Eva-Maria Jakobs, Claas Digmayer und Bianka Trevisan

Methoden der IBK-Forschung: Zum Einfluss von Gebrauchsmuster, Domäne und Nutzer

1 Einführung

In Literatur und Öffentlichkeit wird zum Teil generalisierend von „internetbasierter" Kommunikation (IBK) gesprochen. Sie wird charakterisiert durch Schnellschreib-Phänomene, eine Tendenz zu sprachlicher Ökonomie, Orientierung am Duktus der gesprochenen Umgangssprache, „verschriftete Umgangssprache" und/oder die Verwendung innovativer semiotischer und sprachlicher Formen, die sich in der IBK als Mittel der emotionalen und evaluativen Kommentierung, der Kohärenzsicherung und des spielerischen Rekurses auf Körperlichkeit herausgebildet haben (Emoticons, Inflektive, Adressierungsausdrücke) (u. a. Beißwenger et al. 2012). Es gibt kaum Studien, die Aussagen dazu liefern, in welchem Umfang derartige Phänomene auftreten, d. h. ob diese für internetbasierte Kommunikation tatsächlich repräsentativ sind.

Aussagen dazu dürften erst möglich sein, wenn wir ein genaueres Bild von der Vielfalt des IBK-Haushaltes haben, genauer: seiner Gebrauchsmuster und ihrer Nutzung. Für die Analyse des IBK-Haushalts können wir zum Teil auf etablierte Analysemethoden zurückgreifen, zum Teil erfordert sie neue Ansätze und methodische Zugänge. Der vorliegende Beitrag thematisiert die damit verbundenen Herausforderungen. Im Folgenden wird am Beispiel zweier ausgewählter IBK-Gebrauchsmuster – thematischer Blogkommentar (Kap. 3.1) und Open-Innovation-Portal mit Community-Funktionen (Kap. 3.2) – diskutiert, wie sich Eigenschaften des kommunikativen Gebrauchsmusters, seine Einbettung in gesellschaftliche Handlungsbereiche (Domänen), das behandelte Thema und die Betrachtung von Nutzertypen auf die Methoden der Erhebung, Aufbereitung und Analyse von IBK-Daten auswirken. Die Diskussion stützt sich auf den „Aachener IBK-Ansatz" (Kap. 2.2).

DOI 10.1515/9783110567786-005

2 Framework

2.1 IBK-Gebrauchsmuster

In den vergangenen zwanzig Jahren hat sich im Zuge der Digitalisierung von Kommunikation und durch internetbasierte Technologien ein exponentiell schnell wachsender digitaler Kommunikationsraum entwickelt. Betrachtet man Phänomene wie das „Internet der Dinge und Services“, Facebook oder „Industrie 4.0“, ist klar, dass der Übergang von der Informations- zur Webgesellschaft zumindest auf dem Wege ist. Für die internet- bzw. webbasierte Kommunikation steht inzwischen ein umfangreiches Repertoire von Kommunikationsformen und kommunikativen Gebrauchsmustern zur Verfügung. Dieses Repertoire ist sozio-ökonomisch, kulturell und historisch-zeitlich geprägt (Jakobs 2011). Die zeitliche Prägung ist insofern interessant, als sich die Herausbildung des Repertoires nicht nur außerordentlich schnell vollzog, sondern im Kontext der technologischen Entwicklung in einem hohen Tempo und Umfang weiter ausdifferenziert. Wobei mit Holly (2011) zu fragen wäre, was wen vorantreibt – das technisch Mögliche die Herausbildung neuer Kommunikationsformate oder neue kommunikative Bedarfe die technologische Entwicklung. Viele Formen der internetbasierten Kommunikation sind bislang erst in Ansätzen erforscht (Beißwenger 2013), z. B. unter dem Aspekt ihrer Musterhaftigkeit. Ausnahmen bilden ältere, etablierte Kommunikationsformen wie E-Mail, Chat oder SMS.

In Bezug auf den Teil des kommunikativen Haushalts, der durch internetbasierte Kommunikation abgedeckt wird, interessieren uns insbesondere Aspekte wie Musterhaftigkeit und funktionale Prägung, genauer: (verfestigte) kommunikative Gebrauchsmuster als funktional-thematisch bestimmte Anwendungsformen von Kommunikationsformen. Wir nutzen den Begriff Kommunikationsform im Sinne von Brinker (2010). Kommunikationsformen (wie z. B. Telefonat) sind über situative und mediale Merkmale beschreibbar. Sie geben die Rahmenbedingungen der Interaktion mit Gebrauchsmustern vor und zeichnen sich durch bestimmte Merkmale aus wie etwa Zeichentyp und Kommunikationsrichtung (ähnlich Dürscheid 2005, die zwischen Kommunikationsform und kommunikativer Gattung unterscheidet: „Kommunikationsformen bilden den äußeren Rahmen des kommunikativen Geschehens, kommunikative Gattungen sind die in der Kommunikation konstruierten Handlungsmuster, die den Beteiligten eine Orientierung geben“).

Gebrauchsmuster sind nach Sandig (1997) konventionalisierte kommunikative Standardlösungen für wiederholt auftretende, sozial relevante Probleme. Die Sprachteilhaber wissen, dass sie bestimmte Probleme und Aufgaben unter

bestimmten Bedingungen typischerweise mit bestimmten sprachlichen und visuellen Mitteln bearbeiten und so mit anderen Beteiligten in Kontakt treten können. Die Bedingungen sind Teil des Handlungstyps. Sie umfassen Faktoren wie die Einbettung des Musters in eine bestimmte Domäne (einen gesellschaftlichen Handlungsbereich mit seinen Werten, Normen und Regeln, Brinker 2010) als Teil einer sozio-ökonomisch, zeitlich-historisch und kulturell geprägten Umwelt und dazugehörigen Handlungssituationen (Jakobs 2011), die dort verfügbaren technischen Mittel sowie daran gebundene Codes und Modes. Das Gebrauchsmuster stellt prototypische Mittel für die kommunikative Bearbeitung von Zielen zur Verfügung, wie typische Themen, sprachliche Handlungs- und Visualisierungsmuster oder etwa Vorgaben zu Umfängen.

Die Auseinandersetzung mit IBK-Mustern wird zum Teil durch Zugangsprobleme erschwert. So erhalten Forscher eher selten Zugang zu innerbetrieblich genutzten IBK-Formaten (z. B. Social-Media-Applikationen in Unternehmen). Die IBK-Forschung konzentriert sich deshalb häufig auf privat genutzte und/oder öffentlich zugängliche IBK-Gebrauchsmuster; Studien zu professionellen Nutzungssituationen und Gebrauchsmustern sind vergleichsweise selten (u. a. Beißwenger 2013). Die Erhebung von Daten erfolgt in einem wenig geregelten Rechtsraum. Nach wie vor ist unklar, wer was im Internet erheben darf, wie lange IBK-Daten gespeichert werden dürfen und wem sie gehören (vgl. dazu Beißwenger et al. in diesem Band).

Desiderate der Forschung betreffen nicht nur die Erfassung, Beschreibung und Modellierung des IBK-Haushalts, sondern auch seine Veränderung. Die Beiträge des Handbuchs „Textsorten, Handlungsmuster, Oberflächen. Linguistische Typologien der Kommunikation“ (Habscheid 2011) diskutieren ausführlich den aktuellen Forschungsstand – IBK wird dabei nur am Rande behandelt. Es ließe sich einwenden, dass die theoretische Auseinandersetzung mit kommunikativen Mustern zeitversetzt zu den Entwicklungen in der Welt erfolgen muss; bezogen auf die oben erwähnte Geschwindigkeit der Veränderung digitaler Kommunikationsräume, -formate und -praxen wäre allerdings zu hinterfragen, was „zeitversetzt“ in diesem Kontext bedeutet.

Technisch ist inzwischen vieles möglich. Die verfügbaren Tools erlauben ein umfangreiches Screening der Spuren medialer Wandelprozesse durch die (kontinuierliche) Aufzeichnung von Daten. Schwieriger wird es, wenn es um inhaltliche Fragen geht, wie die Unterscheidung und Modellierung von IBK-Formen und -Gebrauchsmustern. Im Falle hybrider IBK-Formate[1] z. B. fehlen

1 IBK-Formate, die verschiedene Gebrauchsmuster umfassen bzw. kombinieren.

weitgehend adäquate Beschreibungsansätze (vgl. aber Dürscheid et al. 2010, Brommer/Dürscheid 2012) wie auch adäquate Methoden der Erhebung, Aufbereitung und Analyse von Gebrauchsmusterexemplaren. Ähnliches gilt für das Erfassen und Beschreiben konkreter Produzenten und Rezipienten als (Sprach-) Nutzertypen, die Rekonstruktion von Akteur-Konstellationen oder die Betrachtung des Einflusses situativer Parameter (wie Domänenspezifik oder zeitbezogene Phänomene, etwa Moden). Wie am Beispiel der Gebrauchsmuster *themenspezifischer Blogkommentar* und *Open-Innovation-Portal mit Community-Funktion* zu zeigen sein wird, ist der Zugang zu den genannten Größen oft nur über Umwege möglich.

2.2 Der Aachener IBK-Ansatz

Der vorliegende Beitrag stützt sich auf den Aachener IBK-Ansatz. Der Ansatz betrachtet internetbasierte Kommunikation aus verschiedenen Perspektiven. Die Forschung richtet sich auf

1. die Beschreibung und Analyse von IBK-Formen und -Gebrauchsmustern (z. B. Hypertextmuster, Jakobs 2011; Question-Answer-Systeme und Social Media in Unternehmen, Digmayer/Jakobs 2014; Facebook, Wirtz-Brückner 2015; Blogartikel und -kommentar, Trevisan 2014; Tweet, Korioth 2011)
2. die Gestaltung, Nutzung und Bewertung von IBK-Gebrauchsmuster (-exemplar-)en (Open-Innovation-Plattform mit Community-Funktionen, Digmayer/Jakobs 2012a, 2012b, 2012c, Digmayer 2016; Reiseinformationssysteme als Self Services, Jakobs 2012, Wirtz/Jakobs 2013, Digmayer et al. 2015a)
3. die Analyse von Äußerungen in IBK-Formaten für Zwecke der Technikwahrnehmungs- und der Risikoforschung (u. a. Facebook, Trevisan et al. 2014, Trevisan/Jakobs 2015; Blogs, Digmayer et al. 2015b).
4. die Methodenentwicklung für 1–3.

Die erstgenannte Perspektive schließt die Frage ein, ob und wie vorliegende Beschreibungsansätze für Kommunikationsformen und Gebrauchsmuster auf IBK-Formate anwendbar sind und welcher Modifikationen sie bedürfen (u. a. Jakobs 2003, 2011, Trevisan 2014). Das spezielle Interesse gilt professionellen Domänen (Handlungskontexten, -aufgaben und -akteuren).

Eine wesentliche Voraussetzung für den Vergleich von IBK-Gebrauchsmustern und ihrer Nutzung ist eine aussagekräftige Datenbasis. Teil des Aachener Ansatzes ist der schrittweise Aufbau eines größeren Gesamtkorpus, das Daten verschiedener Forschungsprojekte zusammenführt. Das Gesamtkorpus

umfasst nicht nur Korpora zu Gebrauchsmusteranwendungen, sondern auch andere Typen von Daten, z. B. Videodaten und Transkripte (aus Nutzertests und -interviews). Letztere sind Teil empirischer Studien, die erheben, wie Musterrealisierungen von Nutzergruppen wahrgenommen und bewertet werden. Die verbalen Spontankommentierungen von Testpersonen, z. B. in kooperativen Aufgabensettings, liefern u. a. Hinweise auf subjektive Theorien (z. B. was als typisches Mustermerkmal oder als „angemessenes" sprachliches und/oder soziales Agieren gilt) oder Unterstützungsbedarf bei der Nutzung komplexer IBK-Angebote, z. B. Formulierungshilfen für die Kommunikation beruflicher Sachverhalte in unternehmensintern genutzten Social-Media-Applikationen (u. a. Digmayer/Jakobs 2014).

Die Bearbeitung der Forschungsfragen bedingt methodische Entwicklungsarbeit. Die Analyse von IBK-Gebrauchsmustern und ihrer Nutzung (Perspektive 2) erfolgt in der Regel durch die Kombination qualitativer und quantitativer Verfahren der digital gestützten Datenerhebung, -aufbereitung und -analyse, durch die Verbindung manueller und digitaler Bearbeitungsschritte und durch Methodentriangulation. Die Analyse sprachlich bewertender Äußerungen in IBK-Musteranwendungen (Perspektive 3) erfordert die Anreicherung von Text-Mining-Verfahren mit linguistischen Verfahren, wie die linguistische Mehrebenen-Annotation (vgl. Kapitel 3.1). Herausforderungen betreffen u. a. die Intermodalität komplexer IBK-Formate, die bislang nur partiell mit verfügbaren Verfahren und Tools zu erfassen und abzubilden ist (vgl. Kapitel 3.2). Teil der methodischen Entwicklungsarbeit ist die (Weiter-)Entwicklung digitaler Tools (z. B. für die *topic detection*).

3 Fallbeispiele

Die folgende Diskussion von Methoden und Tools der Analyse von IBK-Gebrauchsmustern thematisiert exemplarisch am Beispiel zweier Gebrauchsmuster (Themenspezifischer Blogkommentar, Kapitel 3.1; Open-Innovation-Portal mit Community-Funktionen, Kapitel 3.2) methodische Herausforderungen der Berücksichtigung von Gebrauchsmuster, Domäne, Thema und Nutzertyp.

3.1 Themenspezifischer Blogkommentar

Das erste Beispiel ist das Gebrauchsmuster „Themenspezifischer Blogkommentar“ (3.1.1). Im Folgenden wird skizziert, welche Herausforderungen dieses Muster an Methoden und Tools der Datenerhebung (3.1.2), -aufbereitung (3.1.3) und -auswertung (3.1.4) stellt.

3.1.1 Kurzbeschreibung des Gegenstands

Die Kommunikationsform Blog umfasst verschiedene Ausprägungen und Gebrauchsmuster (Blogartikel und Blogkommentar). Die folgenden Ausführungen beziehen sich auf thematische Blogs. Thematische Blogs behandeln in der Regel exklusiv *ein* Thema bzw. *einen* Themenkomplex (z. B. Familie, Religion oder erneuerbare Energien). Unser Beitrag fokussiert thematische Blogkommentare und Anforderungen an ihre Erhebung, Aufbereitung und Analyse.

Blogkommentare eröffnen Sprachteilhabern die Möglichkeit, sich zu äußern und ihre Äußerung anderen zugänglich zu machen (Trevisan 2014: 43/44). Dies kann bezugnehmend auf den Blogartikel und/oder andere Kommentare geschehen wie auch „frei“ – der Blogger nimmt keinen Bezug auf Vorangegangenes, sondern äußert sich zu einem selbst gewählten Thema (das seinerseits wieder Diskussionen auslösen kann, aber nicht muss). Als Motive bzw. Zwecke des öffentlichen Kommentierens werden in der Literatur genannt: *Dokumentieren des eigenen Lebens*, *Ausdruck tief empfundener Emotionen*, *Ideen verbreiten* oder *Bildung und Aufrechterhaltung von Gemeinschaften* (Nardi et al. 2004: 43). Je populärer das Thema und je größer die Sichtbarkeit des Blogs (etwa auf Grund der Popularität des Betreibers), desto häufiger scheinen Blogartikel kommentiert zu werden (Alby 2008). Die Anzahl der Kommentare pro Blogartikel kann dementsprechend stark variieren. Weitere Unterschiede betreffen das Verhalten der Kommentatoren – einige äußern sich sehr häufig, andere dagegen eher selten. Bislang fehlen u. a. Studien, die erheben, wie sich die Postinghäufigkeit auf die Art und Weise des sprachlichen Handelns in Blogkommentaren auswirkt.

Die folgende Diskussion stützt sich auf Daten des interdisziplinären Forschungsprojekts[2] HUMIC. Die Daten wurden erhoben, um Hinweise darauf zu

2 HUMIC: „Akzeptanzbewertung als integraler Bestandteil von Entwicklung und Ausbau komplexer technischer Systeme. Am Beispiel Mobilfunk“, 2009-2012, gefördert von der Exzellenzinitiative des Bundes und der Länder.

erhalten, wie im Internet bestimmte Technologien (Mobilfunksysteme) von Personengruppen wahrgenommen werden, d. h. welche Aspekte der Technologie sie thematisieren (Teilthemen) und wie sie diese diskutieren (neutral oder wertend). Die Identifizierung, Erhebung und Analyse themenbezogener sprachlicher Äußerungen (z. B. in Blogkommentaren) erfolgte mit Text-Mining-Methoden. Der Analyse-Fokus richtete sich auf bewertungsindizierende Äußerungen. Im Folgenden werden am Fallbeispiel ‚Themenspezifischer Blogkommentar' Herausforderungen der Erhebung, der Aufbereitung und der Analyse thematischer Blogkommentar-Korpora beschrieben.

3.1.2 Datenerhebung und Korpusbildung

Die *Datenerhebung* ist der erste Schritt der maschinellen Verarbeitung natürlicher Sprache; Fehler und Versäumnisse dieser Phase haben weitreichende Konsequenzen für alle nachfolgenden methodischen Schritte und deren Ergebnisse. Der Fokus der Erhebung variiert je nach Forschungskontext. Im vorliegenden Beispiel gab das Forschungsprojekt nicht nur einen thematischen Fokus vor – es definierte auch den Erhebungszeitraum. Die Datenerhebung und Korpusbildung unterlag gebrauchsmuster- wie auch domänenbezogenen Herausforderungen. Sie werden im Folgenden beschrieben:

Gebrauchsmusterbezogene Herausforderungen
Eine Herausforderung ist der Aufbau eines repräsentativen Korpus. Aus dieser Perspektive sind insbesondere Blogs mit zahlreichen Blogkommentaren interessant, die jedoch schwierig zu finden sind. Blogs weisen tendenziell weniger Kommentare per Artikel auf als z. B. Foren oder Facebook-Themenseiten. Dies hat Konsequenzen für den Suchaufwand. Im Fallbeispiel erfolgt die Suche nach relevanten Blogs über verschiedene frei verfügbare Suchmaschinen (hauptsächlich: google.search, google.blogsearch und yahoo). Die Zusammenstellung der Keyword-Listen erforderte domänenspezifisches Fachwissen; sie erfolgte in enger Zusammenarbeit mit den ingenieurwissenschaftlichen Projektpartnern und umfasste Recherchen in der themenbezogenen Fachliteratur. Beispiele für themenspezifische deutsche Keywords sind Ausdrücke wie *Mobilfunk*, *Handy*, *Kunde*, *Funkmast*, *elektrisches Feld*, *hochfrequentes Feld* oder *elektromagnetisches Feld*. Die Erhebung fokussiert zwei themenspezifische Blogs als Datenbasis bzw. Ausgangspunkt der Bildung von Textkorpora: *www.elektrosmogblog.de* und *www.heise.de/mobil/* (vgl. Tab. 1).

Das Textkorpus *www.elektrosmogblog.de* umfasst 63 thematische Blogartikel und 28 Blogkommentare aus dem Zeitraum Mai bis Juni 2008; es hat einen Umfang von ca. 6.000 Token. Die Textdaten wurden manuell per Copy&Paste erhoben und im txt-Format gespeichert (Trevisan/Jakobs 2010). Das Textkorpus www.heise.de/mobil/ umfasst 2.541 Blogartikel und 166.034 Blogkommentare aus dem Zeitraum Januar 2008 bis Dezember 2009. Es hat einen Umfang von ca. 16.000.000 Token; die Textdaten wurden automatisch erhoben und im txt-Format gespeichert. Alle in Kapitel 3.1 angeführten Datenbeispiele entstammen den genannten Textkorpora.

Tab. 1: Übersicht zu den themenspezifischen Korpora

Textkorpus / Anzahl	www.elektrosmogblog.de	www.heise.de/mobil/
Blogartikel	63	2.541
Blogkommentare	28	166.034
Token	≈6.000	≈16.000.000

In HUMIC wurde das Textkorpus *www.heise.de/mobil/* kriteriengeleitet in Subkorpora überführt. Genutzt wurden folgende Kriterien:

– *Refinementkriterium 1*: Das Kriterium erfasst das Auftreten eines Netzjargon-spezifischen Ausdrucksmittels, hier: das Auftreten der *interaktiven Einheit Emoticon*, z. B. als Indikator für Emphase und Bewertungen (vgl. Beißwenger et al. 2012). Das Subkorpus erfasst nur Blogkommentare, die mindestens ein Emoticon enthalten (vgl. Tab. 2).

Tab. 2: Übersicht Subkorpus I

Blogkommentare	109
Token	10.043

– *Refinementkriterium 2*: Das Kriterium erfasst den Nutzertyp nach Postinghäufigkeit. Für das zweite Subkorpus wurden die Anzahl von Kommentaren pro Blogger ermittelt und anhand dieses Wertes Bloggertypen gebildet und ausgezeichnet (z. B. Metadatum Nutzertyp I „Blogger mit 1 Kommentar“, Metadatum Nutzertyp II „Blogger mit 20 Kommentaren“, etc.). Für jeden

Nutzertyp wurden randomisiert 50 Blogkommentare extrahiert und darauf bezogen die Tokenanzahl pro Nutzertyp ermittelt (vgl. Tabelle 3).

Tab. 3: Übersicht Subkorpus II

Nutzertyp	1	10	20	max.	∑
Token	2.897	3.083	5.362	4.264	15.606

Die Anordnung von Blogartikel und -kommentar in einem Blog (Blogstruktur) kann je nach Content-Management-System variieren, was spezifische Anforderungen an die Datenerhebung stellt. Das Fallbeispiel umfasst Quellen, die unterschiedlich strukturiert sind: in *www.elektrosmogblog.de* erfolgt die Listung der Blogkommentare direkt unter dem Artikel in antichronologischer Reihenfolge (der neueste Beitrag erscheint oben); im Falle von *www.heise.de/mobil/* sind die Blogkommentare zu einem Artikel nur über einen Link erreichbar. Die strukturellen Unterschiede erhöhen den Aufwand der semi- wie auch der vollautomatischen Datenerhebung. Die Erhebungsmethode muss blogabhängig angepasst bzw. modifiziert werden. Wer mehrere Blogs zugleich erheben will, benötigt ein dementsprechend breites Wissen um Strukturierungsprinzipien. Bei einer Grundmenge von 100 Blogs muss die Struktur von 20% der Blogs bekannt sein, um den Crawler so trainieren zu können, dass er 80% der Blogs findet und dort enthaltene Texte (z. B. Blogkommentare) zuverlässig extrahiert (*80/20 principal*, Koch 2011). Ist das 80/20-Prinzip nicht gesichert, treten Fehler auf: Texte werden nicht identifiziert und fehlen in der erfassten Textmenge, sie werden nur zum Teil erkannt und dann unvollständig extrahiert oder es werden mit dem Text nicht-analyserelevante Anteile erhoben (z. B. Ankertexte). Im Falle großer Textkorpora potenzieren sich Fehler über die weiteren Verarbeitungsschritte, da manuelle Korrekturen und Korpusüberarbeitungen nur stichprobenartig möglich und damit bedingt hilfreich sind.

Eine weitere Herausforderung der Datenerhebung ergibt sich aus der potentiell begrenzten Verweildauer von Blogs im Netz. Blogs können aus dem Netz „verschwinden“, etwa wenn der Autor aus persönlichen Gründen den Blog schließt oder wenn der Betreiber feststellt, dass das Blogthema nicht mehr öffentlichkeitsrelevant ist. In diesem Fall kommt der von Hyperlink zu Hyperlink suchende Crawler zu einem Punkt, an dem der Suchprozess automatisch abbricht. Wiederholte Datenerhebungsabfragen, etwa infolge veränderter Fragestellungen, sind nach dem Verschwinden von Blogs nicht mehr möglich. Datenerhebungsverfahren, die auf zeitbezogene Phänomene abheben (z. B.

Trendanalysen) und daher iterative Abfragen erfordern, müssen aus den genannten Gründen regelmäßig in kürzeren Abständen erfolgen. Dies erhöht den personellen Aufwand.

Das Ziel, Korpora verschiedener Forschungsprojekte für spätere Forschungsfragen zusammenzuführen (z. B. für Einzelkorpora übergreifende Analysen), bedingt die Entwicklung eines möglichst leistungsfähigen Konzepts für die Datenbenennung und Langzeitspeicherung, das sich an Anforderungen wie Eindeutigkeit, Flexibilität, Nachhaltigkeit und Aufwand orientiert. Eine Grundanforderung an die Archivierung ist die originäre Sicherung der im Internet erhobenen Daten; sie werden „unbereinigt", d. h. vollumfänglich (inklusive Kontext- und Menüelemente, Navigationslinks, Metadaten etc.) gespeichert. Das Verfahren bietet den Vorteil, dass bei sich veränderndem Forschungsfokus Seitenelemente, die im Augenblick der Datenerhebung nicht analyserelevant erschienen, berücksichtigt werden können. Parallel oder alternativ dazu können bzw. sollten – wie beim Fallbeispiel (vgl. Kap. 3.1.3) – die erhobenen Daten bereinigt (d. h. ohne Zusatzinformationen wie z. B. Ankertexte) in einem reinen Textformat (z. B. .txt) abgespeichert werden, um das Korpus bei Bedarf in verschiedene Formate (z. B. .exb) überführen zu können.

Der Aufbau eines Gesamt-Korpus, das verschiedene Korpora integriert, erfordert eine *einheitliche*, *eindeutige* und *nachhaltige* Benennungssystematik, die Hinweise auf die Textquelle, das Erscheinungsjahr und das Gebrauchsmuster liefert, z. B. „Blogname_Jahr_durchlaufende Nummer.txt". Das Verfahren wurde im Fallbeispiel auf die erhobenen Daten angewandt (z. B. Heise2009_000019651.txt). Alternativ kann das tatsächliche Erscheinungsdatum (des Datums) die fortlaufende Nummerierung ersetzen.

Domänenbezogene Herausforderungen

Aufgaben wie Langzeitarchivierung und Speicherung erfordern die Klärung rechtlicher Fragen. IBK-Forschung setzt den freien und/oder lizenzierten Zugriff auf Korpora voraus, Regeln für den Zugang zu Plattformen und wissenschaftlichen Communities und/oder das Einverständnis des „Datenbesitzers" (Autors, Betreibers eines Dienstes etc.). Die Praxis sieht häufig anders aus. Die Erhebung von Daten in Unternehmenskontexten (z. B. firmenintern genutzte Applikationen) bedingt in der Regel langwierige Klärungsprozesse mit der Unternehmensleitung (eine wesentliche Voraussetzung ist die Zustimmung des Personalrates); häufig müssen die erhobenen Daten nach einem vereinbarten Zeitraum gelöscht werden. Die Freigabe von Daten für die Scientific Community ist generell problematisch. Die beschriebenen Probleme sind ein wesentlicher Grund für die

geringe Anzahl linguistischer oder kommunikationswissenschaftlicher Studien zu diesem Bereich der IBK-Nutzung.

In öffentlichen Domänen ist die Erhebung scheinbar unkompliziert – die Daten werden öffentlich oder halb-öffentlich produziert und rezipiert, was die Annahme nahelegt, dass jeder Forscher diese Daten für seine Zwecke erheben und nutzen darf. Tatsächlich bewegt sich hier IBK-Forschung in einer rechtlichen Grauzone. Äußerungen auf Facebook-Seiten und in anderen Social-Media-Umgebungen gehören den Betreibern und den Verfassern, deren Einverständnis einzuholen wäre, was praktisch aber kaum möglich ist[3]. Infolgedessen ist die *Bereitstellung aufbereiteter Textkorpora* für den Download und die Datenanalyse durch Dritte (z. B. auf forschungsnahen Plattformen wie CLARIN) bisher rechtlich nicht eindeutig geklärt. Fragen, die sich in diesem Kontext ergeben, lauten u. a.: Wie sichere ich Daten für Forschungszwecke (auf einem Einzel-PC, in einer Netzwerkumgebung mit Zugang für andere Forscher)? Wer darf auf was Zugriff haben? In welchem Umfang darf ich Daten zur Wiederverwendung Dritten anbieten? Die aktuelle Rechtslage hat auf diese Fragen nur bedingt Antworten; die anhaltenden Diskussionen zeigen, dass bislang keine abschließende Klärung in Sicht ist.[4]

3.1.3 Datenaufbereitung

Die Datenaufbereitung unterlag im Fallbeispiel gebrauchsmuster- und nutzertypbezogenen, themen- und domänenbezogenen Restriktionen.

Gebrauchsmusterbezogene Herausforderungen

Das Gebrauchsmuster schafft unterschiedliche Ausgangsbedingungen für die Datenaufbereitung (etwa die Bereinigung der Daten und die Anreicherung von Text mit Metadaten, PoS-Tags und Mehrebenen-Annotation). Vom Gebrauchsmuster ist u. a. abhängig, ob die Musterrealisierung Hinweise auf Autor, Quelle und Veröffentlichungsdatum liefert, die als Metadaten in die Datenaufbereitung eingehen können (Datenanreicherung). Die Qualität und/oder die Art und Wei-

3 Die Ausführungen beziehen sich auf einen Vortag von Nikolaus Forgó zum Thema auf der 5. Arbeitstagung des DFG-Netzwerks *Empirikom* an der Universität Hamburg (25.–26.04.2013).

4 Siehe aber Beißwenger et al. (in diesem Band), die über ein Rechtgutachten zur Bereitstellung des *Dortmunder Chat-Korpus* in CLARIN-D und die daraus resultierenden Konsequenzen für die Datenaufbereitung und -repräsentation berichten.

se dieser Hinweise entscheiden maßgeblich, ob und wie sie für Metadaten genutzt werden können und wie belastbar darauf basierende Analyseergebnisse sind. So hat typischerweise jeder Blogkommentar einen Autor, der häufig jedoch nicht anhand des im Blog genannten Autorennamens rekonstruierbar ist, da viele Autoren ihre Identität hinter einem Nickname verbergen und/oder über Nicknames eine neue Identität schaffen (Bolander 2013: 80). Die Auswertung erlaubt daher nur selten Hinweise auf soziale Gruppen. Dies ist im Falle von Social-Networking-Sites (SNS) anders. Hier hinterlassen die Nutzer typischerweise sozio-demographische Daten zu ihrer Person (Trevisan et al. 2014). Herausforderungen betreffen hier eher die automatische Erhebung dieser Daten, da SNS-Anbieter wie *Facebook* ihre Nutzer partiell vor externem Datenmissbrauch schützen, z. B. indem sie Ausleseverfahren abbrechen.

Der Einfluss des Gebrauchsmusters zeigt sich besonders deutlich auf der Ebene des Part-of-speech-Tagging (PoS), des Auszeichnens mit morpho-syntaktischen Kategorien. Bestimmte Gebrauchsmuster, z. B. Zeitungsartikel, folgen weitgehend den Normen des schriftsprachlichen Standards (Trevisan et al. 2013a), weshalb sie gern zum Training von Tools genutzt werden. Von Nutzern verfasste Beiträge, wie z. B. Blogkommentare, variieren dagegen je nach Gebrauchsmuster zum Teil stark – das Spektrum der eingesetzten Mittel reicht von einem stark schriftsprachlich orientierten Duktus bis hin zu mehr oder weniger stark (und intendiert) davon abweichenden Ausdrucksformen (wie etwa fehlende Interpunktion, elliptische Sätze, Netzjargon). Die Abweichungen müssen auf der morpho-syntaktischen Annotationsebene abgebildet werden, z. B. durch das PoS-Tagging von *interaction signs* (Beißwenger et al. 2012 in IBK-spezifischer Erweiterung des Konzepts der ‚interaktiven Einheiten' aus Zifonun et al. 1997). Herausforderungen betreffen insbesondere die Adaption bestehender Tools und Annotationsschemata für Social-Media-Texte. Aktuelle Entwicklungen zur Adaption gängiger Tagger für PoS-Tagging verfolgen zwei Verfahrensweisen: (i) das dem Tagger zugrundeliegende Tagset wird durch Tags für die Auszeichnung Netzjargon-spezifischer Mittel und Ausdrücke erweitert und die Daten werden mit dem erweiterten Tagset ausgezeichnet (Bartz et al. 2013); (ii) die Verwendungsregeln des bestehenden Tagsets werden für die Auszeichnung Netzjargon-spezifischer Mittel und Ausdrücke modifiziert bzw. erweitert und anschließend Daten anhand der neu definierten Tagsets annotiert (WebTagger; Neunerdt et al. 2013a,b).

Social-Media-basierte Gebrauchsmuster erfordern bei der Mehrebenen-Annotation (Trevisan 2014) Anpassungen auf allen linguistischen Ebenen. Sie betreffen u. a. die Annotation von Tilgungen und Klitisierungen (morphologische Ebene), von interaktiven Einheiten und onomatopoetischen Ausdrücken

(graphematische Ebene) oder etwa von Formen des IRONISIERENs (pragmatische Ebene). Ihre Verarbeitung – z. B. für Zwecke der Sentiment Analysis – bedingt spezifische Verfahren der Annotation und Auswertung, die bisher erst in Ansätzen existieren (Trevisan 2014).

Nutzertypbezogene Herausforderungen

Die Bestimmung und Identifikation von Nutzertypen erfordert Kriterien, die je nach Forschungsinteresse variieren. Eine Forschungsintention könnte z. B. sein, Nutzertypen anhand des Aktivitätspotentials zu bestimmen und in einem zweiten Schritt zu prüfen, ob und wie sich das Aktivitätspotential auf das sprachliche und/oder soziale Verhalten auswirkt. Das Aktivitätspotential erfasst zum Beispiel, wie häufig jemand in den erhobenen Blogs Kommentare postet (dies setzt voraus, dass er denselben „Namen" verwendet). Die Kommentierungsfrequenz wird in diesem Fall zum Metadatum. Im Fallbeispiel zeigte sich u. a., dass häufig kommentierende Nutzer hohe Anforderungen an Qualitätskriterien wie Sachlichkeit und sprachliches Ausdrucksvermögen stellen und dementsprechend agieren (vgl. auch Neunerdt et al. 2011). Nutzer mit einem geringen Aktivitätspotential zeichnen sich im Vergleich dazu häufig durch sprachliche Abweichungen von der Norm aus und werden dann ggf. von erfahrenen (hochfrequenten) Kommentierenden zurechtgewiesen. Dazu zwei Beispiele:

Zitat 1: „Doppelt gemoppelt? Entweder "beim" oder "bei dem" oder? Jaja, ich bin gut, ich habe einen Grammatikfehler im Artikel gefunden und fühle mich nun auch besser... Es ist Sonntag, steinigt mich^^ wenn ich mir den roten Balken so ansehe, scheint das zu funktionieren :-D, besser als manches Getue! Aber Du hast recht, ich finde solche Fehler viel zu häufig, um noch daran glauben zu können. dass sich Autoren Mühe gebe. (auf meiner Site gibts auch genug, da bin ich sicher :-))

Zitat 2: Da man im Glashaus nicht mit Steinen werfen sollte, finde ich aber dort auch noch Unverständlichkeiten: "Ich mecker echt ungern über textliche Fehler, aber deine Grammatik sind ja mal unter aller Sau." Es gibt noch viele Legastheniker auf der Welt, und man sollte es denen nicht zum Vorwurf machen. Ich mache auch viele Rechtschreibfehler. Naja, der Originaltext beinhaltete irgendwas mit 'Höherer Umsatz verursachen höhere Kosten.' Jetzt solltest du auch meinen "Fehler" verstehen ^^. Und der ursprüngliche Satz hat nichts mit Legasthenie zu tun, sondern ganz einfach mit falscher Grammatik.

Eine andere Möglichkeit der Nutzertypbildung ist die Einordnung anhand von *Einstellungstendenzen*. In HUMIC erfolgte sie über die manuelle Ermittlung

von Polaritätsindizes der veröffentlichten Beiträge (positive vs. negative Gesamttendenz der Blogkommentare). Die Herausforderung ergibt sich bei diesem Ansatz in der Behandlung divergierender methodischer Anforderungen: die Einordnung eines Nutzers nach Einstellungstendenz erfordert eine ausreichend hohe Anzahl von Äußerungen, die Handhabbarkeit des Ansatzes (da manuell) dagegen eine möglichst geringe Menge.

Themen- und domänenbezogene Herausforderungen
Bei der Datenaufbereitung für Blogkommentare werden themenspezifische Begriffe zu Wortfeldern gruppiert (klassifiziert). Die Wortfelder bilden Einträge in themenspezifischen Lexika, die später z. B. für Frequenz- oder Sentiment-Analysen (vgl. *Datenanalyse*) genutzt werden. Herausforderungen ergeben sich aus der Art der Erstellung von Wortfeldern. Im Fallbeispiel wurden Begriffe aus dem Material extrahiert und zu Wortfeldern zusammengestellt. Die Domänenspezifik des Themas erschwert u. a. das Erkennen von Über- und Unterordnungsrelationen, von Synonymen und Antonymen etc. Das fehlende Wissen muss durch Nachfragen bei Experten oder Recherchen in Fachtexten gedeckt werden, was den zu betreibenden Aufwand deutlich erhöht.

3.1.4 Datenanalyse

In der Phase der Datenanalyse des Fallbeispiels dominierten gebrauchsmuster-, nutzertyp- und themenbezogene Anforderungen die methodische Umsetzung.

Gebrauchsmusterbezogene Herausforderungen
Die Analyse von Blogkommentaren dient im Fallbeispiel der Ermittlung von Einstellungen und Polaritäten der öffentlichen Diskussion zu einem bestimmten Thema (hier: Mobilfunk). Sie erfolgt auf Basis der in Kapitel 3.1.3 beschriebenen Mehrebenen-Annotation. Im Mehrebenen-Annotationsmaterial wird untersucht, welche sprachlichen Indikatoren bzw. Mittel in welcher Kombination eine sprachliche Bewertungshandlung indizieren. Zu diesem Zweck werden statistische Methoden sowie Verfahren der Mustererkennung angewandt (z. B. *Support Vector Machine*). Herausforderungen betreffen unterschiedliche Aspekte: Zum einen hängen sprachliche Muster des Bewertens in hohem Maße vom jeweiligen Gebrauchsmuster ab (z. B. Gutachten vs. Blogkommentar). Der Grad der Musterhaftigkeit sprachlichen Bewertens ist umso höher, je normkonformer sprach-

liche Realisierungsmuster sind (*Ich finde, dass das lächerlich ist*. vs. *LOL*); ihre Beschreibung und automatisierte Identifikation und Analyse wird dadurch erheblich erleichtert. Zum anderen hängt der Grad der Musterhaftigkeit sprachlichen Bewertens von der realisierten Sprachhandlung ab: Die Identifikation und Beschreibung sprachlicher Indikatoren, die die Bewertungshandlungen RHETORISCHES FRAGEN oder IRONISIEREN konstituieren, unterscheiden sich wesentlich in ihrer Eindeutigkeit und Auftretenshäufigkeit (Trevisan 2014).

Nutzertypbezogene Herausforderungen

Die Möglichkeiten der Bestimmung von Nutzer- bzw. Akteurprofilen (*Akteur-Analyse*) und ihre Berücksichtigung in der Analyse hängen wesentlich von dem jeweiligen Gebrauchsmuster und der Domäne ab. In vielen Fällen werden Gebrauchsmuster anonym genutzt (z. B. Blogkommentar), andere Muster erfordern konkrete Angaben zur Person des sich Äußernden (z. B. Facebook). Gleichzeitig kann die Domäne die Vollständigkeit von Nutzer-Profilen einschränken, z. B. im Falle von Intranet-Blogs oder geschlossenen Facebook-Gruppen. Die Aussagekraft von Nutzerprofilen variiert damit abhängig von der Zugänglichkeit sowie der Qualität und Quantität der im Netz zu ermittelnden Nutzerdaten.

Themenbezogene Herausforderungen

Im Fallbeispiel wird bei der Datenanalyse wortfeldbezogen (Kap. 3.1.3) ermittelt, welche Begriffe (z. B. *Mast*, *Strahlen*, *Handy*) wie gewichtet sind. Die Ermittlung der Auftretenshäufigkeit erfolgt über Frequenz-Analysen; sie erlauben Aussagen über Themen- und Bewertungsschwerpunkte der im Netz geführten Diskussion zu einem Thema (Trevisan/Jakobs 2012: 198/199). Herausforderungen ergeben sich insbesondere durch Wortfeldbegriffe, die nicht im Schritt der Datenaufbereitung erfasst wurden (unbekannte Synonyme oder Wortneuschöpfungen, z. B. *Apfel-Phone* statt *iPhone*) und dementsprechend bei der Datenanalyse unberücksichtigt bleiben. Die daraus resultierenden Lücken wirken sich nachteilig auf nachgelagerte Analysemethoden, z. B. *Trend-Analysen*, aus, die auf frequenzanalytischen Voruntersuchungen basieren. Trend-Analysen erheben, wie sich die Wahrnehmung und Bewertung eines Themas und seiner Merkmale im zeitlichen Verlauf verändern (Trevisan et al. 2013b). Sie können zurückliegende Entwicklungen über die Zeit rekonstruieren, Veränderungen der öffentlichen Diskussion und ihrer Themenschwerpunkte zu einem Themenbereich aufzeigen sowie Zeitpunkte und Verschiebungen in der öffentlichen Wahrnehmung und Bewertung eines Themas aufdecken. Werden Themenbe-

griffe nicht erfasst, verschieben sich Themengewichtungen. Aussagen, die sich aus diesen Ergebnissen ableiten, bilden Diskussionen im Internet nicht adäquat ab und können (insbesondere in anwendungsnahen Kontexten) zu falschen Implikationen führen.

3.2 Open-Innovation-Portal

Das zweite Beispiel ist das Gebrauchsmuster „Open-Innovation-Portale mit Community-Funktionen“ (3.2.1). Im Folgenden wird skizziert, welche Herausforderungen dieses an sich komplexe Muster an Methoden und Tools der Datenerhebung (3.2.2), -aufbereitung (3.2.3) und -auswertung (3.2.4) stellt.

3.2.1 Kurzbeschreibung des Gegenstands

Open-Innovation-Portale sind digitale, plattformbasierte Gebrauchsmuster, die an die kommunikative Funktion der Ideengenerierung und -entwicklung geknüpft sind. Unternehmen setzen Open-Innovation-Portale in Innovationsprozessen ein, um Ideen von „außen“ (z. B. von Kunden, Tüftlern und Experten) zu erhalten, die für die Entwicklung neuer Produkte und Services genutzt werden. Das Konzept der *Open Innovation* (*OI*) öffnet den firmeninternen Innovationsprozess, indem er gezielt potentielle Kunden als Wissensressource in den Entwicklungsprozess einbindet.

Zu den Methoden des Open-Innovation-Ansatzes gehören so genannte *Innovationswettbewerbe*. In diesen werden die Teilnehmer aufgefordert, in einem festgelegten Zeitraum zu einem bestimmten Problem bzw. einer ausgelobten Aufgabe Ideen zu entwickeln und einzureichen. Ziel des Verfahrens ist das Sammeln von Bedarfs- und/oder Lösungsinformationen für neue Produkte und Services. Die besten Ideen werden nach Ablauf der Einreichungsfrist durch eine Jury oder Peer Reviews ermittelt und honoriert (Hallerstede/ Bullinger 2010). Innovationswettbewerbe können offline oder online erfolgen; online werden sie als *Open Innovation Portal* (*OIP*) realisiert. OIP unterscheiden sich von offline durchgeführten Wettbewerben durch die Möglichkeit der Integration von Community-Funktionen. Community-Funktionen ermöglichen den Teilnehmern die Bewertung eingereichter Ideen (Vorselektion der besten Ideen) wie auch ihre diskursive Weiterentwicklung (Co-Creation), z. B. durch Kommentierung oder den Austausch von Nachrichten. Der Erfolg von Open-Innovation-Portalen mit Community-Funktionen (COIP) hängt u. a. davon ab, wie es gelingt, Personen zum Einreichen von Ideen zu motivieren und eine aktive Community aufzubau-

en. Die Nutzer der Plattform können auf der Plattform verschiedene Rollen einnehmen: einige reichen nur Ideen ein, andere beschränken sich auf das Kommentieren von Ideen, wieder andere übernehmen beide Rollen.

Die folgende Diskussion stützt sich auf Daten, die in dem interdisziplinären Verbundprojekt *Offene Innovationsplattform für altersbezogene Dienstleistungen* (*OpenISA*)[5] erhoben wurden. Im Projekt wurden vier Open-Innovation-Wettbewerbe als Portale durchgeführt. Das im Folgenden diskutierte Fallbeispiel bezieht sich auf das Open-Innovation-Portal *Einfach Telefonieren*, das als Wettbewerbsidee die Entwicklung eines „Senioren-Mobiltelefons der Zukunft" auslobt und dabei dezidiert ältere Teilnehmer als Ideengeber und Kenner der Zielgruppe adressiert.

Das Portal *Einfach telefonieren* bietet verschiedene Typen von Funktionen an (vgl. Abb. 1 unten): System-bezogene Funktionen für die Anmeldung auf der Plattform und die Registrierung im Wettbewerb, die Wettbewerb-bezogene Funktion „Ideen einreichen" sowie Community-bezogene Funktionen, die partizipative Handlungen ermöglichen, wie das Bewerten und Kommentieren von Ideen, Selbstdarstellung (über Profile) und den Austausch untereinander (Nachrichten).

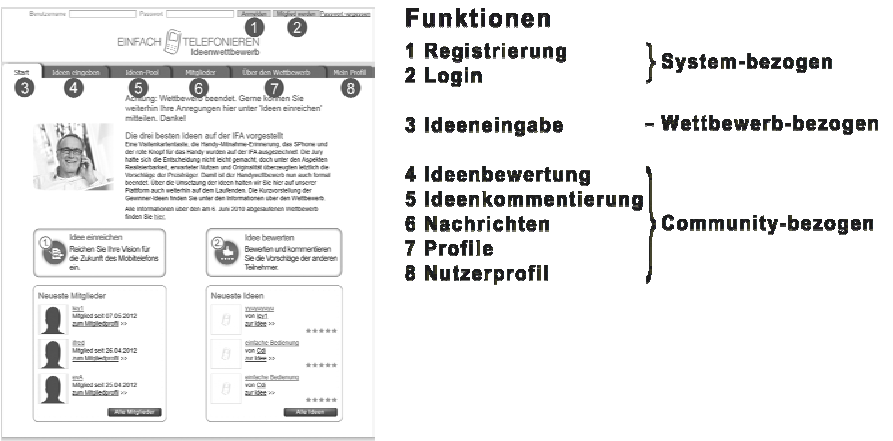

Abb. 1: Funktionen des COIP Einfach Telefonieren

5 Gefördert im Ziel 2-Programm des Bundeslandes NRW aus Mitteln der Europäischen Union: Europäischer Fonds für regionale Entwicklung – Investition in unsere Zukunft.

Die Funktion „Ideen einreichen“ unterstützt verschiedene Formate der Ideeneingabe bzw. -entwicklung (Abb. 2).

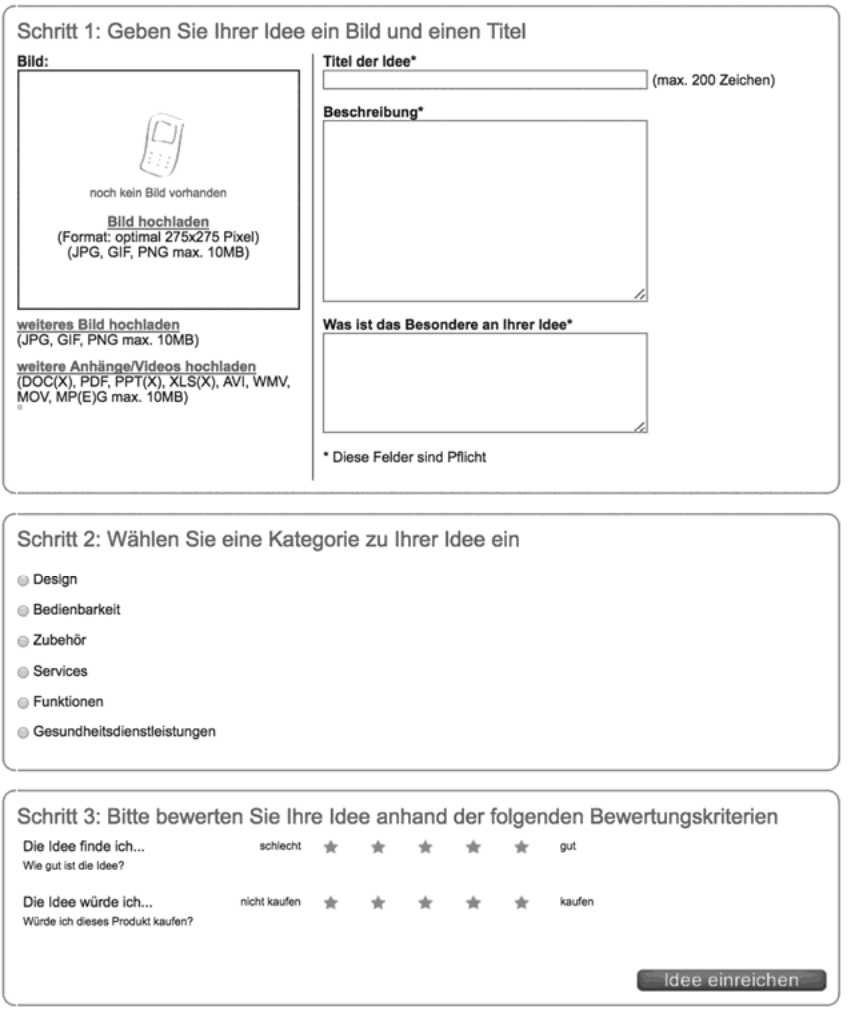

Abb. 2: Eingabeformular für Ideenbeschreibungen

1. Die Wettbewerbsteilnehmer können ihre Idee schriftlich beschreiben und mit Hilfe eines Eingabeformulars hochladen (vgl. Abb. 2). Das Eingabeformular strukturiert die Ideen-Beschreibung durch die Eingabefelder *Titel*, *Beschreibung* sowie *Besonderheiten*. Optional kann der oder die Einreichende die Angaben des Eingabeformulars durch das Hochladen weiterer Datei (-typ-)en ergänzen (Dateien mit Abbildungen, Präsentationsfolien, Tabellenkalkulationen oder Videos). Er bzw. sie wird abschließend gebeten, die eingereichte Idee thematisch einzuordnen. Die Einordnung erfolgt anhand vorgegebener Kategorien (im Fallbeispiel: Aspekte von Mobiltelefonen wie Design, Bedienbarkeit, Zubehör, Services, Funktionen und Gesundheitsleistungen).
2. Die Wettbewerbsteilnehmer können ergänzend oder alternativ mit Hilfe eines interaktiven Baukastensystems (*Toolkit*) – dem Handy-Konfigurator – Designvorschläge entwickeln und einreichen (vgl. Abb. 3). Teil des Gebrauchsmusters ist damit eine spezifische Form maschineller Interaktivität: Der Konfigurator bietet verschiedene Gestaltungskomponenten und -varianten an. Die „Entwickler" können Komponentenausprägungen auswählen und per Drag&Drop für die Entwicklung eines Handy-Designs nutzen. Über zwei Texteingabefelder im Toolkit (Titel und Beschreibungstext) können die Nutzer den Designvorschlag ergänzend verbal beschreiben.

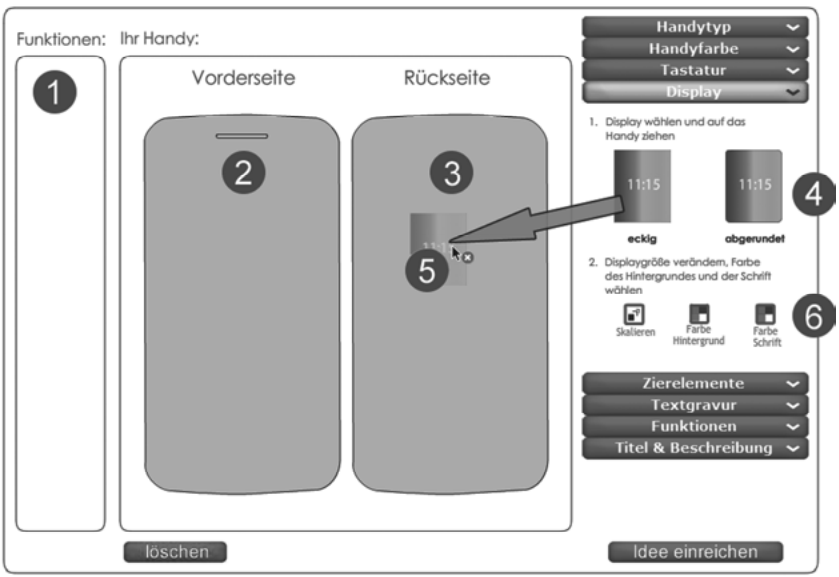

Abb. 3: Das Toolkit „Handy-Konfigurator"

Das Gebrauchsmuster weist methodisch relevante Besonderheiten auf. Sie betreffen den „Lebenszyklus“ des Portals und des ausgelobten Wettbewerbs, die Darstellungsmöglichkeiten für Ideen, das integrierte Toolkit sowie die Interaktion in der Community. Im Fallbeispiel richtete sich das Forschungsinteresse auf den Zusammenhang von Gestaltungsmerkmalen der Plattform und Nutzerverhalten (Welche Funktionen werden wie genutzt? Wie wirkt sich die Gestaltung einer Funktion auf die Nutzung aus? Wo benötigen die Nutzer Unterstützung? Wie interagieren die Community-Mitglieder?). Die Ergebnisse wurden u. a. für die Ableitung von Gestaltungshinweisen für derartige Wettbewerbe genutzt.

3.2.2 Datenerhebung

Im Fallbeispiel wurde eine integrative Methodik genutzt: Um Hinweise auf die tatsächliche Nutzung zu erhalten, wurde die Nutzung der realen Plattform erfasst. Um festzustellen, wie Teile des Gebrauchsmusters und seiner Realisierung von Nutzergruppen (z. B. der Zielgruppe Senior-Experte) wahrgenommen werden, wurden Nutzertests durchgeführt.

Aufzeichnung von Aktivitäten auf der Plattform: Das Forschungsinteresse richtete sich im Fallbeispiel auf eine möglichst umfassende Abbildung der Portalnutzung (sprachliche und nicht-sprachliche Aktivitäten). Die Daten müssen aus der Plattform selbst gewonnen werden. Forscher haben in der Regel keinen Zugriff auf die Erfassungs- und Verwertungsformen der den Wettbewerb betreibenden Unternehmen. Im Fallbeispiel wurden die Nutzeraktivitäten bezogen auf die Hauptfunktionen (vgl. Abb.1) durchgängig (ab Wettbewerbsbeginn) als Logfiles aufgezeichnet; die eingestellten Inhalte (Ideenbeschreibungen, Kommentare und Nachrichten) wurden extrahiert und offline verfügbar gemacht. Die Aufzeichnung der Logfiles war im Fallbeispiel Teil der Forschungsvereinbarungen mit den Partnerunternehmen (d. h. den auslobenden Firmen und dem Portalbetreiber).

Die Extraktion der nutzergenerierten Inhalte ist sehr aufwändig. Im Fallbeispiel wurde das Portal „Seite für Seite“ durchgesehen; die Nutzerbeiträge (Ideenbeschreibung, Kommentare, etc.) wurden manuell identifiziert und auf einem Speichermedium abgelegt. Die manuelle Erhebung ist notwendig, wenn – wie im Fallbeispiel – die Seiten per Javascript erzeugt werden. Die Extraktion ist je nach Datenbestand zeit- und arbeitsintensiv: das Portal *Einfach telefonieren* enthält mehrere Hundert Ideen; jede Idee wird auf einer eigenen Unterseite mit den dazugehörigen Bewertungen und Kommentaren dargestellt.

Themenbezogene Herausforderungen

Ein Thema (z. B. eine eingereichte Idee) kann in verschiedenen Datei-Formaten (.JPG, .GIF, .PNG, .DOC(X), .PDF, .PPT(X), .XLS(X), .AVI, .WMV, .MOV, .MP(E)G) beschrieben und eingereicht werden (vgl. Abb. 4). Zu den Herausforderungen der Datenerfassung gehört, alle zu einer Idee gehörenden Dateien vollständig zu erfassen.

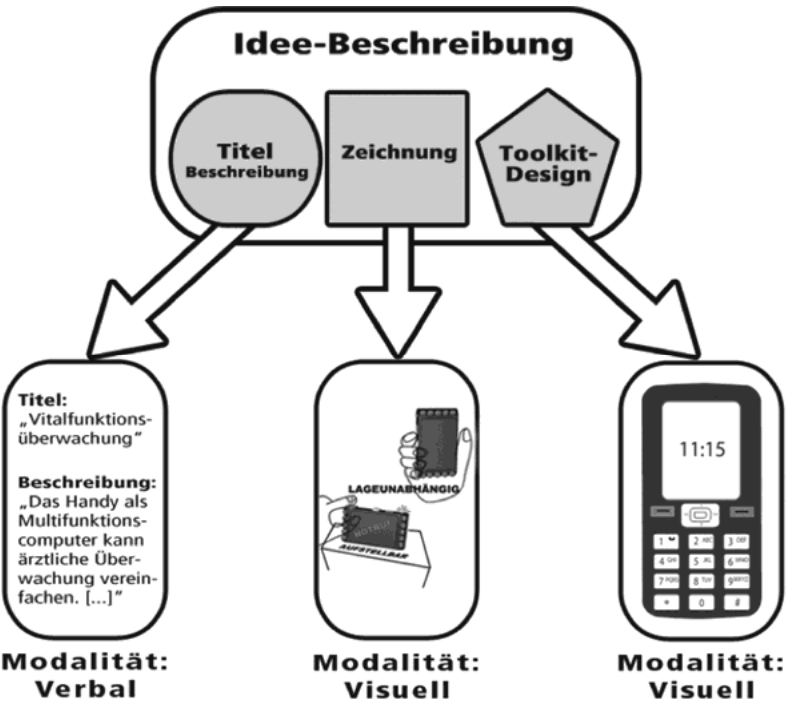

Abb. 4: „Verteilte" Themen (splitted topics)

Je nach Gegenstandsbereich (Ideen für Handy-Design, -Funktionen, -Services) präferieren die Teilnehmer eher eine Modalität (textuelle, visuelle, interaktive Darstellung) oder eher ihre Kombination. Während Textteile im Browser relativ einfach mit der Option „Seite speichern unter..." heruntergeladen werden können, vergrößert die Sammlung von Bildern, Videos und Dokumenten den Erhebungsaufwand deutlich. Einige Dateiformate lassen sich nur mit Mehraufwand (z. B. manuelles Herunterladen von Dokumenten), andere nur mit Hilfslösungen erfassen (z. B. Abfilmen von Videos).

Gebrauchsmusterbezogene Herausforderungen

Die Erhebung des COIP-Datenbestands unterliegt *zeitlichen* Restriktionen. Wird das Portal nach Wettbewerbsende aus dem Netz genommen (offline geschaltet), ist der Zugriff nicht mehr möglich. Ideal wäre, den Prozess der Erzeugung von Inhalten kontinuierlich erfassen zu können, was in der Regel nicht machbar ist. Eine (im Fallbeispiel praktizierte) Alternative ist die punktuelle, zeitlich begrenzte Extraktion von Inhalten, die in einem engen Zeitkorridor kurz vor Wettbewerbsende erfolgen muss, um Ansprüchen zu genügen wie Datenreichtum (möglichst viele eingestellte Nutzerbeiträge) und Vermeiden von Datenlücken (möglichst wenige nicht erfasste Beiträge). Wesentlich ist ein gutes Erhebungskonzept; Erhebungsfehler können im Nachhinein kaum kompensiert werden.

Das im COIP *Einfach Telefonieren* erhobene Datenkorpus umfasst 372 Ideen (268 textuell beschriebene Ideen und 104 mit dem Toolkit produzierte Baukasten-Ideen). Zu den Ideen wurden 713 Kommentare abgegeben – 581 zu textuellen Ideenbeschreibungen, 132 zu Baukastenideen (vgl. Tab. 4, unten).

Tab. 4: Daten-Übersicht COIP Einfach Telefonieren

	Textuelle Ideenbeschreibung	Baukasten-Idee	Gesamt
Ideen (Anzahl)	268	104	372
Kommentare (Anzahl)	581	132	713

Nutzertypbezogene Herausforderungen

Hinweise auf Nutzerdaten liefert die Logfile-Registrierung von Zugriffen auf die Unterseiten des Portals. Sie erfordert – wie erwähnt – die Zustimmung des Portalbetreibers. Bei der Logfileregistrierung werden automatisiert die IP-Adresse des Nutzers, Datum, Uhrzeit, URL und weitere Parameter gespeichert. Die automatisierte Erfassung von Nutzerdaten muss vom Portalbetreiber vor der COIP-Freischaltung eingerichtet werden, um Datenverluste vermeiden und Nutzeraktivitäten bezogen auf Wettbewerbsphasen erfassen zu können. In der Anfangsphase des Wettbewerbs finden sich z. B. besonders viele Aneignungshandlungen (Bucher 2004), die der Erschließung des COIP-Angebotes bei Erstkontakt (*initial contact situation*, Wirtz/Jakobs 2013) dienen (z. B. sich registrieren, sich Überblick verschaffen). Wird eine Phase nicht über Logfiles erfasst, müssen

Phänomene, die diese auszeichnen, aufwändig (z. B. in Nutzertests) rekonstruiert werden.

Die Erhebung von Logfiles für nutzerbezogene Analysezwecke ist ein sensibler Punkt: Da das Logfile Daten enthält, die Rückschlüsse auf den Portalnutzer zulassen (IP-Adresse, persönliche Daten aus der Registrierung, u. a.), kann der Portalbetreiber die Herausgabe von Logfiles aufgrund rechtlicher (und ethischer) Bedenken verweigern. Eine Form des Datenschutzes ist die Anonymisierung der Daten (z. B. durch Nummerierung der IP-Adressen), die jedoch Mehraufwände für beide Seiten erzeugt: Der Betreiber muss im Logfile Nutzerdaten anonymisieren, der Forscher muss bestehende Logfileanalysetools der neuen Formatierung anpassen.

Wie Nutzer die Gebrauchsmusterrealisierung (z. B. Gestaltungseigenschaften der Wettbewerbsfunktionen und des Toolkits) wahrnehmen und bewerten, kann nicht aus authentischen Daten abgeleitet werden (nur, wenn sich Nutzer zu diesem Thema in der Community äußern). Eine methodische Alternative bieten Nutzertests, in denen Zielgruppenvertreter COIP-typische Aufgaben bearbeiten und verbal kommentieren. Gebrauchsmusterspezifische Herausforderungen beginnen spätestens bei der Rekrutierung von Testpersonen. Wenn die Zielgruppe – wie im Fallbeispiel – die der „älteren Tüftler“ ist, entfällt die häufig in der Forschung praktizierte Lösung, für Analysezwecke Studierende zu rekrutieren. Der Anspruch, Testaufgaben realitätsnah zu gestalten, ist ebenfalls schwer umsetzbar; die wenigsten erreichbaren älteren Testpersonen sind „Erfinder“. Bestimmte Anteile der Portalnutzung, wie das Agieren in Online-Communities (kommentieren, interagieren, auf andere eingehen), lassen sich nur bedingt simulieren.

3.2.3 Datenaufbereitung

Die extrahierten Inhalte werden in eine Datenbank überführt. Ihre Ablage erfordert ein konsistentes, verständliches und prägnantes Benennungsschema, um spätere Suchvorgänge und Analysen effizient durchführen zu können (z. B. das Ermitteln des durchschnittlichen Zeitraums zwischen Ideeneinreichung und erstmaligem Auftreten eines bestimmten Kommentartyps, z. B. „Verbesserungsvorschlag“). Die Ablage erfordert ein schlüssiges relationales Datenbankkonzept: Zu jeder Tabelle muss ein angemessenes Relationsschema entwickelt und die Tabelle zu anderen Tabellen in Beziehung gesetzt werden (Kemper/ Eickler 2004). Eine von mehreren Herausforderungen ist, Anforderungen späterer Ana-

lysen vorausschauend zu berücksichtigen: Fehler bzw. Defizite des Datenbankkonzepts verursachen zeitintensive Nachbesserungs- bzw. Erweiterungsarbeiten.

Gebrauchsmusterbezogene Herausforderungen
Sind die Unterseiten eines COIP vollständig erhoben, müssen die mit HTML ausgezeichneten Beschreibungen und Kommentierungen von Ideen und die darin enthaltenen Angaben (u. a. Autor, Datum, Referenzen zu Dateien) vom so genannten Boilerplate (HTML-Tags, PHP-Skripte, etc.) getrennt werden. Die Trennung ist aufgrund des konsistenten Seitenaufbaus automatisiert per Skript möglich: Bestimmte Bestandteile des HTML-Quelltexts (HTML-Tags) umschließen nutzergenerierte Beiträge, die anhand der Tags leicht identifiziert und extrahiert werden können. Die erfassten Inhalte werden mit Metadaten beschrieben. Sie beziehen sich auf die Art des Inhalts (z. B. Idee versus Kommentar) sowie die Beziehungen zwischen Inhalten (z. B. X gehört zu Idee Y; X ist Kommentar zu Kommentar Y). Die Auszeichnung mit diesen Metadaten ist nur manuell möglich.

Themenbezogene Herausforderungen
Die Erfassung der Beziehung zwischen extrahierten Inhalten ist insofern wichtig, als – wie oben dargestellt – die Beschreibung von Ideen häufig über mehrere Typen von Dateien und Dokumenten eines Nutzers verstreut erfolgen kann (splitted topic description) bzw. in den Kommentaren anderer Portalnutzer weiterentwickelt wird.

Domänenbezogene Herausforderungen
Wenn die Metadatenbeschreibung Hinweise auf die Inhalte der eingereichten Ideen liefern soll (z. B. Verschlüsselungsverfahren für Mobiltelefone oder typische Hardware-Komponenten), erfordert dies in der Regel domänenspezifisches Wissen, das Zusatzaktivitäten bedingt, wie den Austausch mit Experten. Die konsistente Vergabe derartiger domänenspezifischer Metadaten bedingt den Aufbau und die Pflege eines Lexikons.

Nutzertypbezogene Herausforderungen

Analyseaufgaben wie die Identifikation von Vertretern der primären Zielgruppe (hier: Seniorexperte) oder *Lead Usern* erfordern die Verfügbarkeit nutzerspezifischer Daten wie Pseudonym, Name, Geschlecht, Alter, Nutzertyp (z. B. bestimmt nach der Anzahl eingereichter Ideen oder Kommentare pro Nutzer), die in der Datenbank als Metadaten erfasst werden. Die Ermittlung nutzerbezogener Daten stößt an ihre Grenzen, wenn Nutzer bei der Registrierung erfundene Selbstauskünfte geben, wenn sie keine ergänzenden freiwilligen Angaben machen oder wenn nicht alle im COIP registrierten Daten öffentlich zugänglich sind.

Eine andere Herausforderung betrifft die Aufbereitung von Logfiles: Die Logfiles müssen für die Auswertung zunächst bereinigt werden, d. h. automatisierte Zugriffe von Suchmaschinen und nicht intendiertes Nutzungsverhalten (z. B. Hackerangriffe) identifiziert und entfernt werden. Die Unterscheidung von tatsächlichen COIP-Nutzern, Suchmaschinencrawlern und Personen, die das COIP zweckentfremdet nutzen, erfordert eine umfangreiche Analyse des Logfiles. Der Umfang ergibt sich aus der gebrauchsmusterbedingt hohen Anzahl von Logfileeinträgen sowie der Notwendigkeit, jeden Fileeintrag manuell zu überprüfen. Die Identifikation nicht analyserelevanter Fileeinträge erfolgt anhand von Bestandteilen, die sich auf den Zugreifenden (z.B. IP-Adresse) beziehen sowie den Zugriffsort im Portal (z. B. robots.txt, Adminbereich). Das Aufbereiten der Logfiles ist wichtig für die Qualität von Analyseergebnissen (z. B. für den Vergleich von Unterseiten nach der Häufigkeit ihres Aufrufs).

3.2.4 Datenanalyse

Gegenstand der Datenanalyse ist die Betrachtung von Nutzungsaktivitäten in der Zeit, des Zusammenhangs von Design und Portalnutzung, die Rekonstruktion der Entstehung von Inhalten (z. B. über die Analyse von Idee-Kommentar-Strängen), das Erfassen gebrauchsmusterspezifischen Handelns (z. B. die Tonalität von Ideenbeschreibungen, Kommentaren und Nachrichten) sowie die Ermittlung von Unterstützungsbedarf (etwa bezogen auf das Toolkit). Ein anderer analyserelevanter Komplex betrifft die Formen und Regeln des sprachlich-sozialen Agierens und Interagierens in der Plattform-Community (z. B. über die Analyse intertextueller Äußerungen und/oder von Äußerungen, die das soziale Verhalten der Plattformnutzer kommentieren).

Gebrauchsmusterbezogene Herausforderungen

Die Analyse der sich in den Wettbewerbsphasen anteilig verändernden Nutzeraktivitäten erfolgte anhand der Logfileeinträge, sie nutzte die dort enthaltenen zeitlichen Metadaten (*time stamps*). Im Fallbeispiel zeigte sich, dass in frühen Phasen die Aneignung des Systems das Einstellen von Ideen und der Aufbau der Community im Vordergrund stehen, in späten Phasen das Diskutieren, Interagieren und Weiterentwickeln eingereichter Ideen in der Community.

Logfileanalysen liefern u. a. Hinweise darauf, wie sich Gestaltungsmerkmale auf das Nutzerverhalten auswirken. Im Einzelnen wurde geprüft, wo (auf welchen Unterseiten) die Nutzer bevorzugt in das Portal einsteigen bzw. aussteigen, welche Funktionen und Inhalte häufig aufgerufen werden und welche nicht sowie wo besonders viele „Rückschritte" auf dem Bewegungspfad auftreten (z. B. als Indikator für Nutzungsprobleme und „sensible" Gestaltungsanteile). Die Analysen ergaben u. a., dass 91,76% der Nutzer den Portalbesuch beim Registrierungsprozess abbrachen. Die Gründe dafür zeigten sich erst im Nutzertest. Die Testpersonen äußerten ein massives Unbehagen, persönliche Daten preiszugeben und Angst vor Datenmissbrauch (Digmayer/Jakobs 2012a).

Wie das Beispiel zeigt, erschließen sich Phänomene oft erst in der Kombination quantitativer und qualitativer Methoden. Rein quantitative Angaben sind häufig interpretationsbedürftig. Im Fallbeispiel ergab die Logfileanalyse für den Toolkit-Handy-Konfigurator geringe Zugriffe und zahlreiche Rückschritte in den Nutzerpfaden. Die Gründe zeigten sich erst im Nutzertest. Das Tool überforderte insbesondere ältere Nutzer (durch zu viele Gestaltungsoptionen und unklare Handlungssequenzen); sie brachen deshalb häufig die Nutzung ab. Eine anschließende Testreihe mit multimodalen eTutorials ergab, dass diese COIP-Nutzer in komplexen interaktiven Nutzungssituationen wirkungsvoll unterstützen können (z. B. durch Handlungsabfolgen begleitende verbale Anweisungen und ihre Visualisierung; Digmayer/Jakobs 2012 b).

Themenbezogene Herausforderungen

Teil des Gebrauchsmusters ist die Erwartung, dass die Teilnehmer die auf der Plattform eingereichten Ideen bewertend kommentieren. Teil der Analyse sprachlicher Bewertungshandlungen (Ripfel 1987, Sandig 2003) ist die Identifizierung des Bewertungsgegenstands (Was wird bewertet?) bzw. -aspektes (Welche seiner Eigenschaften wird bewertet?), von Vergleichsgrößen (Womit wird er verglichen?) und Bewertungsmaßstäben. Die Rekonstruktion des Bewertungsgegenstandes bzw. -aspektes wird erschwert, wenn ein Thema (hier: Idee) – wie oben dargestellt – auf mehrere Dateien verteilt beschrieben wird und/oder

wenn sich die Bewertung auf Sachverhalte bezieht, die auf externen Websites expliziert werden (Beispiel: „Die Idee ist nicht neu, siehe hier [Link]“). Andere Herausforderungen ergeben sich im Falle „unvollständiger Bewertungshandlungen“, z. B., wenn das Bewertete nicht benannt wird (Beispiel „Das finde ich toll“).

Die themenbezogene Analyse der Nutzerbeiträge indiziert, dass das sprachliche Verhalten der Teilnehmer themen- und rollenbezogen variiert (vgl. Digmayer 2016). Die Beschreibung von Ideen erfolgt primär in einem sachlich-neutralen Stil, an Normen des Schriftsprachlichen orientiert. In den Kommentaren variiert der Ton; er kann – z. B. im Falle eigener Betroffenheit – emotional gefärbt sein und vom Standardsprachlichen abweichen, wie das Zitat im folgendem Beispiel illustriert: Der Verfasser wehrt sich gegen den Vorwurf, eine Idee von einem anderen Nutzer übernommen zu haben. Der Ton des Kommentars weicht merklich von den sachlichen Ideenbeschreibungen ab.

Zitat: „Ich glaube nicht, dass ich es nötig habe irgendetwas "abzukupfern" ;-)=) - es kann sein, dass auch andere ähnliche Ideen haben - na und. Die meisten Sachen geistern mir schon seit Jahren(!) im Kopf herum und jetzt werden sie niedergeschrieben. Ein anderer hat's zuerst veröffentlicht - na und ? Ist die Idee deshalb schlecht? Wenn's jetzt um den Nachweis geht, wer hat was wann zuerst ... naja, das interessiert MICH an dieser Stelle herzlich wenig.“

Nutzertypbezogene Herausforderungen

Bezogen auf die Weiterentwicklung von Ideen (Co-Creation) richtete sich das Forschungsinteresse auf die Ermittlung von Nutzertypen. Im Fallbeispiel erfolgte die Bildung nach den Kriterien Häufigkeit der eingereichten Ideen und Kommentare. Die Zuordnung zu Typen ist anhand der Angaben in der Datenbank maschinell möglich. Die Analyse ergab vier Typen: Nutzer, die primär Ideen einreichen (Erfinder), Nutzer, die primär kommentieren (Kommentatoren), Nutzer, die Ideen einreichen wie auch kommentieren (kommentierende Erfinder) sowie Passive, die Aktivitäten im Portal lediglich beobachten. Alternativ sind andere Kriterien denkbar, z.B. die Einteilung nach dem Interaktionsverhalten (z. B. Reagieren auf andere Kommentare und/oder auf das Verhalten anderer Nutzer) oder linguistischem Profil (z. B. Grad der Schriftsprachlichkeit, Tonalität). Die Nutzertypermittlung nach diesen Kriterien ist nur manuell möglich.

Domänenbezogene Herausforderungen: Die Identifikation von Nutzertypen mit hohem Innovationspotential (Lead Usern) ist insbesondere für Unternehmen von Interesse, sie erfordert Parameter, die nur in der Zusammenarbeit mit Domänen-Experten definierbar sind: Neben Postinghäufigkeit, Interaktionsver-

halten und linguistischem Profil muss das Domänenwissen des Nutzers bewertet werden. Dies ist ohne Experten des jeweiligen Wissensgebiets kaum möglich.

4 Fazit

Vergleicht man die oben beschriebenen IBK-Beispiele und die damit verbundenen methodischen Herausforderungen, so zeigen sich Gemeinsamkeiten wie Unterschiede, wobei die Unterschiede deutlich überwiegen. Gemeinsamkeiten ergeben sich u. a. bezogen auf den Bedarf nach leistungsfähigen Systemen für das Datenmanagement, z. B. flexible und zugleich belastungsfähige Benennungs- und Metadatensysteme. Unterschiede des methodischen Herangehens zeigen sich in allen Phasen – von der Datenerhebung bis zur Datenanalyse. Sie ergeben sich zum einen durch Charakteristika der Gebrauchsmuster, zum anderen durch nutzer-, domänen- und themenbezogene Herausforderungen.

Datenerhebung: Methodisch relevante Unterschiede ergeben sich bereits aus dem Gegenstand: im Falle thematischer Blogkommentare interessieren nicht nur die Kommentare eines Blogs, sondern alle Beiträge zu einem Thema. Die Erhebung erfolgt dementsprechend verteilt über mehrere Blogs (als Quellen; vgl. Abb. 5). Im Falle von COIP hat man eine Quelle, in der die Themenbehandlung in verschiedenen, aufeinander bezogenen Bereichen erfolgt (Ideeneinreichung vs. Ideenkommentierung/-weiterentwicklung in der Community). Im Falle des thematischen Kommentars erfolgt die Themenbehandlung in einem Text, der sich auf andere beziehen kann, jedoch nicht muss. Dies gilt ähnlich für die COIP-Kommentare; in anderen Teilen des COIP (Ideeneinreichung) erfolgt die Themendarstellung dagegen häufig verteilt auf verschiedene Dateien und Modalitäten. Im Falle thematischer Blogs dominiert eine kommunikative Ressource – geschriebene Sprache.

Bei beiden Gebrauchsmustern sind unterschiedliche Dateitypen zu berücksichtigen. Blogkommentare können in drei verschiedenen Formaten (.html, .jpeg, .avi) vorliegen und zwei Modalitäten (textuell vs. audio-visuell). Die Heterogenität der zu erhebenden Datenformate (vgl. Abb. 5 unten) ist im Falle von COIP wesentlich höher; sie umfasst 16 Datenformate und drei Modalitäten (textuell, visuell, audio-visuell).

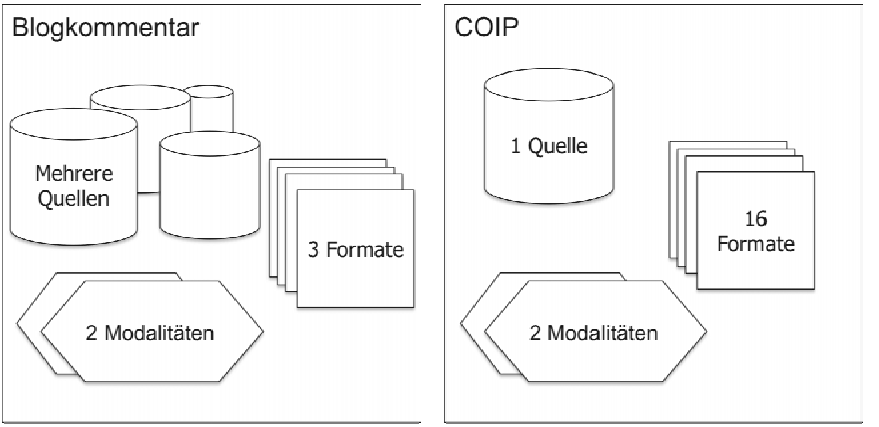

Abb. 5: Gegenüberstellung: Blogkommentar vs. COIP

Die Herausforderungen der Datenerhebung variieren *gebrauchsmusterbezogen*. Bei beiden betrachteten Gebrauchsmustern repräsentieren Musterrealisierungen zeitlich-begrenzte kommunikative Ereignisse. Ein daraus für die Datenerhebung resultierendes Risiko ist die (unangekündigte) Beendigung des Ereignisses (Offline-Schaltung durch den Betreiber). Das Eintreten einer Offline-Schaltung hat gebrauchsmusterspezifische Konsequenzen: Im Falle einzelner abgeschalteter Blogs (dass alle Blogbeiträge zu einem Thema zeitgleich abgeschaltet werden, dürfte eher selten der Fall sein) kommt es zum partiellen Verlust von Daten. Wird ein COIP geschlossen, gehen alle Portaldaten verloren.

Auch die Aufwände der Datenerhebung variieren gebrauchsmusterspezifisch. Im Falle *thematischer Blogkommentare* erhöhen unterschiedlich strukturierte Quellen den Aufwand der semi- wie auch der voll-automatischen Datenerhebung durch die notwendige Anpassung der Erhebungsmethodik. Bei COIP, bei denen die Unterseitenerzeugung mit Java-Script erfolgt, ist die Datenerhebung (mit Ausnahme von Logfileaufzeichnungen) nur manuell möglich und dadurch sehr aufwändig.

Datenaufbereitung: Auch hier unterscheiden sich die Aufwände *gebrauchsmusterbezogen*. Im Falle *thematischer Blogkommentare* müssen Ankertexte entfernt werden. Im Falle von COIP ist der Aufwand höher: es müssen irrelevante Einträge identifiziert und bereinigt sowie Inhalte von der Boilerplate getrennt werden. Je nach Gebrauchsmuster ergeben sich andere zeitliche und inhaltliche Aufwände. Der Aufwand ist bei Blogkommentaren geringer im Vergleich zu COPI, da Metadaten, die durch den Blogbetreiber bereitgestellt werden, genutzt werden können. Bei COIPs werden keine Daten durch den Betreiber bereitge-

stellt; sie müssen vom Daten Aufbereitenden selbst entwickelt und den Datensätzen zugeordnet werden. Die Datenaufbereitung thematischer Blogkommentare erzeugt Mehraufwände durch Aufbereitungsschritte wie Annotation (PoS Tagging, semi-automatische Mehrebenen-Annotation), Datenanreicherung (Lexikon) und Datenverfeinerung (Bildung von Subkorpora). Im Falle von COIPs entfällt die Annotation.

Datenanalyse: Die stärksten Unterschiede zeigen sich – gebrauchsmusterbedingt – in der Analysephase. Sie wurden umfangreich in Kapitel 3.1.4 und 3.2.4 beschrieben.

5 Ausblick

Insgesamt zeigt sich, dass gerade im methodischen Bereich in naher Zukunft viel an Forschungsarbeit zu leisten ist. Dies zeigten u. a. die Diskussionen des Empirikom-Netzwerkes. Der gegenstandsbedingt starke Bedarf nach computergestützten Erhebungs-, Aufbereitungs- und Analysemethoden bzw. -tools erfordert eine intensive Zusammenarbeit von (angewandter) Linguistik und Informatik bzw. informatiknahen Disziplinen, etwa Texttechnologie und Computerlinguistik. Wie aktuelle Studien zur Verbindung traditioneller Forschungsmethoden (etwa der linguistischen Diskursanalyse) mit neuen computergestützten Verfahren (etwa des Textmining bzw. NLP) zeigen (u. a. Niehr et al. 2015), gibt die verschränkte Methodenentwicklung wichtige Impulse für beide Seiten. Generell ist davon auszugehen, dass von der Zügigkeit methodischer Innovationen abhängen wird, wie schnell wir eine genauere Vorstellung davon haben, wie sich die Landschaft der internetbasierten Kommunikation gestaltet, was sie auszeichnet und wie sie sich verändert.

Literatur

Alby, Tom (2008): Web 2.0. Konzepte, Anwendungen, Technologien. München: Hanser.

Bartz, Thomas et al. (2013): Optimierung des Stuttgart-Tübingen-Tagset für die linguistische Annotation von Korpora zur internetbasierten Kommunikation: Phänomene, Herausforderungen, Erweiterungsvorschläge. In: *JLCL* 28(1), 155–198. Online unter: http://jtei.revues.org/476 (21.07.2017).

Beißwenger, Michael et al. (2012): A TEI Schema for the Representation of Computer-mediated Communication. In: Journal of the Text Encoding Initiative (jTEI) 3. Online unter: http://jtei.revues.org/476 (21.07.2017).

Beißwenger, Michael (2013): Das Dortmunder Chat-Korpus. In: Zeitschrift für germanistische Linguistik 41, 161–164.
Beißwenger, Michael, Harald Lüngen, Jan Schallaböck, John H. Weitzmann, Axel Herold, Pawel Kamocki, Angelika Storrer und Julia Wildgans (in diesem Band): Rechtliche Bedingungen für die Bereitstellung eines Chat-Korpus in CLARIN-D: Ergebnisse eines Rechtsgutachtens.
Brinker, Klaus (2010): Linguistische Textanalyse. Eine Einführung in Grundbegriffe und Methoden. Berlin: Erich Schmidt Verlag.
Bolander, Brook (2013): Language and Power in Blogs: Interaction, disagreements and agreements. John Amsterdam: Benjamins Publishing Company.
Brinker, Klaus (2010): Linguistische Textanalyse. Berlin: Erich Schmidt.
Brommer, Sarah und Christa Dürscheid (2012): Mediennutzung heutiger Jugendlicher – Generation Facebook? In: Thema Deutsch, Band 12: Sprache der Generationen, 271–293.
Bucher, Hans-Jürgen (2004): Online-Interaktivität – ein hybrider Begriff für eine hybride Kommunikationsform. Begriffliche Klärungen und empirische Rezeptionsbefunde. In: Bieber, Christoph und Claus Leggewie (Hrsg.): Interaktivität. Ein transdisziplinärer Schlüsselbegriff. Frankfurt am Main: Campus Verlag, 132–167.
Digmayer, Claas (2016): Communitybasierte Open Innovation – Plattformen für ältere Nutzer. Kommunikative Usability, Sociability und eTrust. Dissertation. RWTH Aachen University.
Digmayer, Claas und Eva-Maria Jakobs (2012a): Innovationsplattformen für Ältere. In: Marx, Konstanze und Monika Schwarz-Friesel (Hrsg.): Sprache und Kommunikation im technischen Zeitalter. Wieviel Internet (v)erträgt unsere Gesellschaft? Berlin, Boston: de Gruyter, 143–165.
Digmayer, Claas und Eva-Maria Jakobs (2012b): Help Features in community-based Open Innovation Contests. Multimodal Video Tutorials for the Elderly. In: Proceedings of the 30th International ACM SIGDOC Conference.
Digmayer, Claas und Eva-Maria Jakobs (2012c): Interactive Video Tutorials as a Tool to remove Barriers for Senior Experts in Online Innovation Contests. In: Proceedings of the 6th International Technology, Education and Development Conference (INTED), 5407–5416.
Digmayer, Claas und Eva-Maria Jakobs (2014): Corporate Lifelong Learning 2.0: Design of Knowledge Management Systems with Social Media Functions as Learning Tools. In: Proceedings of the IEEE International Professional Communication Conference 2014, Pittsburgh (USA), 123–131.
Digmayer, Claas et al. (2015a): Designing Mobility Apps to Support Intermodal Travel Chains. In: Proceedings of the ACM SigDoc 2015, 16.–17.07.2015, Limerick (IRL).
Digmayer, Claas et al. (2015b): Medusa and Pandora meet the Web 2.0: How Risk Types influence the Communication in Social Media. In: Proceedings of the ProComm 2015, 12.–15.07.2015, Limerick (IRL): 171–178.
Dürscheid, Christa (2005): Medien, Kommunikationsformen, kommunikative Gattungen. In: Linguistik online 22, Heft 1.
Dürscheid, Christa et al. (2010): Wie Jugendliche schreiben. Schreibkompetenz und neue Medien. Berlin/New York: De Gruyter.
Habscheid, Stephan (2011): Textsorten, Handlungsmuster, Oberflächen: Linguistische Typologien der Kommunikation. Berlin, New York: de Gruyter.
Hallerstede, Stefan und Angelika Bullinger (2010): Do You Know Where You Go? A Taxonomy of Online Innovation Contests. In: Proceedings of the XXI ISPIM Conference 2010.

Holly, Werner (2011): Medien, Kommunikationsformen, Textsortenfamilien. In: Habscheid, Stephan (Hrsg.): Textsorten, Handlungsmuster, Oberflächen. Linguistische Typologien der Kommunikation. Berlin, Boston: de Gruyter (=de Gruyter Lexikon), 144–163.

Jakobs, Eva-Maria (2003): Hypertextsorten. In: Zeitschrift für germanistische Lingusitik 31 (Themenheft "Deutsche Sprache im Internet und in den neuen Medien"), 232–273.

Jakobs, Eva-Maria (2011): Hypertextuelle Kommunikate. In: Moraldo, Sandro M. (Hrsg.): Neue Sprach- und Kommunikationsformen im WorldWideWeb. Medialität, Hypertext, digitale Literatur. Rom: Aracne, 57–79.

Jakobs, Eva-Maria (2012): Kommunikative Usability. In: Marx, Konstanze und Monika Schwarz-Friesel (Hrsg.): Sprache und Kommunikation im technischen Zeitalter. Wieviel Internet (v)erträgt unsere Gesellschaft? Berlin, Boston: de Gruyter, 119–142.

Kemper, Alfons und André Eickler (2004): Datenbanksysteme. Eine Einführung. München: Oldenburg.

Koch, Richard (2011): The 80/20 Principle: The Secret to Achieving More with Less. New York: Doubleday.

Korioth, Daniel (2011): Microblogging in der Unternehmenskommunikation. Eine sprachlich-kommunikative Untersuchung der Kommunikationsform Tweet am Beispiel der DAX30-Unternehmen. Magisterarbeit. RWTH Aachen University.

Nardi, Bonnie A. et al. (2004): Why we Blog. In: Communication of the ACM 47, 41–46.

Neunerdt, Melanie et al. (2011): Ontology-based corpus generation for web comment analysis. In: Proceedings of the HT 2011, 6.–8. Juni, University of Technology (TU/e), Eindhoven (NL).

Neunerdt, Melanie et al. (2013a): Focused Crawling for Building Web Comment Corpora. In: Proceedings of the Consumer Communications and Networking Conference (CCNC, IEEE), 676–679.

Neunerdt, Melanie et al. (2013b): Part-of-Speech Tagging for Social Media Texts. In: Proceedings of the International Conference of the German Society for Computational Linguistics and Language Technology (GSCL) 2013, 139–150.

Niehr, Thomas et al. (2015): Neue Wege der linguistischen Diskursforschung. Computerbasierte Verfahren der Argumentanalyse. In: Zeitschrift für Diskursforschung (2), 113–136.

Ripfel, Martha (1987): Was heißt bewerten? In: Deutsche Sprache. Zeitschrift für Theorie, Praxis, Dokumentation 1 (15), 151–177.

Sandig, Barbara (1997): Formulieren und Textmuster. Am Beispiel von Wissenschaftstexten. In: Jakobs, Eva-Maria und Dagmar Knorr (Hrsg.): Schreiben in den Wissenschaften. Frankfurt a. M.: Lang, 25–44.

Sandig, Barbara (2003): Formen des Bewertens. In: Bobrowski, Ireneusz (Hrsg.): Anabasis. Festschrift für Krystyna Pisarkowa. Kraków: Lexis, 279–287.

Trevisan, Bianka et al. (2013a): Sprachgebrauch in den neuen Medien: Textsortenspezifische Phänomene. Vortrag auf der GAL 2013 vom 19. bis 20. September an der RWTH Aachen University (D).

Trevisan, Bianka et al. (2013b): Web Comment-based Trend Analysis on Deep Geothermal Energy. In: Proceedings of the IEEE International Professional Communication Conference (IPCC) 2013, 15.–17.07.2013, Vancouver (CA).

Trevisan, Bianka (2014): Bewerten in Blogkommentaren. Dissertation. RWTH Aachen University.

Trevisan, Bianka et al. (2014): Facebook as a source for human-centred engineering. Web Mining-based reconstruction of stakeholder perspectives on energy systems. In: Proceed-

ings of the 1st International IBM Symposium on Human Factors, Software, and Systems Engineering (AHFE 2014), 180–191.

Trevisan, Bianka und Eva-Maria Jakobs (2010): Talking about Mobile Communication Systems. Verbal Comments in the Web as a Source for Acceptance Research in Large-scale Technologies. In: Proceedings of the IEEE International Professional Communication Conference (IPCC) 2010, 7.–9.07.2010, Twente (NL), 93–100.

Trevisan, Bianka und Eva-Maria Jakobs (2012): Probleme und Herausforderungen bei der Identifikation von Bewertungen in großen Textkorpora. Am Beispiel Mobilfunk. In: Braukmeier, Sabrina et al. (Hrsg.): Wege in den Sprachraum. Frankfurt a.M. u. a.: Lang, 189–209.

Trevisan, Bianka und Eva-Maria Jakobs (2015): Linguistisches Text Mining. In: Keller, Bernhard et al. (Hrsg.): Zukunft der Marktforschung. Entwicklungschancen in Zeiten von Social Media und Big Data. Heidelberg: Springer Gabler, 167–185.

Wirtz-Brückner, Simone (2015): Social Networking Sites als Gebrauchsmuster. Prototypische Merkmale aus Experten- und Nutzersicht. Dissertation. RWTH Aachen University.

Wirtz, Simone und Eva-Maria Jakobs (2013): Improving User Experience for Passenger Information Systems. Prototypes and Reference Objects. In: IEEE Transactions on Professional Communication 56, 120–137.

Zifonun, Gisela et al. (1997): Grammatik der deutschen Sprache. 3 Bände. Berlin/New York: de Gruyter [Schriften des Instituts für deutsche Sprache; 7].