

Linguistische Aspekte der Website-Evaluation

Sprachwissenschaftliche Anforderungen und Verfahren im Nutzerest

1 Einführung

*A: also ich finds nicht gut, ich versteh das nämlich alles nicht
B: ja ich auch nicht, da muss man nämlich schon ahnung haben
(Äußerungen zweier Testpersonen im Nutzerest einer Website)*

Das Internet hat wesentlich zur Erweiterung des kommunikativen Haushalts von Gesellschaften beigetragen. Im Laufe der letzten Jahre sind viele Hypertextsorten (Jakobs 2004) entstanden, die in vielfältigen Kontexten des privaten, öffentlichen und beruflichen Lebens eine breite Palette von Aufgaben, Interessen und Bedürfnissen bedienen. Die Angewandte Linguistik hat sich in den letzten Jahren mit verschiedenen Aspekten online verfügbarer, hypertextuell realisierter Kommunikation befasst, dazu gehört unter anderem auch die Bewertung von Websites (Bucher 2001, 2005; Jakobs 2005; Jakobs/Lehnen 2005).

Die Auseinandersetzung mit der Qualität von Websites besitzt gesellschaftliche Relevanz: Mit der breiten Etablierung und Nutzung von Firmenportalen, Online-Shops und Informationssystemen – um nur einige Beispiele zu nennen – wächst die Nachfrage nach gut gestalteten Webangeboten und nach Verfahren der Qualitätskontrolle. Letztere werden in professionellen Domänen primär von Institutionen mit einem betriebswirtschaftlichen Hintergrund entwickelt und in Bewertungungsverfahren wie Rankings oder Benchmarks angewendet. Als Beispiel wäre unter anderem zu nennen das jährliche ibi Website Rating für Webaufritte deutschsprachiger Finanzdienstleister.

Im Vordergrund stehen dementsprechend betriebswirtschaftlich ausgerichtete Analysekonzepte, die sich Websites aus der Sicht der Betreiberinteressen nähern. Als bewertungsrelevant gelten z. B. Aspekte, die das Kaufverhalten von Kunden fördern oder der Unterstützung von Vertriebskanälen dienen. Die Berücksichtigung der Nutzerinteressen und -bedürfnisse beschränkt sich meist auf Kategorien der „klassischen“ Usability-Forschung, wie die Klicktiefe oder unterstellte Nutzungsstrategien und ihre Unterstützung durch Gestaltungselemente der Website. Die Qualität der sprachlich-visuellen Gestaltung – vom zugrunde liegenden Kommunikationskonzept bis hin zu seiner Umsetzung auf Makro-, Meso- oder Mikrolevel – spielt kaum eine Rolle. Sie liegt weder im Fokus des Interesses, noch im Bereich der Möglichkeiten der bewertenden Institutionen. Die Angewandte Linguistik könnte diese Lücke vieler Bewertungsverfahren schließen.

Unser Fokus auf die Qualität von Websites orientiert sich zum einen an der Frage, wie dieser Beitrag der Angewandten Linguistik ausssehen könnte, zum anderen an der Frage, was die Auseinandersetzung mit der Bewertung von Websites der Linguistik bringen kann. Ausgehend von einem gängigen Verfahren der Usability-Forschung, dem Nutzerest (Kap. 2), werden linguistisch relevante Aspekte des Verfahrens diskutiert. Sie betreffen zum einen den Testgegenstand und die genutzten Bewertungsmaßstäbe (Kap. 2.1), zum anderen sprachliche Anteile des Tests. Die Diskussion beschränkt sich auf ausgewählte Aspekte: die Formulierungsleistungen der Testenden (2.2) und die Auswertung von Nutzer-Kommentaren mit linguistischen Methoden (2.3). Anhand der Kommentare wird exemplarisch diskutiert, welche Typen von Daten entstehen und welche Rückschlüsse gesprächsanalytische Auswertungsverfahren zulassen.

2 Sprachlich konstituierte Aspekte von Nutzerests

Die Bewertung von Websites erfolgt in der Regel mit Methoden der Usability-Forschung. Die Usability-Forschung nutzt verschiedene methodische Ansätze, fokussiert aber meist psychologische, technische, informations- und betriebswirtschaftliche Aspekte und ist dabei stark auf quantitative Befunde ausgerichtet; sprach- und kommunikationswissenschaftliche Fragestellungen – dazu gehören Aussagen zur *Kommunikationsqualität* (BUCHER, 2001, 46) der elektronisch gestützten Kommunikationen spielen eine untergeordnete Rolle (JAKOBS/LEHNEN 2005; RICHTER 2002).

Quantitative Befunde geben wertvolle Hinweise auf die Häufigkeit, Wiederkehr oder Länge von Nutzerhandlungen, etwa über „typische“ Navigationsmuster; sie liefern aber wenig Aufschluss über Handlungsgründe und -motive der Nutzer oder die kommunikativen Konzepte, Erwartungen und subjektiven Theorien, die dem Verhalten und den Bewertungen der Nutzer zugrunde liegen. Letztere können mit Elementen der medial-kommunikativen Sozialisation von Sprachteilhabern wie: medien-spezifischen Rezeptionserwartungen, erlernten oder adaptierten Rezeptions- und Interaktionsstrategien (Kompetenz) und an (Hyper-)Textumstern orientierten Rezeptionsverfahren. Die medial-kommunikative Sozialisation von Sprachteilhabern hat wesentlichen Einfluss darauf, wie kommunikative Web-Angebote genutzt und bewertet werden.

Aussagen zur sprachlich-kommunikativen Qualität und Akzeptanz von Web-Angeboten bedingen – so der Ausgangspunkt des Beitrags – die Integration qualitativer kommunikationsorientierter Methoden, z. B. text- oder gesprächslinguistische Methoden, die in der Usability-Forschung erst in Ansätzen genutzt werden (BUCHER 2001, BUCHER/JÄCKEL 2002, RICHTER 2002). Bei genauerer Betrachtung zeigt sich zudem, dass viele gängige Testverfahren, wie z. B. der Nutzertest, hochgradig sprachlich konstituierte Verfahren sind, die spezifische sprachlich-kommunikative Kompetenzen erfordern. Die an Testverfahren gebundenen schriftlichen und mündlichen Handlungsaufgaben sind in der (Angewandten) Linguistik bisher kaum untersucht (Kap. 2.2, 2.3).

Nutzertests zu Websites simulieren Rezeptionssituationen mit dem Ziel, Hinweise darauf zu erhalten, wie das kommunikative Angebot vom Nutzer „wahrgenommen“ wird. Sie erlauben (über die Bewertung von Gestaltungsmerkmalen) Rückschlüsse auf Mustervorstellungen und -präferenzen des Adressaten, auf Optimierungsbedarf hinsichtlich der Webseitengestaltung, aber auch die Rekonstruktion von Rezeptionsstrategien und mentalen Modellen, die nutzerseitig das (kommunikative) Handeln mit hypertextuell strukturierten Kommunikationen steuern.

Das Untersuchungsdesign von Nutzertests variiert je nach Untersuchungsinteresse. Es gibt Tests, in denen der Nutzer das Kommunikat ohne Vorgaben explorativ erkundet, und Tests, in denen die Testperson vorgegebene Aufgaben löst. Die Auswahl der Testperson orientiert sich typischerweise an der (den) Zielgruppe(n) des Kommunikats, linguistisch formuliert: am Adressaten der dem Kommunikat zugrunde liegenden Websitesorte (z. B. Unternehmenswebsite vs. Online-Shop; Jakobs 2005). Frei explorierende Tests dienen – wie der Ausdruck bedeutet – dazu, Hinweise auf untersuchungsrelevante Items zu erhalten. Aufgabengeleitete Tests dienen dagegen der gezielten Überprüfung von Hypothesen. Die Aufgabensituation berücksichtigt in der Regel prototypische Funktionen des Kommunikationsangebots (im Falle einer Bibliothekswebsite z. B. das Informieren über Standorte von Büchern, das Registrieren von Literaturbestellungen etc.) sowie daran gebundene Nutzerhandlungen (z. B. das Suchen nach Informationen zum Standort eines Buches, das Durchführen des elektronischen Bestellvorgangs).

Das Untersuchungsdesign basiert in der Regel auf einem Mehrmethodenansatz (Bucher 2005). Er umfasst typische Einzelmethoden wie mündliches und/oder schriftliches Befragen der Testperson vor dem Test, Stellen von Testaufgaben, Beobachten und Dokumentieren der Aufgabenlösung. Der Test schließt häufig mit einer retrospektiven Befragung, in der die Testperson zu ihren Eindrücken während des Tests, zu ihrem Lösungsweg und/oder dem getesteten Kommunikat befragt wird. Je nach Untersuchungsdesign wird die Testperson anhand elektronischer oder schriftlicher Dokumentationen mit ausgesuchten Testpassagen (Bedien- und Rezeptionshandlungen) und/oder mit Beobachtungen des teilnehmenden Testleiters konfrontiert und gezielt zu Phänomenen befragt.

Aufgabenbasierte Nutzererfahrungen – wie oben erwähnt – der Überprüfung von Hypothesen zu ausgewählten Phänomenen der Nutzung (Rezeption) und/oder der Wirkung von Gestaltungselementen des Kommunikations. Die Identifizierung untersuchungsrelevanter Aspekte erfolgt theoretisch und/oder empirisch anhand der Analyse des zu testenden Kommunikations. Beide Varianten setzen eine theoretisch fundierte Modellierung des Kommunikations voraus.

Wir gehen davon aus, dass Websites Realisierungen spezifischer Muster sind, d.h. mustergeteilt produziert und rezipiert werden. Folgt man dieser Position, bedingt die Entwicklung von Hypothesen wie auch die Analyse konkreter Websitebeispiele ein tragfähiges Hypertext- und Webmustermodell (vgl. Jakobs 2004, 2005, Jakobs/Lehnen 2005).

Die hier vertretene Modellierung stützt sich auf die Modellierung von Hypertext als elektronisch gestützte, funktional-thematische mediale-strukturelle Darstellungsform. Im Web verfügbare hypertextuell realisierte Kommunikate werden als Websites bezeichnet. Die ihnen zugrunde liegenden Webmuster sind Teil des kommunikativen Wissens von Gemeinschaften: Ihre Mitglieder wissen, dass bestimmte Probleme unter bestimmten außersprachlich gegebenen Bedingungen kommunikativ mit Hilfe von Websites gelöst werden können. Der außersprachliche Handlungszusammenhang umfasst das zu lösende Problem (und den Sinn der Problemlösung), den die Problemlösung einbettenden zeitlichen, kulturellen und domänenspezifischen Kontext, die gegebene Situation, die kommunikativ Interagierenden (Websitebetreiber und -adressat) sowie die Wahl einer geeigneten Kommunikationsform. Die Websiteart gibt prototypische Gestaltungsinventare vor. Dazu gehören neben themenhierarchischen Strukturierungs-, Formulierungs- und Visualisierungsmustern für die Gestaltung von Content und interaktive Elemente.

Gestaltungsinventare geben Möglichkeiten für die Gestaltung des Websiteartenexemplars als Ganzes (globale Ebene) wie auch für die Gestaltung von Teilnetzen des Exemplars (Meso-Ebene) oder einzelner Seiten vor. Einzeelseiten (Knoten) werden bausteinartig gestaltet und rezipiert (Jakobs 2004). Websites setzen sich aus einer Vielzahl solcher Bausteinarten zusammen, die spezifische Beiträge zur Realisierung von Funktionen des Websiteartenexemplars leisten: funktional systembezogene Bausteine dienen primär der Orientierung und Bewegung im System (z. B. Einstiegs- und Überblickseiten), Bausteine mit thematischer Funktion (z. B. Produktbeschreibung) dienen primär der Umsetzung von Themen im Gegensatz zu Bausteinen, die primär den Vollzug nichtsprachlicher Handlungen erlauben (etwa die Login-Seite einer Bankwebsite). Der Bausteincharakter setzt sich auf lokaler Ebene fort, hier: die Ebene der Gestaltung von Einzeelseiten (Knoten). Typische Gestaltungselemente oder Bausteine der Login-Seite für geschützte Teile von Bankwebsites sind z. B. Titel, sprachliche Instruktion (für die Eingabe), Eingabemaske (mit den Feldern PIN und Nutzerkennung) sowie Textpassagen, die der (rechtlichen) Absicherung von Websitebetreiber und -nutzer dienen. Die Identifizierung von Bausteinarten unterschiedlicher Granularität ermöglicht in Kombination mit computerlinguistischen Methoden die (halb)automatische Analyse großer Websitekorpora, die Hinweise auf konventionale, - und Standardisierungsstendenz geben.

Die oben beschriebene Modellierung schärft den Blick für die Vielfalt potentiell analysierrelevanter Beschreibungs- und Bewertungsaspekte wie auch für die Herausforderung, Bedienung, Abweichung und Variation von Mustern. Die Instanzierung des Ansatzes für analytische und bewertende Zwecke bedingt Entscheidungen hinsichtlich der Frage, welche Kriterien und Maßstäbe sich aspektbezogen eignen und wie die Analysierenden ihre Maßstäbe gewinnen: durch theoretisch gestützte Setzung (normativ: wie sollte etwas typischerweise aus diesem oder jenem Grund gestaltet sein, z. B. die Formulierung von Titeln, Links, Funktionen) oder empirische Rekonstruktion (deskriptiv: was gilt aus der Sicht der Adressatengruppe als gelungene Realisierungsvariante) (Jakobs 2005).

Die Komplexität des Gegenstands verlangt das Zusammenführen verschiedener Perspektiven. Aus der Sicht ihrer funktional-thematischen Bestimmung von Websiteartenexemplaren ist zu prüfen, wie sprachlich-visuelle Gestaltung kommunikative Ziele und Aufgaben unterstützt, aus der Sicht der medialen Prägung des Kommunikations ist zu bewerten, wie die mediale Gestaltung die Rezeption des kommunikativen Angebots unterstützt. Sowohl in Bezug auf die Erfassung und Beschreibung von

Hypertextmuster und -sorten als auch daran gebundener (sozial, kulturell und zeitlich variierender) Bewertungskriterien und -maßstäbe besteht erheblicher Forschungsbedarf. Er betrifft u. a. Studien zur kulturellen Prägung von Mustern und Bewertungsmaßstäben und ihrer Veränderung in der Zeit. Bucher (2004) belegt anhand von Testergebnissen, dass sich die Gestaltungspräferenzen europäischer Webproduzenten und -rezipienten (hier: für Webportale) deutlich von denen chinesischer unterscheiden.

2.2 Sprachliche Anteile der Testvorbereitung und -durchführung

Aufgabenbasierte Nutzerests bedingen seitens des Testenden spezifische sprachliche Leistungen und Fähigkeiten in den Phasen der Testkonzeption, -aufbereitung und -auswertung (Lehnen 2005). Dazu zählen u. a. das:

- Formulieren von Fragebogen und Interviewleitfäden: Anweisungen, Fragen und Aussagen müssen knapp, neutral und eindeutig formuliert sowie sinnvoll sequenziert werden. Der Abstraktionsgrad (Wortwahl etc.) soll die sprachlichen Voraussetzungen der Testpersonen berücksichtigen. Die Antwortangaben sind kohärent sprachlich oder visuell darzustellen.
- Formulieren von Aufgaben: Die Formulierungsaufgabe umfasst das sprachliche Skizzieren anschaulicher Szenarien (Einbetten der Aufgabe in einen Handlungskontext), eindeutiges, verständliches, präzises Instruieren, klare Zielangaben sowie das Strukturieren und Beschreiben von Teilschritten.
- Formulieren eines Drehbuchs (Test szenario), das den Testablauf für Testende festlegt: Die Darstellung soll für alle Testenden gleichermaßen verständlich, eindeutig, gut nachvollziehbar sein. Es geht hier um das Beschreiben von Handlungsschritten und das Benennen von Zuständigkeiten.
- Verschriftlichen (Transkribieren) sprachlicher Äußerungen (des Testleiters und der Testperson) nach bestimmten Transkriptionskonventionen.
- (exaktes, gut strukturiertes und nachvollziehbares) Benennen und Beschreiben nicht-verbaler Handlungen der Testperson.
- Formulieren eines Testberichts: Der Testbericht soll adressatenorientiert den Test kurz beschreiben, seine Ergebnisse (auf einem angemessenen Abstraktionsniveau) zusammenfassen und gegebenenfalls Schlussfolgerungen und Empfehlungen für den Auftraggeber formulieren. Er trägt beschreibende und argumentative Züge. Der Bericht soll (trotz Datenvülle) kurz und knapp sein, andererseits jedoch dem Leser alle notwendigen Informationen für Anschlusssentscheidungen wie auch Leseanreize bieten (z. B. Anschaulichkeit).

Die produzierten Texte (Fragebogen, Aufgabenbeschreibung, Interviewleitfäden, Datenverschriftlichung) sind Teil einer komplexen Handlungskette, an deren Ende der Testbericht steht, dessen sprachliche wie inhaltliche Qualität hohe Relevanz besitzen: Die Auftraggeber müssen anhand des Berichts entscheiden können bzw. wissen, ob oder wie das Kommunikat optimiert werden soll, oder wie es im Vergleich zu anderen einzustufen ist.

Mündlich zu realisierende Anteile der Testdurchführung umfassen das Begleiten, Instruieren, Informieren und Verabschieden der Testperson durch den Testenden, Frage-Antwort-Sequenzen oder etwa das Vorlesen von Anweisungen und Aufgaben. Die interaktiven Anteile (Testender → Testperson) erfordern ein erhebliches kommunikatives Geschick. Der Testende begleitet die Testperson. Sein kommunikatives Verhalten soll der Testperson die Scheu vor dem Test nehmen, andererseits aber experimentellen Standards genügen, d. h. neutral bzw. nicht beeinflussend sein, etwa wenn sich die Testperson bei Problemen mit der Aufgabenlösung Hilfe suchend an den Testenden wendet.

2.3 Sprachliche Leistungen der Testperson: Nutzerkommentare und ihre Auswertung

2.3.1 Lautes Denken, Spontankommentierung, Sprechen bei kooperativer Aufgabenlösung

In vielen Nutzerests wird die Testperson aufgefordert, ihr Tun und/oder das Kommunikat laut zu kommentieren. In der psychologisch-kognitiv orientierten Fachliteratur wird dies als Methode des „Lauten Denkens“ bezeichnet, in eher (medien-)linguistisch orientierten Studien als Erzeugen von „Spontanankomentierungen“ (Bucher 2005, 96)². Die Situation wird in der Regel als monologisch bewertet, sie ist – linguistisch betrachtet – eher eine spezifische Gesprächssituation mit festgelegten Rollen: Die Testperson berichtet einem imaginären Partner (Testleiter), was sie tut und dabei „denkt“.

Zur Rolle des Testleiters gehört z. B. die Aufgabe, die Testperson bei Bedarf an die Aufgabe zu erinnern, sich laut zu äußern.

Eine andere Variante ist die Methode der kooperativen Aufgabenbearbeitung durch zwei Testpersonen (Co-Discovery). Die Interaktion zwischen den Testpersonen erzeugt eine natürliche Gesprächssituation (on (JORDAN 2002, 53ff.): die Testpersonen müssen sich in Bezug auf ihr Vorgehen, ihre Annahmen, zu subjektiven Theorien u. a. verständigen. Der Testleiter tritt in den Hintergrund, d. h. aus der Gesprächssituation heraus, bzw. ist Teil einer zweiten Gesprächskonstellation (Testpersonen – Testleiter).

2.3.2 Dokumentation und Analyse multikodaler Datentypen

Die Rezeption von Websites bedingt „regelmäßige, kompetenzbedingte und angebotsabhängige Sequenzen von Aneignungshandlungen“ (BUCHER 2001, 49), die der Beobachtung zugänglich sind:

Wer Hypertexte und Hypertext-Angebote nutzen will, muss den Cursor bewegen, Seiten scrollen, Links aktivieren, zurückknavigieren, auf Dialoganfragen antworten, Seitenadressen eingeben, nach Verknüpfungsstellen suchen usw. (...) Erst die Aneignungshandlungen konstituieren das Programm, die Lektüre und den Rezeptionsgegenstand.

Die im Testverlauf entstehenden Daten sind vielfältig. Ihre Aufbereitung und Auswertung bildet aus sprach- und kommunikationswissenschaftlicher Sicht ein anspruchsvolles Aufgabenfeld. Der Test erzeugt verschiedene Datenmodalitäten: verbale Äußerungen, Körperbewegungen (Körperhaltung, Gestik und Mimik), Augenbewegungen der Testpersonen, operative Handlungen (Manipulation der Tastatur und Maus), Veränderung des Bildschirminhalts (mit Cursor-Bewegungen, Pausen, Klicks, sprachlichen Eingaben etc.), rezepptive Handlungen (Lesen, Durchmusteren des Bildschirms).

Die Datendokumentation erzeugt verschiedene Datenquellen: die Bildschirmoberfläche wird mit Hilfe von Programmen wie Screencam „gefilmt“, eine externe Kamera erfasst (je nach Untersuchungsdesign) den Oberkörper der Testperson (Mimik, Gestik, Körperhaltung), auditive Daten und/oder die Bedienung von Tastatur und Maus. Beide erzeugte Dateien (Screencam und Video) enthalten die Äußerungen der im Raum befindlichen Personen.

Zu den Herausforderungen der Analyse gehört die Aufgabe, Teilaspekte des komplexen Testgeschehens „sichtweise“ zu erfassen (sprachliche Daten transkribieren, Gestik und Mimik „notieren“, nichtsprachliche Handlungen, Bewegungen und Aktionen in der Website kategorisieren und auflisten, etc.) und das Zusammenspiel der Schichten für die Analyse nachvollziehbar darzustellen. Die damit verbundenen Schwierigkeiten sind aus der Gesprächsanalyse bekannt (BERGMANN 1985; FRANCK 1989).

2.3.3 Ziel der Auswertung von Nutzerkommentaren

Ziel der Datenauswertung ist die Rekonstruktion von Problemlöseprozessen, Einstellungen, Erwartungen und Problemlösestrategien. Die Auswertung der Verbalisierungen der Testpersonen lässt Rückschlüsse auf verschiedene Aspekte der Webstierezeption wie auch der Bewertung des Kommunikats zu. Die Rückschlüsse richten sich auf:

- Nutzermerkmale: Erwartungshaltung, Annahmen, Mustervorstellungen und -präferenzen sowie Bedürfnisse,

² An der Methode ist häufig kritisiert worden, dass „Lautes Denken“ strikt genommen kaum möglich ist; die Versuchsperson kommentiert eher ausgewählte Aspekte des Testgeschehens.

A: (8) TP schaut auf den Bildschirm

1 TP: text müller okee (8)

A: TP nimmt das Aufgabenblatt in die linke Hand und liest schweigend die Aufgabe (Dauer: 6 Sekunden) [TP hält während der gesamten Zeit Aufgabenblatt in linker Hand, seine rechte Hand liegt während der gesamten Zeit auf der Maus]

Beispiel (3)

Im Folgenden soll das bisher skizzierte Potential der Methode des Lauten Denkens bzw. des Nutzerkommentars an einem Beispiel exemplarisch dargestellt und kritisch diskutiert werden. Testgegenstand ist die Website eines technischen Forschungsinstituts. Die Testperson – ein Ingenieur Mitte 40 – erhält eine Rechercheaufgabe: „Suchen Sie den Text von M. Müller, Simulation von Schleifverfahren“. Die Aufgabe wird szenisch eingebettet: „Sie befinden sich auf der Startseite [Name des Instituts] und interessieren sich für Literatur zum Thema ‚Schleifverfahren‘“. Ausgangspunkt der Suche ist die auf dem Bildschirm zu sehende Startseite. Der Transkriptausschnitt enthält Äußerungen der Testperson (TP) und des Testleiters (TL) sowie Kommentare zu den Aktionen der Testperson (A) – insbesondere bei Verbalisierungspausen – und zu dem sich verändernden Testobjekt (O).

2.3.5 Potentiale und Grenzen der Methoden Lautes Denken und Nutzerkommentierung

„(da) m probiere ich .. den but- auf der startseite probiere ich den but-ton institut (3) auf der/ in der navigationsliste (3) schaue ich mir die struktur an (1) suche nach einem link auf die forschungsguppe chemiefaser-technik“.

Beispiel 2:

Bei konkreten Aufgaben werden qualitativ andere Sprachhandlungen hervorgerufen, z. B. Äußerungen, die sich auf das eigene Vorgehen beziehen (hier: Suche nach einer Information):

„ich persönlich finde jetzt ganz unabhängig von den informationen die man da sieht das von der aufteilung her gefällt mir persönlich nicht so sehr“.

Beispiel (1)³

Die im Test zu lösenden Aufgaben evozieren unterschiedliche Typen von Informationen. Explorative Aufgaben erzeugen verstärkt Einschätzungs- und Bewertungshandlungen der Testpersonen. Im folgenden Beispiel äußert die Testperson erste Eindrücke beim Betrachten der Startseite:

2.3.4 Äußerungstypen

- Nutzerverhalten: Handlungsmuster, Strategien und Routinen,
- Nutzerprobleme im Umgang mit dem System: Probleme der Orientierung, Bewegung und Suche in der hypertextuellen Struktur,
- Bewertung der inhaltlichen und sprachlich-visuellen Gestaltung des Kommunikats: Vollständigkeit, Relevanz und Nachvollziehbarkeit der Inhalte, Strukturierung und sprachlich-visuelle Gestaltung von Inhalten etc.
- Bewertung der Gestaltung von Bedienungselementen: Art, Anzahl und Qualität von Navigations- und Orientierungshilfen (Suche, Sitemap, Links etc.) u. a. (vgl. JAKOBS/ LEHNEN 2005).

2 TL: also bitte laut denken

3 TP: ja das ist ja eine Literaturschnittstelle äh ne Literatursstelle

A: TP schaut zwischen Aufgabenblatt und Bildschirm hin und und her, klickt auf der Startseite den Link 'Literatur' (Element des linken Navigationsmenüs) an.

4 dann gehe ich auf (1) Literatur mal gucken was ich denn da kriege (3)

O: Es öffnet sich eine Überblicksseite. Unter dem Titel 'Literatur' erscheint eine Auflistung mit 3 Rubriken (Veröffentlichungstypen): "Wissenschaftliche Publikationen", "Technische Veröffentlichungen", "Bücher". Jede Rubrik enthält eine Liste mit Themenschlagwörtern mit Linkfunktion. Das Themenfeld "Schleifverfahren" erscheint gleich zweimal unter verschiedenen Rubriken.

5 Schleifverfahren (4)

A: TP fixiert Bildschirm; TP klickt Link "Schleifverfahren" unter Rubrik "Wissenschaftliche Publikationen" an.
O: Es öffnet sich eine Seite mit dem Titel "Schleifverfahren", unter dem Titel befindet sich eine lange Liste von Literaturbeiträgen (Scroll-Seite). Die Literaturangaben haben das Muster: Name, Vorname (Erscheinungsjahr): Titel, Erscheinungsort. Die Auflistung orientiert sich am Erscheinungsjahr.

6 so: jetzt kriege ich hier eine ganze Menge Veröffentlichungen kann

ich (1) ((Leiser)) kann ich hier irgendwie suchen+ (13)

A: (13) TP scrollt zunächst im oberen Bereich der Seite mit dem Scrollrad der Maus (Blickbewegungen nach oben und unten auf der Seite). TP scrollt mit dem Rad tiefer runter auf der Seite, findet den Eintrag und bewegt den Cursor auf den Eintrag.

8 Schleifverfahren da hab ichs ja (4)

A: (4) TP atmet laut durch, schaut auf Aufgabenblatt

9 so hier hätte ich mir jetzt gewünscht dass ich irgendwas suchen kann

Der Ausschnitt zeigt, welche Schritte die Testperson zur Lösung der Aufgabe unternimmt, was sich auf dem Bildschirm verändert und welche Aspekte sie dabei verbalisiert bzw. kommentiert. Auffällig ist, dass der Verbalisierungsprozess sehr unterschiedliche Funktionen hat. Die Äußerungen dienen dem:

a) Vergewenwärtigen der Aufgabe (Text müller okee)

b) Eingrenzen und Benennen des Aufgaben- oder Suchkontextes (ja ich also ist ja eine Literaturschnittstelle äh ne Literatursstelle)

c) Benennen nichtsprachlicher Handlungen (dann gehe ich auf (1) Literatur mal gucken was ich denn da kriege (3))

d) Beschreiben und Kommentieren von Handlungsergebnissen (so: jetzt kriege ich hier eine ganze Menge Veröffentlichungen)

e) Nachfragen beim Versuchsleiter (kann ich (1) kann ich hier irgendwie suchen)

f) Verbalisieren von Bedürfnissen (so hier hätte ich mir jetzt gewünscht dass ich irgendwas suchen kann).

Die Kommentierung der Testperson richtet sich u. a. auf das Benennen der während der Aufgabenlösung fokussierten bzw. konsultierten Inhalts- und Bedienelemente des Angebots (c). Die Testperson ist sich dessen bewusst, dass dies wichtig für den Test ist, sie zeigt sich dem anwesenden Testleiter gegenüber kooperativ. Diese Kooperationsbereitschaft ist eine wesentliche Voraussetzung für den Ertrag der Methode und Teil einer quasi-dialogischen Handlung (jmd. berichtet jmdm. etwas). Teil dieser kooperativen Haltung ist der Versuch, Phänomene möglichst genau bzw. nachvollziehbar zu benennen und sprachliche Fehlleistungen zu korrigieren („eine literarische Mittstelle äh ne literarische“). Andere Äußerungen bewegen sich im Spannungsfeld von Selbstverständigung und Verständnissicherung gegenüber dem Testleiter (text müllert oke). Hinweise darauf, welche Funktion überwiegt, liefern nicht-sprachliche (Pausen), sprachliche (Partikel wie so) und prosodische Mittel (Lautstärke, Intonation, u. a.). Sie können in der Phase des Pretests wichtige Anhaltspunkte für die Verständlichkeit und Nachvollziehbarkeit der Aufgabenformulierung geben, im Test selbst erlauben sie Rückschlüsse auf die Person der Testperson (z. B. auf den Grad der Vertrautheit mit der Aufgabenstellung, hier: einer Literaturrecherche).

Die Kommentare der Testperson erlauben Rückschlüsse auf ihr Repertoire an Rezeptionsstrategien wie auch – über die Präferenz für Einzelstrategien – auf Strategiehierarchien (vgl. Jakobs/Villiger 1999). Äußerungen wie „kann ich (1) kann ich hier irgendwas suchen“ zeigen, dass der Nutzer das Strukturierungsprinzip der Inhalte auf der Website (Ordnung der Literatur nach Erscheinungsjahr) nicht nachvollziehen kann. Er benötigt 13 Sekunden, um den Eintrag zu finden. Das auf der Seite genutzte Strukturierungsprinzip entspricht offensichtlich nicht seinem Orientierungsmuster (alphabetische Ordnung nach Autorennamen), das konventionell prototypisch für Literaristik ist. Die Äußerungsssequenz zeigt, dass Nutzer- und Produzentenmodell divergieren. Die Interpretation wird durch spätere metakommunikative Äußerungen der Testperson gestützt. Kommentare wie „so hier hätte ich mir jetzt gewünscht dass ich irgendwas suchen kann“ formulieren Erwartungshaltungen, hier: die interaktive „Ausstattung“ des Kommunikations mit einer elektronischen Suchfunktion, die Sortierens- und damit verbundene Suchprobleme kompensiert. Der Vergleich mit Äußerungen anderer Testpersonen zeigt, ob es sich in Fällen wie diesem um in der Zielgruppe geteilte Erwartungen handelt.

Der Vorteil der gesprächsanalytischen Annäherung im Vergleich zur inhaltsanalytischen Auswertung besteht darin, dass Zusatzinformationen der oben beschriebenen Art auf Ausprägungsgrade von Erwartungen deuten: in Zeile 9 weist die Betonung *wünschen* als Sprechakt des Wünschens *oder* des Forderns aus. Im ersten Fall würde die Äußerung auf Gestaltungs*bedürfnisse* deuten, im zweiten auf die Erwartung eines Gestaltungs*standards*.

Eine Schwierigkeit der Analyse betrifft die Identifizierung von Äußerungstypen. Es gibt z. B. Äußerungen, die sich auf das nichtsprachliche Vorgehen der Testperson beziehen *und/oder* kommentieren den Charakter besitzen. Die Äußerung „so: jetzt kriegt ich hier eine ganze Menge veröffentlichtungen“ beschreibt einseitig das (auf dem Bildschirm sichtbare) Suchergebnis, sie kann aber auch bewertenden Charakter haben, im Sinne von: Das Suchergebnis bzw. das Angebot der aktuell konsultierten Seite sind unübersichtlich. Welche Deutung zutrifft, ist häufig erst im weiteren Verlauf der Aufgabenbearbeitung (bei sehr gutem Kenntnis des getesteten Kommunikations) oder im retrospektiven Interview zu klären, wenn Testleiter und Testperson gemeinsam Passagen noch einmal durchgehen. Der zeitlich geringe Abstand zwischen Testdurchführung und retrospektivem Interview verlangt vom Testleiter neben Konzentration ein hohes analytisches Vermögen in der Klassifizierung von Probandenäußerungen und der Identifizierung analyserelevanter Stellen.

Ein wichtiger Indikator für verschiedene Phänomene sind Partikel wie *ah* und *Pause*. Pausen können unterschiedliches „bedeuten“, z. B. Unsicherheit oder Nichtverstehen (die Person prüft mental Möglichkeiten und entscheidet sich für eine) oder aber kognitive Überlastung. Das vermehrte Auftreten von Pausen im Interaktionsgeschehen (Zeile 4, 5, 7, 8) deutet darauf, dass sich die Testperson an diesen Stellen auf den Aufgabenraum „Lösen der am PC zu bearbeitenden Aufgabe“ konzentrieren muss und die Aufgabe, den Testleiter durch lautes Kommentieren an Entscheidungen und Überlegungen teilhaben lassen „vergisst“. Das Phänomen ist u. a. aus experimentellen Settings der Schreibfortsetzung bekannt: die beim (lauten) Formulieren ihres Textes Beobachteten verstummten bei schwierigen Passagen, der Beobachter nimmt dies als Pause war (z. B. Kesselring 1992). Insgesamt fehlt eine systematische Analyse des Pausenverhaltens bezogen auf Aufgaben und Probanden.

Der Gesprächsanalytische Zugang ist auf der einen Seite sehr ergiebig, auf der anderen Seite jedoch zeit- und ressourcenintensiv. Im Sinne von Kosten-Nutzen-Erwägungen ist er nicht durchgängig durchzuführen, sondern der Analyse kritischer (mehrdeutiger) analyserelevanter Passagen vorbehalten.

2.3.6 Vor- und Nachteile der Methode des Co-Discovery

Abschließend sollen kurz Vor- und Nachteile der Methode des Co-Discovery angerissen werden (für eine ausführlichen Darstellung: Rick 2005). Beim Co-Discovery lösen zwei Testpersonen gemeinsam eine Aufgabe. Sie werden gebeten, Handlungen und Eindrücke zu verbalisieren. Die Kooperations-situation hat Vorteile. Sie lenkt partiell von der Künstlichkeit der Testsituation ab. Gegenseitige Kenntnis der Testpersonen reduziert die Angst vor Fehlern und sozialer Bloßstellung, Phänomene werden weniger rationalisiert (JORDAN 1998).

Die Methode des Co-Discovery hat neben der Erzeugung „natürlicher“ Gesprächssituationen weitere Vorteile. Sie eignet sich per se gut für Anwendungen, die für kooperative Nutzungssituationen entwickelt wurden (z. B. firmeninterne Wissensmanagementsysteme), da sie deren Bedingungen besser simuliert (WILSON/BLOSTEIN 1998).

Die im Verlauf der Aufgabenlösung geleisteten Hilfestellungen und Vorschläge der Testpartner erlauben Rückschlüsse darauf, welche Art der Hilfestellung, Erklärung oder Problemlösung eher angenommen wird als andere und/oder erfolgreich ist (MIYAKE 1982, RICK 2005).

Hinweise auf das (tatsächliche) Nutzerverhalten liefern insbesondere Situationen, in denen die Testpersonen ihr Verhalten im Test mit ihrem Verhalten in Realsituationen vergleichen:

„Ja im realfall hätte ich dann auch an der stelle abgebrochen und irgend-wen angerufen und gefragt ist es das jetzt ist es das nicht ich such jetzt xp und da steht nur rwth version“.

Nachteile der Co-Discovery-Methode ergeben sich, wenn die Voraussetzungen der Testpartner differieren (Arten des Lernens, kulturelle Hintergründe, sprachliche Gewohnheiten: einer redet viel und sehr laut, einer wenig und sehr leise; KAHLER 2000, 30) und Muster ihres Umgangs mit dem Computer (NIELSEN 1993, 198). Testpersonen können miteinander nicht kooperieren; sie arbeiten zudem nur dann gut zusammen, wenn sie verstanden haben, dass Zusammenarbeit zu den Zielen des Tests gehört und wie sie aussehen soll (WILSON/BLOSTEIN 1998). Die verbale Interaktion der Testpersonen kann sich anders als gewünscht entwickeln (der zu untersuchende Aspekt wird nicht thematisiert); korrigierende Eingriffe des Versuchsführers sind möglich, sie stören jedoch die Spontanität des Verhaltens der Testpersonen (JORDAN 1998, 54). Mit der Anzahl der Testpersonen wächst zudem die Komplexität des Geschehens und der zu interpretierenden Daten (DUMAS/REDISH 1999, RICK 2005).

5 Ausblick

Das Usability-Testing von Websites eröffnet aus zwei Perspektiven interessante linguistische Anknüpfungspunkte. Die erste Perspektive betrifft die Bewertung relativ neuer sprachlicher Kommunikationsmittel, hier Websites als Realisierung von Webistemustern, und die Frage, woher die Bewertenden ihr Wissen um bewertungsrelevante Aspekte des Kommunikationsmittels und die daran gebundenen Bewertungsmaßstäbe beziehen. Hier ist insbesondere die Hypertextlinguistik gefordert im Sinne der Entwicklung von Modellen zur Erfassung und Analyse von Webistemustern wie auch der Beschreibung ausgewählter Muster bezogen auf Kulturen, Gestaltungsprinzipien und -maximen. Letztere erlauben u. a. Hin- und hergehende Studien zu klären, welche spezifischen sprachlich-kommunikativen Kompetenzen Testentwickler benötigen, u. a. um analytisch reiche Verfahren, wie sie z. B. die Angewandte Gesprächs-

Die zweite Perspektive betrifft die sprachlichen Anteile von Bewertungsverfahren. Hier wäre in weit-

forschung anbietet, in die Auswertung sprachlicher Testdaten einzubeziehen. Insgesamt kann und sollte sich die Linguistik stärker mit der Frage der Bewertung elektronischer Kommunikate befassen und sich stärker als bisher in Bewertungsverfahren der Praxis einbringen.

Literatur

- BERGMANN, JÖRG (1985): Flüchtigkeit und methodische Fixierung sozialer Wirklichkeit: Aufzeichnungen als Daten der interpretativen Soziologie. In: BONS, WOLFGANG/HARTMANN, HEINZ (Hrsg.): Entzauberte Wissenschaft. Göttingen, S. 299-320.
- BUCHER, HANS-JÜRGEN (2001): Von der Verständlichkeit zur Usability. Rezeptionsbefunde zur Nutzung von Online-Medien. In: Osnabrücker Beiträge zur Sprachwissenschaft (OBST) Nr. 63, November 2001: Hypermedien und Wissenskonstruktion, S. 45-66.
- BUCHER, HANS-JÜRGEN (2004): Online-Interaktivität – ein hybrider Begriff für eine hybride Kommunikationsform. Begriffliche Klärungen und empirische Rezeptionsbefunde. In: BIEBER, CHRISTOPH / LEGGOWIE, CLAUDIUS (Hrsg.): Interaktivität. Ein transdisziplinärer Schlüsselbegriff. Frankfurt a. M. S. 132-167.
- BUCHER, HANS-JÜRGEN (2005): Ist das Internet „ready“ für deine Nutzer? Online-Angebote zwischen Gebrauchstauglichkeit und Kommunikationsqualität. In: JÄCKEL, MICHAEL / HAASE, FRANK (Hrsg.): In medias res: Herausforderung Informationsgesellschaft, München, S. 81-116.
- BUCHER, HANS-JÜRGEN/JÄCKEL, MICHAEL (2002): „E-Business-Plattformen im Usability-Test“. In: ROTERS, GÜNNAR/TURBECK, OLIVER/KLINGLER, WALTER (Hrsg.): Content im Internet. Trends und Perspektiven. Berlin, S. 67-82.
- DUMAS, JOSEPH S./REDISH, JANICE C. (1999): A Practical Guide to Usability Testing. Exeter: protobornterierten Gesprächsanalyse. In: Zeitschrift für Sprachwissenschaft, Phonetik und Kommunikationsforschung (42), S. 160-167.
- JAKOBS, EVA-MARIA (2004): „Hypertextsorten“. In: Zeitschrift für germanistische Linguistik 31 (2). Berlin, S. 232-252.
- JAKOBS, EVA-MARIA (2005): Bewertungsperspektiven auf Websites. In: SCHÜTZ, ASTRID/HABSCHEID, STEPHAN/HOLLY, WERNER/KREMS, JOSEF/VOB, G. GÜNTHER (Hrsg.): Neue Medien im Alltag. Befunde aus den Bereichen: Arbeit, Lernen und Freizeit. Lengerich u. a. S. 71-86.
- JAKOBS, EVA-MARIA/LEHNEN, KATRIN (2005): Hypertext – Klassifikation und Evaluation. In: RUNKEHL, JENS/SIEVER, TORSTEN/SCHLOBINSKI, PETER (Hrsg.): Websprache.net. Sprache und Kommunikation im Internet (= Linguistik & Tendenzen 10). Berlin/New York, S. 159-184.
- JAKOBS, EVA-MARIA/VILLIGER, CLAUDIA (1999): „Das versteht kein Mensch...“. Verständliche Gestaltung von Hilfesystemen für Softwareprogramme. In: JAKOBS, EVA-MARIA/KNORR, DAGMAR/POGNER, KARL-HEINZ (Hrsg.): Textproduktion. Hypertext, Text, Kontext, S. 211-227
- JORDAN, PATRICK W. (1998): An Introduction to Usability. London: Taylor & Francis
- KAHLER, HELGE (2000): Constructive Interaction and Collaborative Work: Introducing a Method for Testing Collaborative Systems. In: ACM Interactions 7 (3), S. 27-34
- KESELING, GISBERT (1992): Pause and intonation contours in written and oral discourse: In: STEIN, DIETER (ed.): Cooperating with written text. The pragmatics and comprehension of written texts. Berlin, New York, S. 31-60.
- LEHNEN, KATRIN (2005): Vermittlung berufsbezogener Schreibkompetenzen im Studium. Am Beispiel des Usability Testing. In: JAKOBS, EVA-MARIA/LEHNEN, KATRIN/SCHINDLER, KIRSTEN (Hrsg.): Schreiben am Arbeitsplatz. Wiesbaden, S. 235-250
- MIYAKE, NAOMI (1982): Constructive Interaction. San Diego: University of California, Center for Human Information Processing [Technical Report; 113]

- NIELSEN, JAKOB (1993): Usability Engineering. Boston u. a.
- RICHTER, GERD (2002): Web-Usability. Oder: Wie man die Benutzbarkeit von Web-Seiten untersucht kann. In: Spiegel der Forschung 19. Jg., H2, S. 75-78
- RICK, URSULA (2005): Co-Discovery und Lautes Denken. Ein Methodenvergleich. Magisterarbeit am Institut für Sprach- und Kommunikationswissenschaft, RWTH Aachen.
- RUBIN, JEFFREY (1994): Handbook of Usability Testing: how to plan, design, and conduct effective tests. New York.
- WILSON, CAUCEY/BLOSTEIN, JUDY (1998): Pros and Cons of Co-participation in Usability Studies. In: Usability Interface 4 (4). <http://www.stcsig.org/usability/newsletter/9804-coparticipation.html>

Anhang

Transkriptionszeichen

Die in Beispiel 1-3 verwendeten Transkriptionszeichen:

- ... kurze Pause (die Anzahl der Punkte entspricht der relativen Dauer)
- (4) längere Pause (Zahl in Klammern gibt Länge in Sekunden an)
- : gedehnte Aussprache (z. B. so:)
- / Abbruch der Äußerung (z. B. in dem mentl/ auf der Seite suche ich)
- Wortabbruch (z. B. verka-äh einkaufen)
- ((leise)) Kommentierung des Redeverhaltens; das + markiert das Ende der kommentierten Äußerung