baum im Sinne der Graphentheorie, bei dem jeder Sohn exakt einen Vater hat. Vielmehr treten auch gerichtete azyklische Graphen (engl. Directed Acyclic Graph, DAG) auf.

In [Maier, 93] wird das Problem der Mehrfachfunktionen darauf zurückgeführt, daß bei der Analyse der Diskursstruktur einer Präsentation unterschiedliche Ebenen zu berücksichtigen sind, wobei die *ideationale Ebene* das Sachgebiet, die *interpersonale Ebene* die Hörer-Sprecher-Beziehung und die *textuelle Ebene* die Organisation von Texten betrifft. Demnach können zwischen Präsentationsteilen gleichzeitig mehrere Beziehungen bestehen, die den unterschiedlichen Ebenen zuzuordnen sind. Das hier angeführte Beispiel verdeutlicht jedoch, daß bei Bildern auch auf einer einzigen Ebene (hier: der ideationalen Ebene) Mehrfachfunktionen möglich sind. Der Umstand, daß Bildteile häufig in unterschiedlichen Zusammenhängen wiederverwendet werden, läßt sich nicht zuletzt darauf zurückführen, daß sich die Bildherstellung im allgemeinen wesentlich aufwendiger als die Texterstellung gestaltet.

3.2 Beispielanalysen

Abbildung 3.2 zeigt einen aus einer Bedienungsanleitung entnommenen Dokumentausschnitt, in dem der Benutzer darüber informiert wird, wie der Wasserbehälter einer Espressomaschine zu füllen ist. Es handelt sich hierbei wie bei den meisten Beispieldokumenten dieser Arbeit um eine leicht modifizierte Form einer Gebrauchsanweisung für die Philips Espressomaschine HD 5649.

In den nächsten beiden Abschnitten werden die intentionale, die rhetorische und die attentionale Struktur dieses Dokumentausschnitts analysiert.

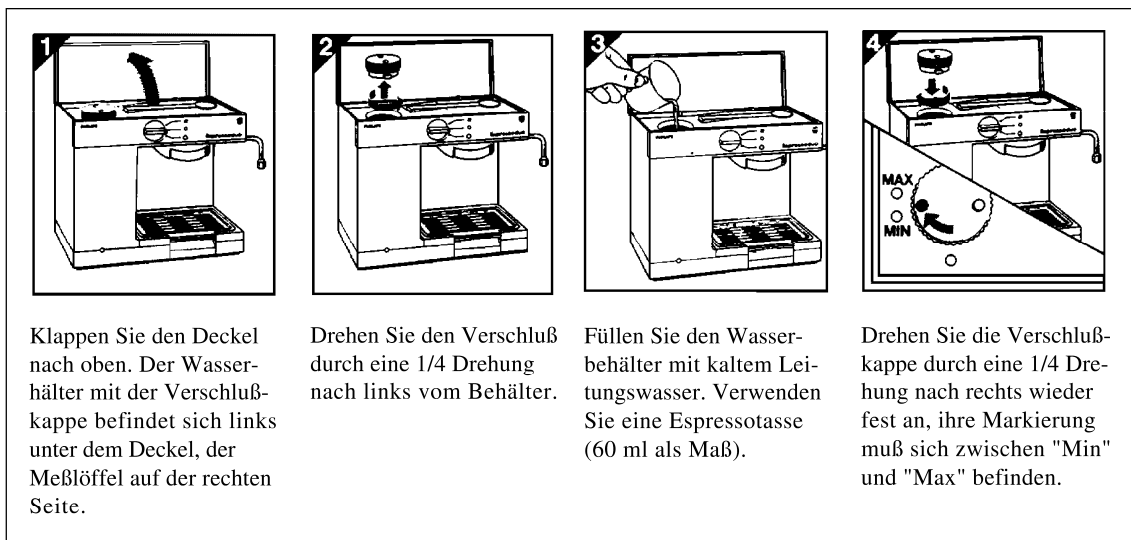


Abbildung 3.2: Dokumentausschnitt

3.2.1 Intentionale Struktur

Unter Verwendung der in Abschnitt 3.1.3.2 eingeführten Notation läßt sich die intentionale Struktur des in Abbildung 3.2 gezeigten Dokumentausschnitts wie in den Abbildungen 3.3 bis 3.7 dargestellt beschreiben.

Mit diesem Dokumentausschnitt wird das Ziel assoziiert, daß der Benutzer einen Plan zur Ausführung der Aktion *fill-container-1* kennt (vgl. Abbildung 3.3), der in diesem Fall als eine Sequenz von vier Teilaktionen, nämlich *lift-1*, *remove-1*, *pour-in-1* und *tighten-1* definiert wird. Das dem Gesamtdokument zugrundeliegende Ziel wird dadurch erreicht, daß dem Benutzer mitgeteilt wird, daß die vier Aktionen auszuführen sind. Das Dokument setzt sich somit aus vier Teilen zusammen, denen jeweils das Ziel (*Done User ...*) zugeordnet wird.

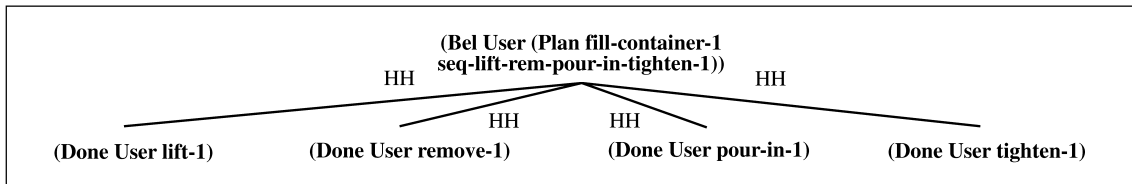


Abbildung 3.3: Füllen des Wasserbehälters - Intentionale Struktur

Zunächst wird der Benutzer gebeten, den Deckel der Espressomaschine nach oben zu klappen (vgl. Abbildung 3.4). Dies soll ihm zu verstehen geben, daß der Präsentator die Ausführung der Aktion *lift-1* wünscht. Mit der Bitte wird somit das Ziel *(Bel User (Goal System (Done User lift-1)))* verbunden. Der Benutzer wird jedoch nur dann der Aufforderung des Präsentators nachkommen, wenn er dazu in der Lage ist, d.h. wenn gilt: *(Cando User lift-1)*. Um dieses Ziel zu erreichen, wurde das Bild generiert.

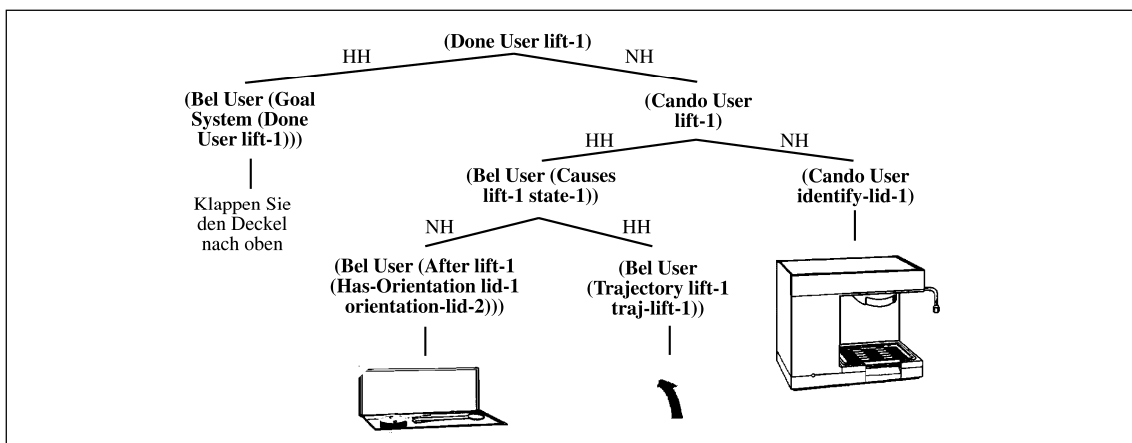


Abbildung 3.4: Hochheben der Klappe - Intentionale Struktur

Auch den einzelnen Bildteilen lassen sich unterschiedliche Ziele zuordnen. So soll durch die Darstellung des Oberteils der Espressomaschine und den Pfeil erreicht werden, daß der Benutzer glaubt, daß die Aktion *lift-1* den Zustand *state-1* verursacht, während die Abbildung der restlichen Espressomaschinenteile ihn dazu befähigen soll, den Deckel *lid-1* zu identifizieren. Das mit dem Bild assoziierte Ziel *(Cando User lift-1)* läßt sich somit in die Teilziele *(Bel User (Causes lift-1 state-1))* und *(Cando User identify-lid-1)* zerlegen. Das erste Ziel setzt sich wiederum aus den Teilzielen *(Bel User (After lift-1 (Has-Orientation lid-1 orientation-lid-2)))* und *(Bel User (Trajectory lift-1 traj-lift-1))* zusammen. D.h. der Benutzer soll glauben, daß nach Ausführen der Aktion *lift-1* der Deckel *lid-1* die Orientierung *orientation-lid-2* hat und daß *traj-lift-1* die Trajektorie des Objekts ist, das durch die Aktion *lift-1* bewegt wird.

Als nächstes wird der Benutzer gebeten, die Verschlusskappe zu entfernen (vgl. Abbildung 3.5). Wie im ersten Teil soll der Benutzer durch die Bitte darauf aufmerksam gemacht werden, daß das Entfernen der Kappe erwünscht ist, während ihm das Bild die Durchführung der Handlung ermöglichen soll. Um die Verschlusskappe zu entfernen, muß der Benutzer zunächst einmal in der Lage sein, diese aufzufinden. D.h. es muß gelten: $(\text{And } (\text{Cando User localize-container-1}) (\text{Cando User localize-cover-1}))$. Zur Realisierung dieses Ziels wird Text verwendet. Eine weitere Bedingung für die erfolgreiche Ausführung der Aktion besteht darin, daß der Benutzer weiß, welcher Zustand durch die Aktion *remove-1* erreicht werden soll und in welche Richtung die Kappe zu drehen ist. Da in diesem Fall durch das Bild eine zusammengesetzte Aktion dargestellt wird, unterscheidet sich die intentionale Struktur von der des ersten Bildes dadurch, daß sie ein weiteres Ziel enthält, nämlich daß der Benutzer einen Plan für die zusammengesetzte Aktion *remove-1* kennt. Dies wird durch $(\text{Bel User } (\text{Plan remove-1 seq-turn-lift-1}))$ ausgedrückt. Zur Realisierung dieses Ziels wurden die Trajektorien der Einzelbewegungen durch zwei Pfeile dargestellt, denen sich die Ziele $(\text{Bel User } (\text{Trajectory turn-1 traj-turn-1}))$ und $(\text{Bel User } (\text{Trajectory lift-2 traj-lift-2}))$ zuordnen lassen.

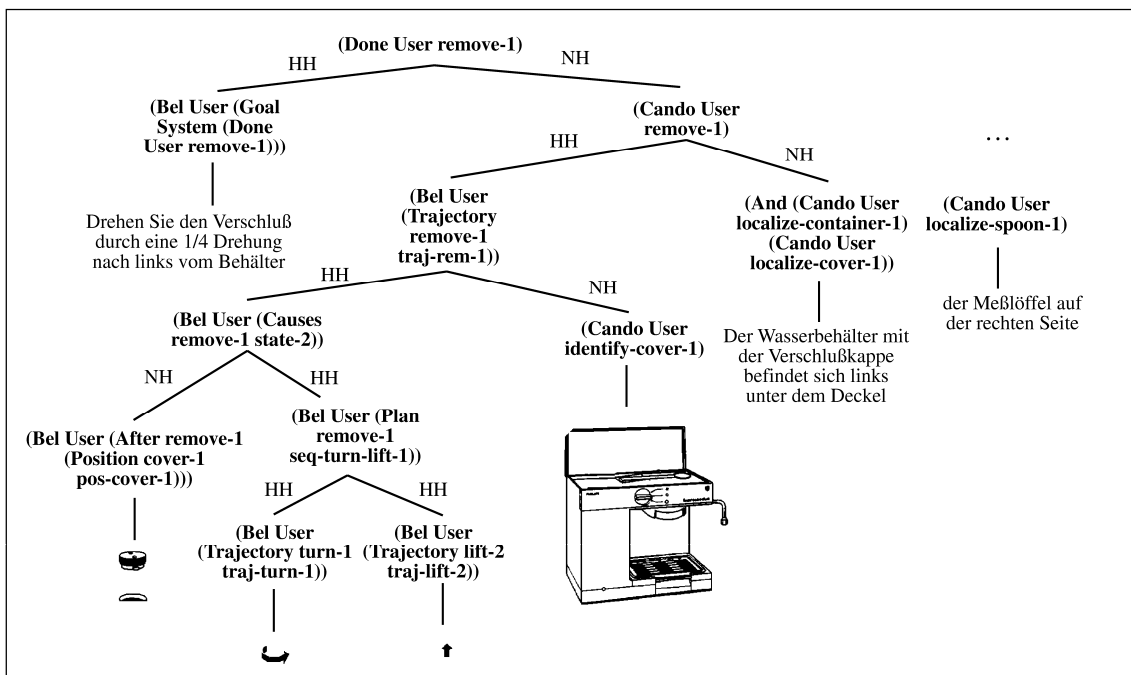


Abbildung 3.5: Abschrauben des Deckels - Intentionale Struktur

Eine Besonderheit dieses Dokumentteils besteht darin, daß das mit dem Text “der Meßlöffel auf der rechten Seite” assoziierte Ziel $(\text{Cando User localize-spoon-1})$ mit keinem anderen Ziel dieses Dokumentausschnitts in direkter Verbindung steht. Es handelt sich um ein Teilziel eines Ziels, das sich auf einen später auftauchenden Dokumentteil bezieht, der in der vorliegenden Arbeit jedoch aus Platzgründen

weggelassen wurde.

Die intentionale Struktur des dritten Teils des Dokumentausschnitts weist eine ähnliche Struktur wie die beiden vorangegangenen Teile auf (vgl. Abbildung 3.6).

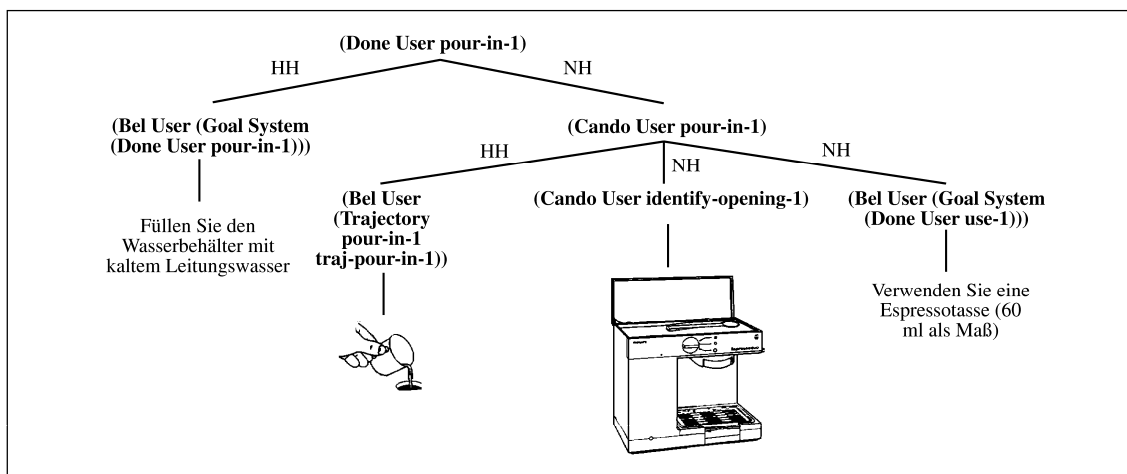


Abbildung 3.6: Eingießen von Wasser - Intentionale Struktur

Dem Benutzer wird durch eine textuelle Bitte zu verstehen gegeben, daß er die Handlung *pour-in-1* ausführen soll. Ein Bild und eine weitere Bitte sollen ihm die Ausführung der Handlung ermöglichen. Durch das Bild soll erreicht werden, daß der Benutzer den Verlauf der Einfüllaktion kennt. Der Text soll hingegen sicherstellen, daß der Benutzer weiß, welche Behälter zum Einfüllen zu verwenden sind. D.h. dem Bild werden die Ziele (*Bel User (Trajectory pour-in-1 traj-pour-in-1)*) und (*Cando User identify-opening-1*) zugeordnet, während mit der zweiten Bitte (*Bel User (Goal System (Done User use-1))*) assoziiert wird.

Der vierte Teil des Dokumentausschnitts weist die komplexeste intentionale Struktur auf (vgl. Abbildung 3.7). Wie in allen vorangegangenen Analysebeispielen wird durch die Bitte zum Ausdruck gebracht, daß der Präsentator die Ausführung einer Aktion, in diesem Fall das Aufsetzen der Verschlusskappe, wünscht. Durch das Bild soll erreicht werden, daß der Benutzer die Trajektorie der aufzusetzenden Verschlusskappe kennt. Auch in diesem Fall ist eine komplexe Aktion, die sich aus zwei Teilaktionen, nämlich *replace-1* und *turn-2*, zusammensetzt, zu visualisieren. Im Gegensatz zu den vorangegangenen Beispielen wird auch der Ausgangszustand für die Aktion dargestellt, wodurch (*Bel User (Before tighten-1 (Has-Position cover-1 pos-cover-1))*) erreicht werden soll. Das Ergebnis der Aktion wird diesmal durch eine Text-Bild-Kombination wiedergegeben. Dem Text ist das Ziel (*Bel User (After tighten-1 (Has-Orientation cover-1 orientation-cover-1))*) zuzuordnen. D.h. der Benutzer soll wissen, in welcher Stellung sich die Verschlusskappe nach dem Aufsetzen befinden soll. Da diese Stellung relativ zu den Positionen "Min" und "Max" angegeben wurde, ist der Text nur dann verständlich, wenn der Benutzer in der Lage ist, diese beiden Positionen zu lokalisieren. D.h. wenn gilt: (*And (Cando User*

localize-min-1) (*Cando User localize-max-1*)). Zur Erreichung dieses Ziels dient das Nebenbild.

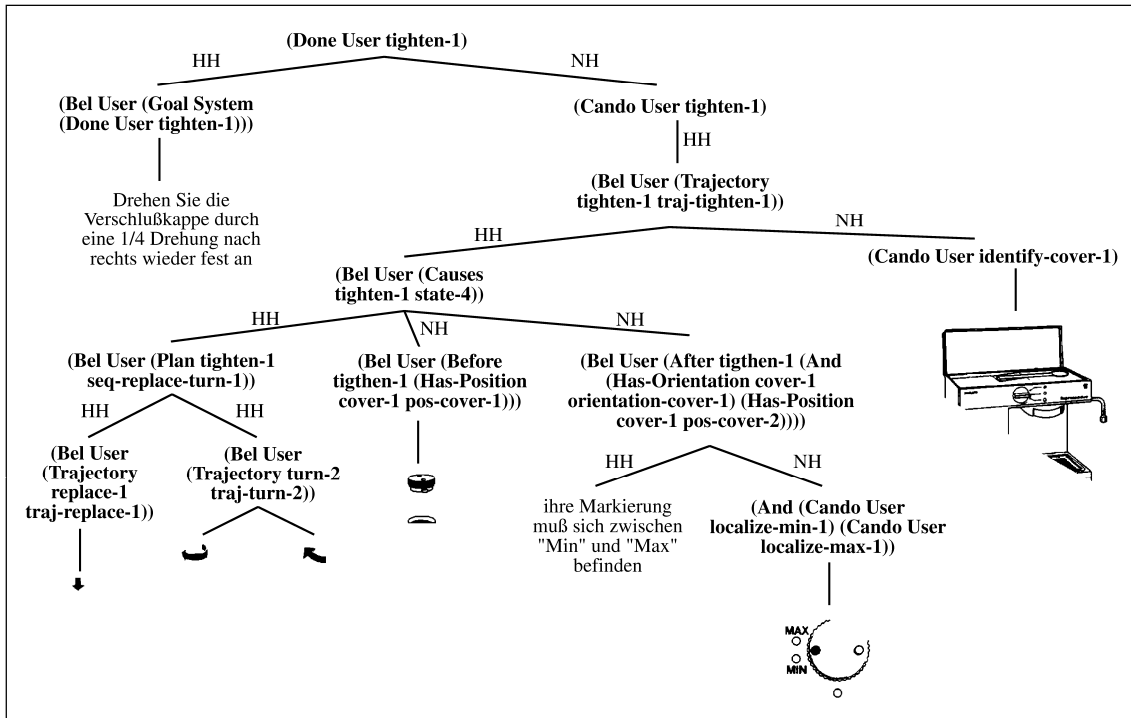


Abbildung 3.7: Aufsetzen der Verschlusskappe - Intentionale Struktur

3.2.2 Rhetorische Struktur

Wenden wir uns nun der rhetorischen Struktur des Dokumentausschnitts aus Abbildung 3.2 zu. Diese ist in den Abbildungen 3.8 bis 3.12 dargestellt. Das Dokument setzt sich aus vier nacheinander zu befolgenden Anweisungen (Describe-Sequence) zusammen (vgl. Abbildung 3.8).

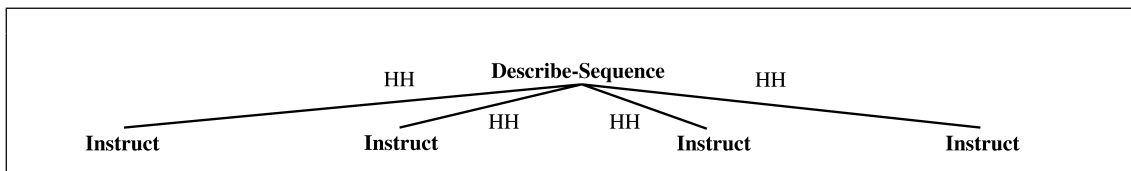


Abbildung 3.8: Füllen des Wasserbehälters - Rhetorische Struktur

Jede dieser Aufforderungen wird durch ein Bild unterstützt, das dem Benutzer die Ausführung der entsprechenden Handlungen ermöglichen soll (Enable). Auch den einzelnen Bildteilen lassen sich unterschiedliche rhetorische Funktionen zuordnen. So informiert der obere Teil des ersten Bildes über eine Cause-Result-Beziehung zwischen der durch den Pfeil dargestellten Aktion und der gezeigten Klappenstellung,

während der untere und der mittlere Teil als Hintergrund (Provide-Background) für den Rest des Bildes dienen (vgl. Abbildung 3.9).

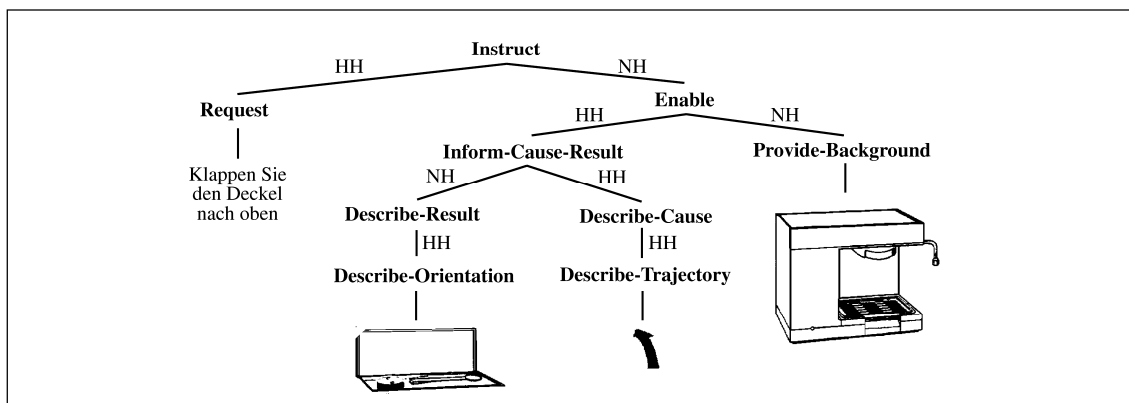


Abbildung 3.9: Hochheben der Klappe - Rhetorische Struktur

Die rhetorische Struktur des zweiten Teils des Dokumentausschnitts ähnelt der des ersten Teils, zeichnet sich jedoch durch höhere Komplexität aus. Während im ersten Bild eine elementare Handlung darzustellen war, wird im zweiten Bild auf eine Handlungssequenz verwiesen (Describe-Sequence), die sich aus zwei aufeinanderfolgenden elementaren Handlungen zusammensetzt. Darüberhinaus wird die Aufforderung durch Text zur Beschreibung der Position des Wasserbehälters und der Verschlusskappe unterstützt (vgl. Abbildung 3.10).

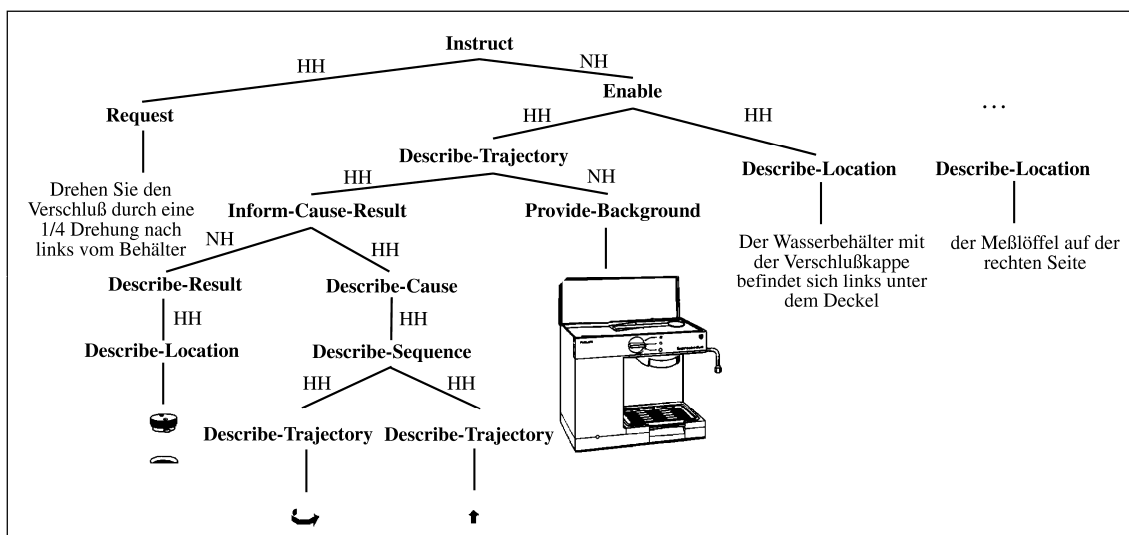


Abbildung 3.10: Abschrauben des Deckels - Rhetorische Struktur

Bereits bei der intentionalen Struktur des Dokumentausschnitts hat sich gezeigt, daß der Text "der Meßlöffel auf der rechten Seite" mit keinem anderen hier aufgeführ-

ten Dokumentteil in direkter Verbindung steht. Das Beispiel verdeutlicht, daß sich rhetorische Beziehungen auch über längere Distanzen hinweg etablieren lassen.

Der dritte Teil des Dokumentausschnitts ist dadurch gekennzeichnet, daß zwei Bitten formuliert wurden, die sich auf dieselbe Aktion beziehen. Während die erste Bitte den Kern darstellt, dient die zweite Bitte lediglich zur Unterstützung und schränkt die Art der auszuführenden Handlung durch Angabe des zu verwendenden Hilfsmittels ein (vgl. Abbildung 3.11).

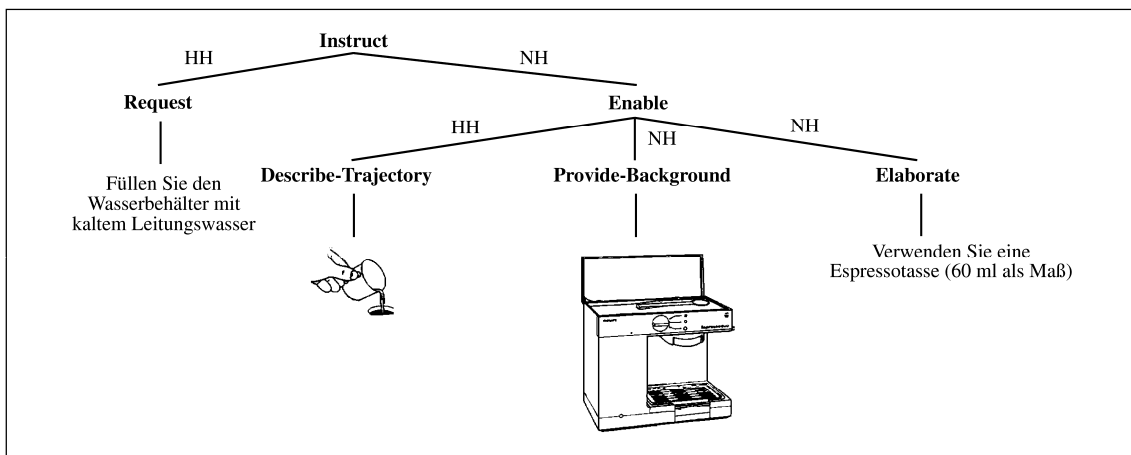


Abbildung 3.11: Eingießen von Wasser - Rhetorische Struktur

Die rhetorische Struktur des vierten Teils unterscheidet sich von der des zweiten Teils dadurch, daß die *Describe-Cause*-Handlung durch eine *Describe-Precondition*- und eine *Describe-Result*-Handlung unterstützt wird (vgl. Abbildung 3.12).

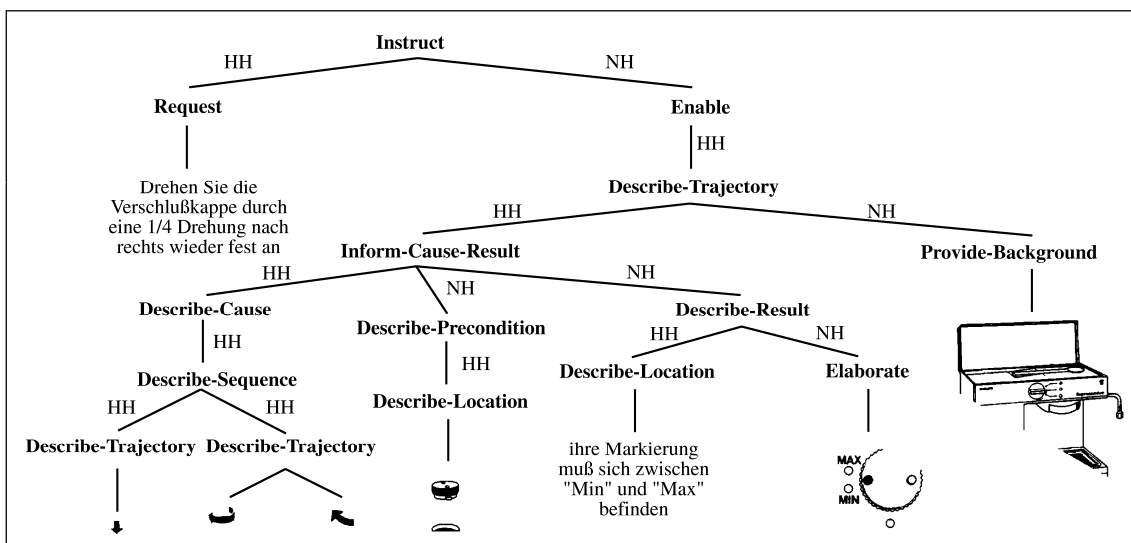


Abbildung 3.12: Aufsetzen der Verschlusskappe - Rhetorische Struktur

Darüberhinaus setzt sich die *Describe-Result*-Handlung diesmal aus einer bildlichen und einer textuellen Handlung zusammen. Der eigentliche Inhalt, nämlich die Angabe der Position der Verschlusskappe nach dem Aufsetzen, wird durch den Text vermittelt. Das Bild elaboriert den Text, indem es genauere Information zu den Positionen “Min” und “Max” liefert.

3.2.3 Attentionale Struktur

Abschließend soll die attentionale Struktur des Dokumentausschnitts ermittelt und gezeigt werden, wie diese die Auflösung von Referenzausdrücken beeinflusst. Wie in Abschnitt 3.1.3.3 beschrieben, ergibt sich die attentionale Struktur unmittelbar aus der intentionalen Struktur.

Jedem intentionalen Ziel wird ein Fokusraum zugeordnet, der Repräsentationen der Objekte enthält, auf die der Benutzer vermutlich seine Aufmerksamkeit richtet, wenn er den entsprechenden Dokumentausschnitt verarbeitet. Beispielsweise korrespondieren die Fokusräume F , $F1-1$, $F2-1$, $F3-1$ und $F4-1$ in Abbildung 3.13 zu den Zielen (*Bel User (Plan fill-container-1 seq-lift-rem-pour-in-tighten-1)*), (*Done User lift-1*), (*Done User remove-1*), (*Done User pour-in-1*) und (*Done User tighten-1*).

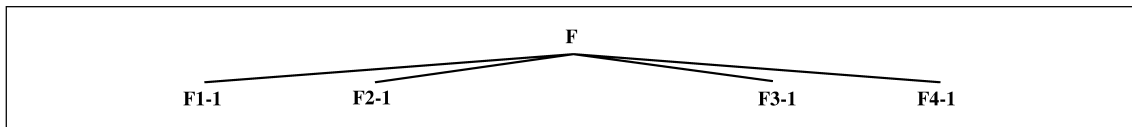


Abbildung 3.13: Füllen des Wasserbehälters - Attentionale Struktur

In Abbildung 3.14 ist die attentionale Struktur des ersten Teils des Dokumentausschnitts dargestellt. Zu jedem Blattknoten wurden die Objekte aufgelistet, die durch die Umsetzung des dazu korrespondierenden Ziels neu ins Spiel gebracht wurden.

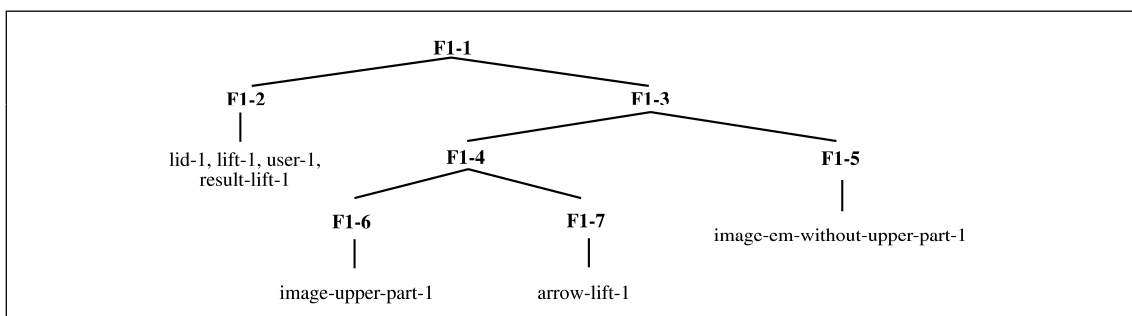


Abbildung 3.14: Hochheben der Klappe - Attentionale Struktur

Beispielsweise ist zu erwarten, daß während der Äußerung “Klappen Sie den Deckel nach oben” die Klappe *lid-1*, der Benutzer *user-1*, die auszuführende Aktion *lift-1* und deren Resultat *result-lift-1* fokussiert werden. Während der Betrachtung des

Bildes wird sich die Aufmerksamkeit des Benutzers auf das Abbild des Oberteils der Espressomaschine *image-upper-part-1*, das Abbild der restlichen Espressomaschinenterteile *image-em-without-upper-part-1* und den Bewegungspfeil *arrow-lift-1* richten. Indirekt gelangen damit jedoch auch die Referenten der graphischen Darstellungen, nämlich *upper-part-1*, *em-without-upper-part* und *lift-1*, in den Fokus.

Die attentionale Struktur des zweiten Teils des Dokumentausschnitts ist in Abbildung 3.15 zu sehen. Bemerkte sei, daß der Fokusraum *F2-6* mit keinem der hier aufgeführten Fokusräume in engerer Beziehung steht. Dies ergibt sich daraus, daß auch das zu *F2-6* korrespondierende Ziel isoliert auftrat.

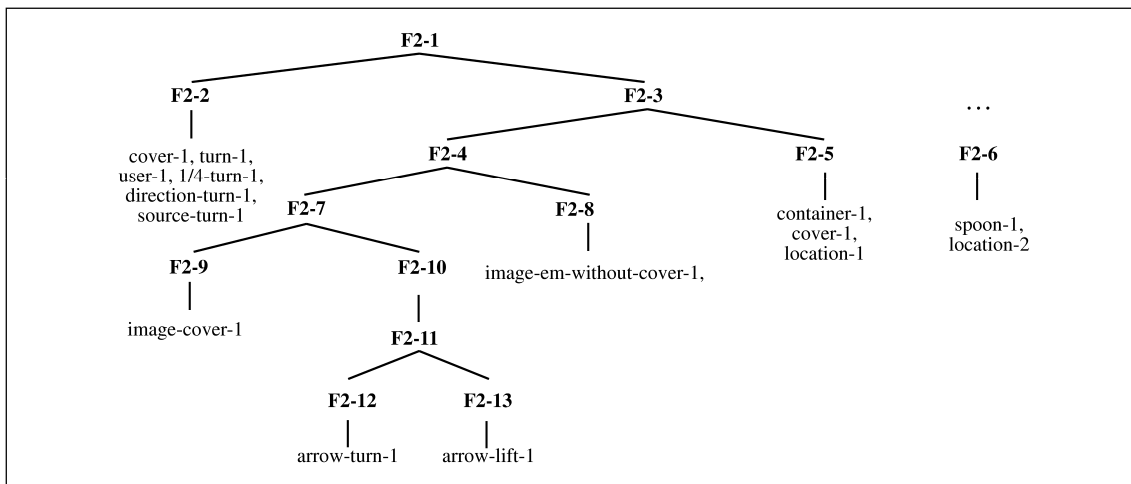


Abbildung 3.15: Abschrauben des Deckels - Attentionale Struktur

Die attentionale Struktur des dritten Teils weist keine Besonderheiten auf. Sie ähnelt stark den attentionalen Strukturen der ersten beiden Teile (vgl. Abbildung 3.14).

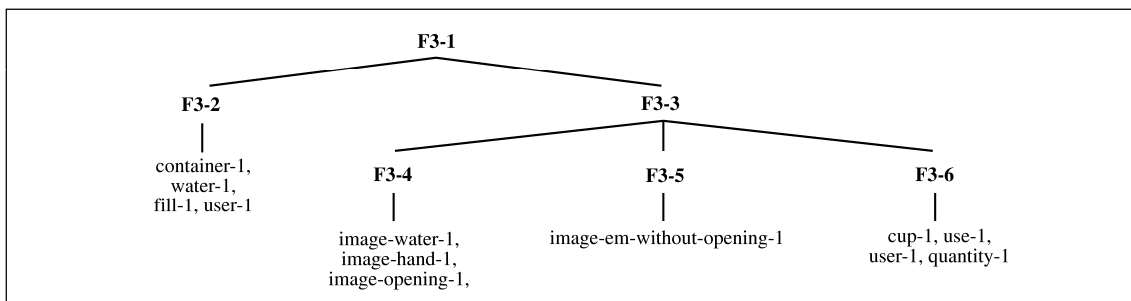


Abbildung 3.16: Eingießen von Wasser - Attentionale Struktur

Die attentionale Struktur des vierten Teils ist ein Beispiel dafür, daß nicht unbedingt jeder Fokusraum zu einem fest abgegrenzten Dokumentteil korrespondieren

muß. Wie in Abbildung 3.17 dargestellt enthält der zu dem Ziel (*Bel User (Trajectory turn-2 traj-turn-2)*) korrespondierende Fokusraum $F4-14$ graphische Objekte aus dem Haupt- und dem Nebenbild, nämlich *arrow-1-turn-2* und *arrow-2-turn-2*.

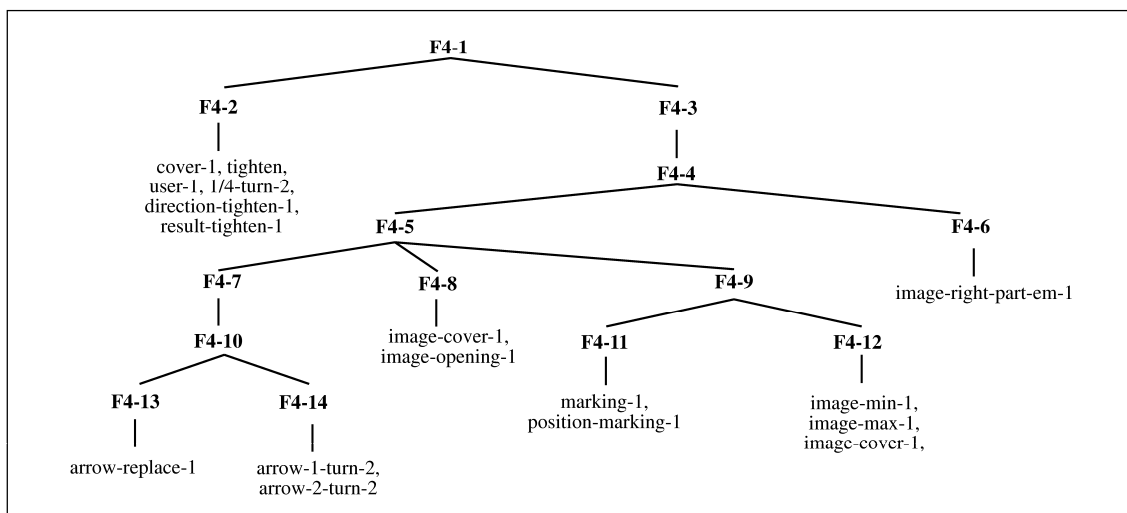


Abbildung 3.17: Aufsetzen der Verschlusskappe - Attentionale Struktur

Um zu verdeutlichen, inwiefern die attentionale Struktur die Analyse von Anaphern beeinflusst, betrachte man die Anapher “ihre” im zweiten Satz der Bildunterschrift des vierten Dokumententeils. Da $F4-2$ neben $F4-11$ der einzige Fokusraum ist, der zu einer textuellen Einheit korrespondiert, ist es wahrscheinlich, daß der linguistische Antezedent zu “ihre” dort zu finden ist. $F4-2$ enthält die Objekte *cover-1*, *tighten-1*, *user-1*, *1/4-turn-2*, *direction-tighten-1* und *result-tighen-1*, von denen *cover-1* als Antezedent identifiziert wird. Da der zu dem Nebenbild korrespondierende Fokusraum $F4-12$ von $F4-11$ aus besser zugänglich ist als andere zu Bildteilen korrespondierende Fokusräume, wird der piktorielle Antezedent zu “Seine” erst einmal hier gesucht, bevor zu anderen Fokusräumen übergegangen wird. $F4-12$ enthält die Objekte *image-min-1*, *image-max-1* und *image-cover-1* und damit auch deren Referenten, nämlich *min-1*, *max-1* und *cover-1*. *cover-1* wird als Antezedent identifiziert, ohne daß andere Möglichkeiten, z.B. die Objekte aus den zu dem Hauptbild korrespondierenden Fokusräumen, $F4-6$, $F4-7$ und $F4-8$, in Betracht gezogen werden. Nicht nur bei der Behandlung von linguistischen, sondern auch bei der Behandlung von visuellen Anaphern kann die Verwaltung von Fokusräumen nützlich sein. Beispielsweise würde man den Antezedenten für das im Nebenbild gezeigte Objekt zunächst im Hauptbild suchen, bevor man andere Abbildungen heranzieht.

3.3 Zusammenfassung

Nicht nur die Erzeugung von Sprache, sondern auch die Präsentation von Multimedia-Dokumenten läßt sich als kommunikative Handlung zur Erreichung

von Zielen auffassen. Multimedia-Präsentationen sind ähnlich strukturiert wie reine Textpräsentationen. Insbesondere ist eine Multimedia-Präsentation charakterisiert durch ihre *intentionale Struktur*, die die Ziele des Präsentators widerspiegelt, ihre *rhetorische Struktur*, die durch rhetorische Relationen bestimmt wird, und ihre *attentionale Struktur*, die sich aus den Entitäten ergibt, auf die der Adressat zu bestimmten Zeitpunkten seine Aufmerksamkeit richtet.