

# Webbasierte linguistische Forschung: Möglichkeiten und Begrenzungen beim Umgang mit Massendaten\*

Britta Juska-Bacher (Basel), Chris Biemann (Darmstadt) und Uwe Quasthoff (Leipzig)

---

## Abstract

Over the past ten to fifteen years, web-based methods of sociological research have emerged alongside classical methods such as interviews, observations and experiments, and linguistic research is increasingly relying upon them as well. This paper provides an overview of three web-based approaches, i.e. online surveys, crowd-sourcing and web-based corpus analyses. Examples from specific projects serve to reflect upon these methods, address their potential and limitations, and make a critical appraisal. Internet-based empirical research produces vast and highly diverse quantities of (speaker-based or textual) data, presenting linguistic research with new opportunities and challenges. New procedures are required to make effective use of these resources.

---

## 1 Einführung

Neben die klassischen sozialwissenschaftlichen Methoden Befragung, Beobachtung und Experiment sind in den letzten 10 bis 15 Jahren webbasierte Formen getreten, die zunehmend auch in der linguistischen Forschung genutzt werden. In diesem Artikel verstehen wir 'webbasierte linguistische Forschung' in dem Sinne, dass das World Wide Web von den Forschenden als Instrument zur Gewinnung empirischer Daten und teilweise auch als Kommunikationskanal genutzt wird. Es geht uns in diesem Beitrag explizit nicht um die Untersuchung spezifischer im Internet verwendeter Sprachformen (z. B. in E-Mail und Chat) oder um das Internet als technisches Mittel (z. B. für Logfile-Analysen).

Das Internet hat verschiedene charakteristische Aspekte, die für die linguistische Forschung von Interesse sind:

1. Die technische Infrastruktur zur Kommunikation und Speicherung ermöglicht völlig neue Formen des Informationsaustauschs. Durch neue Kommunikationsformen wie Foren, Blogs usw. bilden sich sogenannte Communities von Nutzern mit gleichen Interessen.
2. Zahl und Diversität der Internetnutzer sind immens. Auch wenn die allermeisten Nutzer nicht an linguistischen Fragestellungen interessiert sind, verbleiben doch sehr

---

\* Unser Dank gilt den zwei GutachterInnen für ihre äußerst produktiven Kritikpunkte zu einer früheren Version dieses Artikels.

viele Nutzer mit Interesse an ihrer Sprache. Durch die neuen Kommunikationsformen sind sie für die linguistische Forschung zumindest ansprechbar.

3. Menge und Vielfalt der verfügbaren Texte sind unüberschaubar groß. Eine früher unvorstellbar große Textmenge ist frei zugänglich und durch die Suchmaschinen bereits erschlossen.

Damit hat sich die Situation in den letzten 30 Jahren grundlegend gewandelt. Die Menge der auswertbaren Texte als Rohmaterial hat riesige Ausmaße angenommen und linguistisch interessierte Informanten und Helfer (allerdings meist ohne akademische linguistische Ausbildung) stehen zur Verfügung. Mit der Nutzung dieser neuen Ressourcen sieht sich die linguistische Forschung vor neue Herausforderungen gestellt. Verbunden damit sind auch neue Arbeitsabläufe, um diese Ressourcen effektiv zu nutzen.

In den folgenden Abschnitten geben wir einen Überblick über drei webbasierte Methoden: Zwei davon sind am Sprachbenutzer orientiert, nämlich Online-Befragung und Crowdsourcing, eine ist am Sprachprodukt orientiert, i. e. die webbasierte Korpusanalyse. Anhand von konkreten Projektbeispielen aus Areallinguistik, Sprachtechnologie und Lexikografie werden diese Methoden reflektiert, besondere Möglichkeiten und Begrenzungen angesprochen und es wird eine kritische Würdigung dieser Methoden vorgenommen. Während Online-Befragung (Abschnitt 2) und webbasierte Korpusanalyse (Abschnitt 4) auch in webunabhängiger Form möglich und in dieser Ausprägung seit Langem bekannt sind, handelt es sich beim Crowdsourcing (Abschnitt 3) um eine webgebundene, neue Methode. Da sie von den dreien den geringsten Bekanntheitsgrad hat, geben wir dieser Methode etwas größeren Raum und weisen im entsprechenden Abschnitt auch auf eine Reihe von Beispieluntersuchungen hin.

## **2 Online-Befragung in der Areallinguistik**

### **2.1 Definition der Online-Befragung**

Die Befragung ist sowohl in den Sozialwissenschaften als auch in der Linguistik die weitaus häufigste und bekannteste Methode zur Gewinnung empirischer Daten. Unter 'Online-Befragung' verstehen wir die indirekte Befragung von Informanten anhand eines Fragebogens, die unter Nutzung des Internets durchgeführt wird. Die Nutzung dieses – im Vergleich zur postalischen und telefonischen Befragung – neuen Kommunikationskanals und der damit in Verbindung stehenden Probandenrekrutierung hat systematische Folgen im Sinne von positiven wie negativen Methodeneffekten (siehe Abschnitt 2.4). Diese webbasierte Erhebungsmethode wird von der deutschen Sozialforschung seit Mitte der 1990er-Jahre genutzt (Zerback et al. 2009: 15).

Unter der Bezeichnung "Online-Befragung" werden drei verschiedene Vorgehensweisen subsumiert, die sich durch abnehmende Kontrollierbarkeit des Forschungsprozesses auszeichnen:

1. Die Teilnehmenden erhalten den Fragebogen per E-Mail und senden ihn ausgefüllt per E-Mail oder als Ausdruck zurück. Bei diesem Vorgehen wird nur ein neuer Kommunikationskanal genutzt. Da die E-Mails gezielt versandt werden, ist dieses

Vorgehen wie bei der telefonischen und postalischen Befragung als aktive oder gesteuerte Probandenakquirierung zu bezeichnen.

2. Die Teilnehmenden laden den Fragebogen von einem Server herunter und senden ihn per E-Mail oder als Ausdruck zurück. Auch hier ist der Kommunikationskanal neu; je nachdem, ob die Teilnehmenden im Vorfeld angesprochen wurden oder nicht, handelt es sich um eine aktive oder passive Akquirierung.
3. Ein elektronischer Fragebogen wird auf einem Server abgelegt. Über eine oder mehrere Websites wird zum Ausfüllen aufgerufen. Die Teilnehmenden füllen den Fragebogen im Internet aus, die Daten werden elektronisch zum Server übermittelt, dort abgespeichert und können jederzeit vom Untersuchenden abgerufen werden. In Verbindung mit der neuen Kommunikationsinfrastruktur erfolgt eine passive oder ungesteuerte Probandenakquirierung.

Der Unterschied zwischen diesen Vorgehensweisen in der Probandenakquirierung steht im Zusammenhang mit der Steuerbarkeit und Kontrollierbarkeit des Forschungsprozesses durch die Forschenden (detaillierter siehe Abschnitt 2.4). Wir definieren als Online-Befragung in diesem Artikel die dritte Vorgehensweise, d. h. die ungesteuerte Probandengewinnung.

## 2.2 Online-Befragung in der Linguistik

Auch in der Linguistik hat die Online-Befragung inzwischen an Beliebtheit gewonnen.<sup>1</sup> Prinzipiell eignen sich für diese Erhebungsform dieselben Fragestellungen, die bisher anhand von klassischen Befragungen durchgeführt wurden, insbesondere solche, die große Probandenzahlen erfordern. Ein prädestinierter Bereich sind areallinguistische Fragestellungen, die in erster Linie eine deskriptive, z. B. kartografische Darstellung der Daten anstreben. In diesem Bereich wurden Online-Befragungen auf verschiedenen Ebenen ausgeführt: Das im folgenden Abschnitt vorgestellte Projekt untersucht die Ebene des Dialekts; es liegen aber ebenso Online-Befragungen zur Umgangs- oder Alltagssprache (Elspaß/Möller 2006), zu standardsprachlichen Varietäten (Juska-Bacher 2011) und zum interlingualen Vergleich (Juska-Bacher 2009) vor.

## 2.3 Beispielprojekt: Gewinnung wortgeografischer Dialektdaten

Der Einsatz von Fragebögen hat in der Dialektologie eine lange Tradition, siehe z. B. *Deutscher Sprachatlas* (Wenker et al. 1926–1956) und *Deutscher Wortatlas* (Mitzka/Schmitt 1951–1980). Da die Informanten in der Regel nicht linguistisch geschult sind, eignet sich die Fragebogenerhebung besonders für einfach zu notierende wortgeografische Phänomene (Eichhoff 1982: 550; Niebaum/Macha 2006: 15).

Ziel des im Folgenden exemplarisch für eine Online-Befragung aufgeführten Projekts war die Untersuchung sprachgeografischen Wandels in der Schweiz (detailliertere Beschreibung siehe Juska-Bacher 2010a, 2010b). Mit dem *Sprachatlas der deutschen Schweiz* (SDS) (Hotzenköcherle et al. 1962–1997), für den in den Jahren 1939–1958 an 573 Ortspunkten umfangreiche Interviewdaten erhoben wurden, liegt für den Deutschschweizer Sprachraum

---

<sup>1</sup> Die Einführung in Online-Fragebögen und Online-Interviews der University of Leicester ist einsehbar unter: <http://www.restore.ac.uk/orm/site/home.htm>.

umfassendes Kartenmaterial zur Laut-, Wort- und Formengeografie der jüngeren Sprachgeschichte vor. Zum Nachweis diachronen Wandels wurden bereits verschiedentlich Vergleichsdaten erhoben (z. B. Wolfensberger 1967; Christen 1988; Schifferle 1995; Siebenhaar 2000). Da häufig lautliche Veränderungen im Fokus standen und die Erfassung phonetischer Feinheiten linguistisch geschulter Interviewer bedarf, blieb man methodisch bei einer direkten Erhebung und beschränkte sich, aufgrund begrenzter Ressourcen, auf einen Ausschnitt des Sprachraums. Die hier beschriebene Studie hingegen hatte das Ziel, den gesamten Deutschschweizer Raum zu erfassen. Um möglichst viele Dialektsprecher und damit eine möglichst breite regionale Streuung bei überschaubarem zeitlichem Aufwand zu erreichen, wurde mit einer Online-Befragung gearbeitet. Daraus resultierend bot sich eine Beschränkung auf den wortgeografischen Bereich an (s. o.).

Im halb offen gestalteten Fragebogen sollten die Probanden zu 25 Wortschatzfragen die in ihrem Dialekt gebräuchlichste Bezeichnung angeben. Im Einleitungstext wurde um die Angabe einiger soziodemografischer Daten gebeten (u. a. Alter, Geschlecht, Ortsansässigkeit, beschriebene Mundart). Der Fragebogen wurde auf einem Server der Universität Zürich abgelegt und über verschiedene Printmedien, eine Radiosendung sowie die Online-Ausgabe einer kostenlosen Pendlerzeitung wurde zum Ausfüllen aufgerufen. Die Befragung lief von November 2007 bis April 2008. Es beteiligten sich rund 5600 Probanden (gut 0,1% der Deutschschweizer Bevölkerung) von 1258 verschiedenen Ortspunkten der Deutschschweiz.

Eine exemplarische Gegenüberstellung der Karten aus den Online- und den SDS-Daten erlaubte erstmals eine empirisch gestützte Aussage zum Sprachwandel seit den SDS-Erhebungen für den gesamten Deutschschweizer Dialektraum (Juska-Bacher 2010a). In diesen Karten waren die Grobgliederungen des SDS noch deutlich erkennbar (was für die prinzipielle Validität dieser Erhebung spricht), es konnten jedoch auch sehr deutliche wortgeografische Veränderungstendenzen festgestellt werden, die bisher nur für Auszüge des Sprachraums belegt waren. Beispielhaft konnte ein starker standardsprachlicher Einfluss aufgezeigt werden, ohne dass dabei die bisherigen dialektalen Äquivalente gänzlich verschwunden wären.

Die ebenfalls erfassten soziodemografischen Daten wurden zudem mithilfe von statistischen Tests ausgewertet und gaben Hinweise auf die Bedeutung besonders des Alters für die sprachliche "Innovationsfreudigkeit" der Probanden, die auch in anderen Studien belegt wurde. Sie erlauben außerdem eine genauere Beschreibung der Probanden, die für die Methodendiskussion im folgenden Abschnitt von Bedeutung ist. Diese Diskussion beschäftigt sich mit Vorteilen und Kritikpunkten der Online-Befragung.

## **2.4 Vorteile und Kritikpunkte der Online-Befragung**

In Abschnitt 2.1 wurde angedeutet, dass die Nutzung des World Wide Web für die Befragung und besonders für die damit verbundene andere Probandenrekrutierung systematische Folgen hat. Die Vorteile der Online-Befragung lassen sich unter folgenden Punkten in Tabelle 1 zusammenfassen (siehe z. B. Taddicken 2008 und Tourangeau 2004).

Vorteile der Online-Befragung
1. Kosten- und Zeitersparnis
2. Fehlerreduzierung: a. Erläuterungen zum Ausfüllen des Fragebogens b. Reduzierung von Effekten sozialer Erwünschtheit c. Entfallen einer manuellen Dateneingabe
3. attraktive und benutzerfreundliche Gestaltungsmöglichkeit des Fragebogens

**Tabelle 1: Vorteile der Online-Befragung (im Vergleich zur direkten Befragung durch Interviewer und zum postalischen Versand von Fragebögen)**

Die Arbeit mit einem elektronischen, auf einem Server abgelegten Fragebogen, der online ausgefüllt, auf dem Server gespeichert wird und jederzeit abrufbar ist, ist in mehrfacher Hinsicht sehr kostengünstig. Einerseits kann bei der Erhebung sowohl auf Interviewer als auch auf den Kauf von Adresslisten sowie Papier und Porto verzichtet werden. Die Verwendung dieses beschleunigten Kommunikationskanals und das automatische Erfassen der Daten wirken wiederum zeit- und damit kostensparend. Da der Zeitaufwand nicht in direktem Zusammenhang mit der Zahl der akquirierten Probanden steht, bietet dieses Verfahren die Möglichkeit, sehr große Probandenzahlen (mit breiter, z. B. arealer Streuung) zu realisieren. Diese Vorteile fasst Tourangeau (2004: 782) als "marriage of low cost and high capabilities" zusammen.

Die Fehleranfälligkeit kann auf der Probandenseite durch Erklärungen zum Ausfüllen des Fragebogens reduziert werden. Der Verzicht auf die Vermittlungsinstanz Interviewer und die fehlende Personalisierung des Fragebogens schaffen Anonymität und reduzieren damit die Reaktivität und Effekte sozialer Erwünschtheit, sodass davon auszugehen ist, dass der Proband unbeeinflusst, unabhängiger und ehrlicher antwortet. Aufseiten der Forschenden kann eine Fehlerreduzierung durch das Wegfallen der manuellen Datenerfassung erreicht werden. Der Fragebogen selbst erlaubt eine attraktivere und benutzerfreundlichere Gestaltung als ein Ausdruck. Das heißt, er kann durch Abbildungen, Diagramme, Videoclips und auditives Material ergänzt werden. Eine einfache Filterführung erspart dem Probanden die Lektüre für ihn irrelevanten Materials.

Insbesondere wegen seiner Ökonomie stellt der Online-Fragebogen aufgrund der forschungspragmatischen Bedingungen häufig die einzige Möglichkeit dar, Daten von einer großen Menge von Probanden zu erheben.

Kritikpunkte der Online-Befragung
1. mangelnde Repräsentativität der Probanden
2. Fehleranfälligkeit: a. mangelnde Kontrollierbarkeit b. soziale Entkontextualisierung

**Tabelle 2: Kritikpunkte der Online-Befragung (z. B. Taddicken 2008 und Tourangeau 2004)**

Der zentrale Kritikpunkt der Online-Befragung berührt in erster Linie die inferenzstatistische Auswertung der Daten (vgl. Tabelle 2). Damit von einer Stichprobe auf die Grundgesamtheit geschlossen werden kann, muss erstens die Grundgesamtheit definiert sein und zweitens eine repräsentative, d. h. entweder zufällig ausgewählte oder systematisch gezogene Stichprobe

vorliegen. Letzteres ist bei einer Online-Befragung nicht der Fall. Die Grundgesamtheit, über die eine Aussage gemacht werden soll, besteht häufig aus der gesamten Bevölkerung. Die Stichprobe sollte, um repräsentativ zu sein, vom Forschenden aktiv gezogen werden, wobei jede Person der Grundgesamtheit die gleiche Chance hat, in die Stichprobe zu gelangen. Gemäß unserer Definition von Online-Befragung erfolgt die Stichprobenziehung jedoch nicht vom Forschenden kontrolliert, sondern passiv, d. h. die Probanden entscheiden selbst über die Teilnahme (Selbstselektion, man spricht auch von einem 'Convenience Sampling'). Insofern kann nur von einer beschränkten Verallgemeinerbarkeit der Ergebnisse auf die Grundgesamtheit gesprochen werden.

Zentral für die Linguistik ist in diesem Zusammenhang die Frage, ob es im sprachlichen Bereich systematische Unterschiede zwischen der Stichprobe und der Grundgesamtheit (im Idealfall die Gesamtbevölkerung) gibt. Solche systematischen Unterschiede zwischen den beiden Gruppen stellte Bandilla (1999: 15) für Online-Befragungen in den Sozialwissenschaften heraus, indem er ein deutliches Ungleichgewicht zugunsten von jüngeren Menschen, Männern und Personen mit einer höheren Bildung feststellte. Diese Tendenz wurde zehn Jahre später von Maurer und Jandura (2009: 65) bestätigt. Vier sprachwissenschaftliche Online-Befragungen (Elspaß/Möller 2006; Juska-Bacher 2009; Juska-Bacher 2010b; Klosa et al. 2011) stützen diese Verteilungen hinsichtlich des Alters (der Anteil der unter 30-Jährigen lag bei etwa 50%) und der Ausbildung (80–90% Akademiker). Hinsichtlich des Geschlechts allerdings zeichnen sie ein deutlich anderes Bild: Der Anteil der Teilnehmerinnen lag mit einem Prozentsatz von 50–70% deutlich höher als bei den sozialwissenschaftlichen Studien. Diese Abweichung ist entweder auf eine größere Teilnahmebereitschaft von Frauen an Umfragen zu sprachlichen Themen oder aber auf eine höhere Zugriffswahrscheinlichkeit von Frauen auf die entsprechenden Seiten, auf denen zum Ausfüllen des Fragebogens aufgerufen wurde, zurückzuführen.

Es hängt von der Fragestellung der Untersuchung ab, ob und wie groß der Einfluss dieser Faktoren auf das Sprachwissen der Probanden ist. Bei Fragen zu arealen Verteilungen in der Alltagssprache (Elspaß/Möller 2006) oder zum Schweizer Dialektwortschatz (Juska-Bacher 2010b) scheint dies in begrenzterem Maße der Fall zu sein als beispielsweise bei jugendsprachlichen oder Genderthemen. Soziodemografische Daten, die auffällige Verteilungen dokumentieren können, sollten auf jeden Fall in Online-Fragebögen erfasst werden, um systematische Unterschiede gegebenenfalls ausmachen und quantifizieren zu können. Dies bietet bei einer statistischen Auswertung die Möglichkeit, diese Faktoren zu korrigieren.

Die Zusammensetzung der Stichprobe wird maßgeblich beeinflusst durch die Kanäle, über die auf die Befragung aufmerksam gemacht bzw. zur Teilnahme aufgerufen wird. Eine Strategie, dieses Problem zu reduzieren, ist, Aufrufe zum Ausfüllen des Fragebogens in möglichst allgemein zugänglichen Kanälen wie Tages- und Wochenzeitungen zu platzieren, wie dies im Falle der in Abschnitt 2.3 beschriebenen Studie geschehen ist (Juska-Bacher 2010b: 281). Auf diese Weise kann eine möglichst breite Probandengewinnung erfolgen. Andere Strategien der Qualitätssicherung zielen auf die Zuverlässigkeit der Probandenangaben. Sie bestehen u.a. darin, sich nicht auf einen einzelnen Probanden zu verlassen, sondern Antworten von mehreren Probanden (im angeführten Beispiel von denselben Ortspunkten) zu aggregieren

und die häufigste Angabe anzusetzen oder Testfragen in den Fragebogen einzubauen, die Auskunft über die Zuverlässigkeit der Probandenangaben geben.

Da die Teilnehmer von Online-Befragungen frei und eigenaktiv über ihre Teilnahme entscheiden, ist prinzipiell zu erwarten, dass einerseits die Anforderungen an die Gestaltung des Fragebogens hoch sind, um möglichst viele Probanden zum (vollständigen) Ausfüllen zu motivieren (Attraktivität, begrenzte Zahl von Fragen, Filterführung usw.), und andererseits die Probanden sich in der Regel (wie bei anderen Methoden auch) durch ein überdurchschnittliches Interesse für das Thema des Fragebogens auszeichnen.

Der zweite Kritikpunkt an der Online-Befragung betrifft ihre Fehleranfälligkeit (vgl. Tabelle 2). Kritisiert wird die mangelnde Kontrollierbarkeit, sodass – z. B. bei weniger geübten Probanden durch Probleme beim Scrollen – unbeabsichtigt Fehler auftreten können. Ebenso wenig kontrollierbar ist, ob – durch vorsätzlich falsches Ausfüllen oder Mehrfachausfüllen – willentlich Fehler produziert werden. Dieses Problem der Fehleranfälligkeit besteht selbstverständlich in ähnlichem Maße bei Papierfragebögen. Im Falle der Online-Befragung sollte die große Zahl der Probanden ein wirksames Korrektiv darstellen. Ein Teil mutwilliger Falschangaben fällt zudem bei der Datenbereinigung bereits ins Auge, sodass die entsprechenden Datensätze ausgesondert werden können. Unter den Vorteilen der Online-Befragung wurde eine Anonymisierung der Befragungssituation genannt, die zu einer Reduzierung der Effekte sozialer Erwünschtheit führen kann. Kritiker dieser Methode hingegen sehen in der sozialen Entkontextualisierung (zum Begriff siehe Taddicken 2008: 153) die Gefahr, dass Probanden nicht ehrlicher, sondern – weil unkontrolliert – falsch antworten. Auch hier ist auf die großen Probandenzahlen als Korrektiv zu hoffen.

## **2.5 Kritische Würdigung der Online-Befragung**

Die Online-Befragung stellt eine zeit- und kostensparende Möglichkeit der Datenerhebung dar, die mit verschiedenen positiven wie negativen Methodeneffekten verbunden ist. Fehlerreduzierenden Aspekten wie der Möglichkeit multimedialer Hilfestellung beim Ausfüllen des Fragebogens, der Reduzierung von Effekten sozialer Erwünschtheit und dem Entfallen einer manuellen Dateneingabe steht der Kritikpunkt der Fehleranfälligkeit wegen mangelnder Kontrollierbarkeit und sozialer Entkontextualisierung gegenüber. Es ist davon auszugehen, dass große Teilnehmerzahlen im Hinblick auf die Fehleranfälligkeit in vielen Fällen korrigierend wirken.

Der zweite zentrale Kritikpunkt der fehlenden Repräsentativität der Probanden betrifft weniger deskriptive Darstellungen der Ergebnisse (z. B. in Form von Karten) als eine statistische Auswertung. Eine Strategie, diesen Punkt in den Griff zu bekommen, sind die Erfassung und Berücksichtigung soziodemografischer Daten, die systematische Abweichungen der Stichprobe von der Grundgesamtheit dokumentieren können, um die Ergebnisse für diese Faktoren korrigieren zu können. Bei der Interpretation der Testdaten ist dennoch immer eine gewisse Vorsicht geboten.

Werden für die Beantwortung einer Fragestellung große Probandenzahlen benötigt, wie zum Beispiel im Bereich der Areallinguistik, ist eine Online-Befragung aufgrund ihrer geringen Kosten und ihres Potenzials, große Probandenzahlen zu akquirieren (bei in der Regel

begrenzten Mitteln), häufig der einzig gangbare Weg. Bei einem Abwägen, für welche linguistischen Fragestellungen diese Erhebungsart geeignet ist, und bei einem reflektierten Umgang mit ihren methodenspezifischen Eigenheiten bietet die Online-Befragung hier "Chancen für die Dialektologie [und nicht nur für sie! – Anm. der Verfasser], die noch vor kurzer Zeit unvorstellbar waren" (Elspaß/Möller 2006: 123).

### **3 Crowdsourcing**

#### **3.1 Was ist Crowdsourcing?**

Der Begriff des Crowdsourcings wurde von Howe (2006) geprägt und vereint in einem Wort die Konzepte des Outsourcings und der Crowd Intelligence (Schwarmintelligenz). Ursprünglich wurde Crowdsourcing vornehmlich für die Ideenfindung eingesetzt, z. B. beim Entwurf und der Bewertung von Logos, d. h. von Aufgaben, die normalerweise an externe Kreativagenturen vergeben werden. Anstelle des Outsourcings an einzelne Dienstleister wird die Aufgabe beim Crowdsourcing von einer nicht notwendigerweise untereinander vernetzten Gruppe von Crowdarbeitern gelöst. Crowdsourcing in seiner Ursprungsform kann hier als Online-Befragung verstanden werden, da die Besucher bestimmter Websites z. B. über Entwürfe von Logos abstimmen können. Anders als bei der Online-Befragung jedoch können die zur Abstimmung vorgelegten Objekte (z. B. Logo-Entwürfe) von den Teilnehmern selbst eingebracht werden; die Befragten werden so zu Befragern, wenngleich in einem eng abgesteckten Rahmen.

Die Kernidee von Crowdsourcing und anderen Web-2.0.-Instrumenten ist dieselbe wie bei der Online-Enzyklopädie Wikipedia: Das Kollektiv in seiner Gesamtheit leistet genauso viel oder sogar mehr als einzelne Experten (vgl. Giles 2005), Fehlentscheidungen werden durch das Prinzip "Viele Augen sehen mehr als zwei" vermieden und eine komplexe Gesamtaufgabe kann durch die Beiträge vieler Einzelner gelöst werden. Illustriert wird dies in einem Projekt zum kollaborativen Entziffern von handschriftlichen Notizen (Little et al. 2009): Hier wird gemeinschaftlich so lange an der elektronischen Verschriftlichung einer handschriftlichen Notiz gefeilt, bis Konsens erreicht wird und keine Änderungen mehr vorgenommen werden. Keiner der Einzelbeitragenden muss die gesamte Notiz bearbeiten, Teilbeiträge helfen und die für gut befundenen Teilentzifferungen bleiben bestehen.

Wurde die Teilnahme an Crowdsourcing-Aufgaben in der Anfangszeit allein durch deren Neuheitsfaktor oder durch kleine Anreize wie Gewinnspiele motiviert, trat an die Stelle von quasi freiwilligen Crowdarbeitern schnell ein Marktmodell, welches Anbieter von Crowdaufgaben und Anbieter von Crowdarbeitskraft innerhalb einer Crowdsourcing-Plattform vermittelt. Bereits im Jahr 2005 ging mit Amazon Mechanical Turk ([www.mturk.com](http://www.mturk.com)) die erste Crowdsourcing-Plattform online. Dort können Auftraggeber (sogenannte Requester) Aufgaben formulieren, die von den Crowdarbeitern (sogenannten Workern) bearbeitet werden. Die Plattform stellt dafür Designelemente für Aufgaben, Tools für die Aggregation von Antworten sowie für die Abwicklung der Finanzströme von Requester zu Worker bereit. Auftraggeber können den Pool ihrer Crowdarbeiter beschränken, etwa durch das Herkunftsland, durch einen Mindestanteil von für gut befundenen Aufgaben oder durch Qualifikationstests.



Neben den eben erwähnten offenen Plattformen entstand eine Reihe von Dienstleistern, welche Crowdarbeit nutzen, um Datenerhebungsaufträge zu realisieren. Hier bekommt der Kunde lediglich die Ergebnisse; die Aufteilung der Gesamterhebung in einzelne Crowdsourcing-Aufgaben sowie das Qualitätsmanagement sind Teil der Dienstleistung. An dieser Stelle seien exemplarisch zwei Anbieter genannt: das deutsche Angebot Clickworker ([www.clickworker.com](http://www.clickworker.com)), welches sich durch sprachliche Vielfalt der registrierten Crowdarbeiter auszeichnet, und das Non-Profit-Projekt SamaSource (<http://samasource.org/>), welches Crowdarbeit zu fairen Konditionen in Entwicklungsländer vergibt.

Auch wenn Crowdsourcing die Art der zu bearbeitenden Aufgaben nicht a priori einschränkt, so gibt es doch Aufgabenarten, welche für Crowdsourcing geeigneter sind als andere. Dies hat mit Möglichkeiten der automatisierten Qualitätskontrolle zu tun (siehe Abschnitt 3.3). Typische Crowdsourcing-Aufgaben können in sehr kurzer Zeit erledigt werden und haben eine geringe Zahl an Antwortmöglichkeiten auf klar definierte Fragen, welche ohne spezielles Training beantwortet werden können. Typischerweise wird dieselbe Aufgabe an mehrere Crowdarbeiter vergeben und die Antwort aggregiert, z. B. durch Mehrheitsentscheidung oder durch Mittelwertbildung. Häufige Erhebungen finden zum Beispiel im Bereich Relevanzbewertung für Suchanfragen von Suchmaschinen (z. B. Stone et al. 2011) und für das Marketinginstrument der Sentimentanalyse statt.

Wer sind diese Crowdarbeiter? In frühen Surveys wurde ermittelt, dass Crowdarbeiter auf Amazon Mechanical Turk vornehmlich aus den USA stammen und eine Demografie ähnlich der Durchschnittspopulation aufweisen, jedoch mit überdurchschnittlich vielen Frauen und überdurchschnittlicher Repräsentation der Altersgruppe 25–34. Inzwischen hat sich das Bild gewandelt (Ross et al. 2010; Ipeirotis 2010); die Crowd wird zunehmend internationaler und besteht zunehmend aus männlichen Crowdarbeitern. Insbesondere gibt es ein großes Angebot an Arbeitskraft aus Niedriglohnländern wie Indien, wo die vergleichsweise geringen Entgelte einen größeren relativen Gegenwert darstellen.

Im Folgenden wenden wir uns der Frage zu, ob und wie das Instrument des Crowdsourcings auch für die linguistische Forschung genutzt werden kann. Ferner beleuchten wir Maßnahmen zur Qualitätssicherung und diskutieren Vorteile und Nachteile dieser Datenerhebungsmethode.

### **3.2 Crowdsourcing für linguistische Datenerhebungen**

Zunächst geben wir einen breiten Überblick darüber, für welche Art linguistischer Annotationen Crowdsourcing bereits eingesetzt wurde. Diese unterteilen sich in klassische linguistische Fragestellungen, sprachtechnologische Fragestellungen sowie die Verbindung von Sprache mit anderen Modalitäten. Die Studie von Snow et al. (2008) zeigt, dass ab einer gewissen Redundanzstufe bei den erhobenen Daten keine Qualitätsunterschiede zwischen Crowdarbeitern und Expertenannotatoren festgestellt werden. Diese Ergebnisse können jedoch nur für relativ einfache Annotationsaufgaben bestätigt werden. Ferner stellen wir exemplarisch ein Crowdsourcing-Projekt zum Erstellen eines Wortbedeutungsinventars (Biemann 2012) vor, welches dieses komplexe Problem in einfache Teilschritte zerlegt, um es für Crowdsourcing bearbeitbar zu machen.

### 3.2.1 Crowdsourcing für linguistische Fragestellungen

Innerhalb der Linguistik wird Crowdsourcing zur Feststellung von Grammatikalität eingesetzt, z. B. von Madnani et al. (2011). Dies ist eine geeignete Crowdsourcing-Aufgabe: Es gibt nur zwei mögliche Antworten, die Frage lässt sich einfach formulieren, die Beantwortung geht schnell und es gibt hohe Korrelationen der Antworten verschiedener Crowdarbeiter. Ganz ähnliche Charakteristika besitzt die Entscheidung über die Anbindung von Präpositionalphrasen, wie sie etwa in Jha et al. (2011) ausgeführt wird. Im eher semantischen Bereich wurde Crowdsourcing von Parent und Eskenazi (2010) eingesetzt, um Definitionen aus verschiedenen Wörterbüchern zu alignieren, was insbesondere für mehrdeutige Wörter interessant ist. Der Versuch, das semantisch stark strukturierte Online-Wörterbuch *FrameNet* mit Crowdsourcing zu erweitern, wird in Hong/Baker (2011) beschrieben. Hier konnten aufgrund der relativ komplexen Aufgabe nur Teilerfolge erzielt werden: Die Anzahl von Crowdarbeitern, welche die Aufgabe überhaupt durchführten, war gering; ferner mussten größere Anstrengungen unternommen werden, um den linguistischen Formalismus in eine für Nichtlinguisten verdauliche Form zu bringen.

Auch im Zusammenhang mit Fragestellungen mit gesprochenem Material wird Crowdsourcing erfolgreich eingesetzt, in etwa in Kunath/Weinberger (2010): Hier wurden Muttersprachler von Nichtmuttersprachlern erfolgreich unterschieden; ein ähnlicher Versuchsaufbau ließe sich z. B. für die Zuordnung von Dialekten zu Sprachproben verwenden.

### 3.2.2 Crowdsourcing für sprachtechnologische Fragestellungen

In den wenigen Jahren seit seiner Entstehung hat sich Crowdsourcing für die Erstellung und Verbesserung von sprachtechnologischen Verarbeitungskomponenten bereits fest etabliert. Hier werden typischerweise Annotationsaufgaben, für die man früher speziell trainierte Annotatoren einsetzte, in den Crowdspace verlagert. Dies spart vor allem Kosten pro Annotationseinheit, was größere Umfänge ermöglicht. Annotierte Texte können mithilfe maschinellen Lernens dazu verwendet werden, automatische Sprachverarbeitungskomponenten zu erstellen, wobei (bei gleicher Datenqualität) größere Mengen an Annotationen im Allgemeinen bessere Systeme ermöglichen. Wegen der großen Zahl an Crowdsourcing-Projekten in der Sprachtechnologie werden an dieser Stelle nur einige exemplarische Anwendungsfelder genannt.

In der Datenerhebung von Übersetzungen (Bloodgood/Callison-Burch 2010; Zaidan/Callison-Burch 2010) konnte kein Unterschied zwischen maschinellen Übersetzungssystemen, welche auf professionell übersetzten Satzpaaren beruhen, und solchen, die durch mittels Crowdsourcing erstellte Satzpaare trainiert wurden, festgestellt werden. Von Chen und Dolan (2011) wurde Crowdsourcing erfolgreich zum Sammeln von Paraphrasen eingesetzt.

Einen hybriden Ansatz beschreiben Voyer et al. (2010) für die Eigennamenerkennung. Die Annotationsaufgabe wurde in zwei Schritte geteilt: Die Markierung der korrekten Namensspannen erfolgte durch trainierte Annotatoren, da insbesondere für Organisations- und Produktnamen komplexe Richtlinien kommuniziert werden mussten. Das Klassifizieren in verschiedene Typen von Namen jedoch konnte erfolgreich mit Crowdsourcing durchgeführt werden. Auch für das Sammeln von gesprochener Sprache fanden Lane et al. (2010) es

notwendig, Crowdsourcing-Arbeiter stärker zu trainieren als auf den Plattformen üblich. Bei der Transkription von Audiodaten (Novotney/Callison-Burch 2010) gab es jedoch keine derartigen Probleme.

### **3.2.3 Crowdsourcing zur Verbindung von Sprache mit anderen Modalitäten**

Durch die vielfältigen Möglichkeiten, schnell eine große Menge an Crowdarbeitern zu erreichen, und dank der Unterstützung multimodaler Inhalte seitens Crowdsourcing-Plattformen kann Crowdsourcing auch dazu eingesetzt werden, eine Verbindung zwischen Sprache und anderen Modalitäten herzustellen. Dies reicht von der Ermittlung der Assoziationen von Wörtern und Farben (Mohammad/Turney 2011) über die Erhebung von Emotionalität bei Wörtern (Mohammad/Turney 2011) bis hin zur Generierung von 3D-Welten aus textuellen Beschreibungen (Rouhizadeh et al. 2011): Hier werden den Crowdarbeitern 3D-Szenen zur textuellen Beschreibung präsentiert, um mögliche verbale Deskriptoren zu sammeln.

### **3.2.4 Beispielprojekt: Erstellen eines Wortbedeutungsinventars**

Nun diskutieren wir am Beispiel eines Projektes zum Erstellen eines Wortbedeutungsinventars (Biemann 2012) die Vorgehensweise, um komplexe Datensammelaufgaben in geeignete Einzelschritte zu zerlegen. Das Anwendungsziel dieses Projektes war die Realisierung semantischer Suche mit lexikalischer Substitution: Texte sollen für bestimmte Zielwörter mit Synonymen angereichert werden. Dies geschieht in Abhängigkeit vom Satzkontext, um für mehrdeutige Lexeme keine falschen Erweiterungen zu erhalten, was zu irrelevanten Suchergebnissen führen würde. Hierzu wird ein Wortbedeutungsinventar benötigt, welches Bedeutungen nach der Granularität der Substitution unterscheidet.

Die Erstellung eines Bedeutungsinventars ist normalerweise Lexikografen vorbehalten und eine stark subjektive Aufgabe (siehe z. B. Kilgarriff 1999). Von Crowdarbeitern kann nicht erwartet werden, dass sie diese komplexe Aufgabe, welche Expertenwissen über semantische Unterscheidungen voraussetzt, in einer einzigen Aufgabenstellung befriedigend lösen. Ferner sind mögliche Bedeutungsunterscheidungen eines Wortes, z. B. *Magazin* im Sinne von 'Vorratsort' und 'Publikation' (welche man z. B. in das Publikationsorgan und das konkrete Objekt unterscheiden kann), viel zu vielfältig, als dass verschiedene Antwortmöglichkeiten einfach kombiniert werden könnten. Deshalb wurde die Aufgabe der Erstellung eines Bedeutungsinventars für ein gegebenes Zielwort in drei Teilaufgaben untergliedert, welche durch Crowdsourcing-Unterstützung gelöst und gegebenenfalls mehrfach durchgeführt werden.

In Schritt 1 werden eine Handvoll Sätze aus einem Textkorpus extrahiert, welche das Zielwort enthalten. In einer ersten Crowdsourcing-Aufgabe werden Crowdarbeiter nach mehreren möglichen Substitutionen für das Zielwort gefragt. Obwohl hier die potenzielle Antwortmenge sehr groß ist, gibt es doch starke Überlappungen zwischen mehreren Crowdarbeitern, die denselben Satz bearbeiten. Ebenso ähneln sich die Antworten im Falle von Sätzen, welche das Zielwort in derselben Bedeutung enthalten. Nach Schritt 1 erhalten wir also pro Zielwortsatz eine Charakterisierung durch eine Liste von Substitutionen,

gewichtet durch Mehrfachnennung. Für *Magazin* bekämen wir z. B. für einen Satz 'Depot' (3), 'Vorratskammer' (2), 'Getreidespeicher' (1) und 'Lager' (1).

In Schritt 2 werden Sätze mit ähnlicher Charakterisierung automatisch gruppiert, um sie nach ihrer Bedeutung zusammenzufassen. Dabei werden nur Sätze mit hoher Ähnlichkeit in ihren Charakterisierungen zusammengefasst. Hierdurch entstehen Gruppen von Sätzen, welche (durch die hohe Ähnlichkeitsschwelle) höchstwahrscheinlich je einer Bedeutung entsprechen; gegebenenfalls erhält man jedoch mehrere Gruppen pro Bedeutung. Um dies zu bereinigen, werden in einem zweiten Schritt Sätze aus verschiedenen Gruppen präsentiert, wobei das Zielwort markiert ist. Crowdarbeiter entscheiden nun per Multiple-Choice, ob das Zielwort in derselben Bedeutung vorliegt oder nicht. Das Ergebnis dieser Aufgabe erlaubt es, Gruppen zu vereinigen und somit eine Gruppe pro bisher aufgetretene Bedeutung aufzustellen. Für jede Gruppe wird ein Satz als Repräsentant der Bedeutung ausgewählt.

Zu diesem Zeitpunkt haben wir bereits ein Bedeutungsinventar für unser Zielwort: Jede Bedeutung wird charakterisiert durch einen Repräsentanten sowie durch eine gewichtete Liste von Substitutionen, welche sich aus der Vereinigung der Substitutionen der jeweiligen Gruppe ergibt. Es ist jedoch zu diesem Zeitpunkt nicht garantiert, dass alle wichtigen Bedeutungen des Zielwortes durch das Inventar abgedeckt sind.

Dies wird in Schritt 3 sichergestellt: Aus unserem Korpus werden eine größere Anzahl (z. B. 100) Sätze mit dem Zielwort extrahiert und zusammen mit allen Repräsentanten angezeigt. Crowdarbeiter können nun per Multiple-Choice auswählen, welcher Repräsentant dieselbe Zielwortbedeutung beinhaltet oder ob die vorliegende Bedeutung noch nicht durch das Inventar abgedeckt ist, z. B. in *Das Magazin der Kamera enthielt keinen Film*. So erhalten wir eine größere Anzahl von Sätzen mit Bedeutungsmarkierung, mit denen z. B. maschinelle Lernsysteme trainiert werden können, und garantieren gleichzeitig eine hohe Bedeutungsabdeckung.

Solange neue Bedeutungen gefunden werden, werden neue Iterationen angestoßen. Die Sätze mit den neuen Bedeutungen bilden das Material für Schritt 1; in Schritt 2 wird die neu entstandene Gruppierung verifiziert und in Schritt 3 werden wieder weitere Beispiele gesammelt, bis keine neuen Bedeutungen mehr auftreten. Durch das Aufteilen in eine Freitextaufgabe mit großen Überlappungen (Schritt 1) und zwei Multiple-Choice-Aufgaben können Antworten verschiedener Crowdarbeiter leicht aggregiert werden. Die Einzelaufgaben sind durch kurze Instruktionen leicht verständlich und haben eine geringe Bearbeitungsdauer. Die kognitive Last des Erstellens eines Bedeutungsinventars wird so operationalisiert und im Kollektiv gelöst.

Auf die eben beschriebene Weise wurde eine Ressource zur lexikalischen Substitution, welche insbesondere als Wortbedeutungsinventar dienen kann, für über 1000 hochfrequente englische Nomen erstellt. Dafür wurden zunächst fünf Crowdarbeiter pro Aufgabe eingesetzt; aus diesen wurde ein Pool besonders produktiver und qualitativ hochwertiger Arbeiter rekrutiert, sodass die Redundanz auf drei gesenkt werden konnte. Pro Zielwort kostete die Erstellung durchschnittlich etwa 10 US-Dollar und resultierte in durchschnittlich etwa 170 bedeutungsmarkierten Vorkommen pro Zielwort. Die lexikalische Ressource und ein darauf trainiertes automatisches System für lexikalische Substitution stehen zum Download frei zur

Verfügung (<http://www.langtech.tu-darmstadt.de/software/twsi-sense-substituter/>), eine detailreichere Beschreibung der Erstellungsmethode findet sich in Biemann (2012).

### 3.3 Qualitätssicherung beim Crowdsourcing

Schon mehrfach wurden die Wichtigkeit von Redundanz bei der Durchführung von Crowdsourcing-Aufgaben sowie die Notwendigkeit der Aggregation von Antworten erwähnt. Ein Grundprinzip beim Crowdsourcing ist, niemals einem einzelnen Crowdarbeiter zu vertrauen – dieser könnte absichtlich zufällig antworten, um schnell an die Bezahlung zu kommen –, sondern immer irgendeine Art von Qualitätssicherung einzusetzen (Eickhoff/de Vries 2011), dies ganz nach dem Motto: "Never trust a single turker – trust the crowd!" Diesem wichtigen Thema, welches als Hauptproblem bei der Erstellung von Daten mit Crowdsourcing angesehen wird, wenden wir uns in diesem Abschnitt zu.

Ein früh eingesetzter Mechanismus zur Qualitätsüberprüfung sind die Berechnung des Durchschnittswertes und das Ausschließen von notorisch abweichenden Crowdarbeitern, eingesetzt beispielsweise von Amazon für die Relevanzbewertung von Produkt-Suchergebnissen auf einer Skala von 1 bis 5. Doch die Crowd lernt schnell: Antworten im mittleren Relevanzbereich haben die geringste Durchschnittsabweichung und je mehr Crowdarbeiter absichtlich den Mittelbereich wählen, desto näher ist diese Antwort wiederum am Durchschnittswert – ganz im Gegensatz zu echten Bewertungen, welche durch diesen Mechanismus sogar bestraft werden. Aus diesem Grund wird auf Crowdarbeiterforen wie dem Mechanical Turk Forum (<http://turkers.proboards.com/>) oder Turker Nation (<http://www.turkernation.com/>) vor Auftraggebern mit diesem Mechanismus gewarnt.

Einen systematischen Ansatz zur Qualitätssicherung verfolgt z. B. die Plattform Crowdfunder (<http://crowdfunder.com/>). Hier werden gezielt Testaufgaben in die Menge der eigentlich zu bearbeitenden Aufgaben eingestreut. Crowdarbeiter mit einer zu hohen Fehlerrate werden von der weiteren Bearbeitung ausgeschlossen. Dieses Vorgehen hat jedoch auch Nachteile: So muss auch die Arbeit an den Testaufgaben bezahlt werden, und Antworten von Crowdarbeitern, die insgesamt nur wenige Aufgaben eines Typs bearbeiten, können aufgrund einer geringen Anzahl erledigter Testaufgaben nur schwer auf ihre Qualität hin beurteilt werden. Ein mehrstufiges Modell zur Einschätzung von Crowdarbeitern präsentieren Tai et al. (2011): Nur Crowdarbeitern, die bei einer kleineren Menge gute Arbeit leisteten (hier gemessen durch Testaufgaben), wird es erlaubt, eine größere Anzahl von Aufgaben dieser Art zu bearbeiten. Durch dieses Vorgehen konnte die Durchschnittsfehlerrate um etwa die Hälfte gesenkt werden.

Eine aufwendigere Lösung zur Erhöhung der Datenqualität ist das Durchführen von Validierungsaufgaben, welche wiederum mit Crowdsourcing durchgeführt werden. Das Projekt Anawiki (Chamberlain et al. 2009) zur Annotation von Koreferenzen geht zweistufig vor: Erst werden Koreferenzen im Text frei annotiert, dann werden sie einzeln in einer nachgeschalteten Validierungsaufgabe verifiziert, welche durch die binäre Entscheidung mit relativ geringer Redundanz auskommt. Im Extremfall kann die Crowd als ein initialer Pool verstanden werden, aus dem mit Pilotaufgaben gezielt kooperative und hochqualitativ arbeitende Crowdarbeiter mittels manueller Überprüfung rekrutiert werden, wie es beim im vorhergehenden Abschnitt beschriebenen Projekt (Biemann 2012) geschehen ist.

### 3.4 Kritische Würdigung des Crowdsourcings

Wann ist also der Einsatz von Crowdsourcing für die linguistische Forschung zu empfehlen? Das wichtigste Kriterium ist das Vorhandensein einer Möglichkeit, die Qualität durch Mechanismen der Redundanz oder der Verifikation durch Folgeaufgaben sicherzustellen (vgl. Eickhoff/de Vries 2011). Dies bedeutet im Allgemeinen, dass komplexe Aufgaben in kleine Teilschritte zerlegt werden müssen, welche nur eine geringe Anzahl von Antworten zulassen. Schwierig ist dies insbesondere bei Aufgaben, die die Sprachproduktion von Crowdarbeitern fordern. Es muss möglich sein, die Aufgabenstellung kurz und prägnant darzustellen. Die Bearbeitungszeit pro Einheit sollte möglichst kurz sein. Ferner dürfen die Aufgaben nicht zu repetitiv sein, da Crowdarbeiter als langweilig empfundene Aufgaben nur in kleinen Volumina durchführen. Vorhandene Methoden zur linguistischen Datenerhebung können also nicht direkt in den Crowdspace übertragen werden, sondern benötigen meist eine deutliche Überarbeitung hinsichtlich dieser Richtlinien.

Eine weitere Hürde ergibt sich, wenn die Versuchspersonengruppe örtlich oder nach demografischen Parametern kontrolliert werden soll, da Crowdarbeiter gern anonym bleiben und freiwillige Angaben nicht überprüft werden können. Dieses Problem der mangelnden Repräsentativität ergab sich schon für die Online-Befragung und wird beim Crowdsourcing einerseits durch die wachsende Anzahl von Crowdarbeitern aus prekären Verhältnissen noch verschärft. Andererseits können mit dem Ersetzen der Freiwilligkeit als intrinsischer Motivation durch pekuniäre Vergütung auch neue Zielgruppen erschlossen werden. Mit Bezahlung können überdies höhere Datenvolumina realisiert werden; die finanziellen Anreize führen aber eben auch zum Missbrauch – Crowdarbeiter können nicht als kooperativ angenommen werden.

Häufig wurde Crowdsourcing als Ausbeutungsinstrument kritisiert, da die meist geringe Bezahlung für viele Aufgaben zu Stundenlöhnen unterhalb der Mindestlohngrenze führt (siehe Fort et al. 2011). Instrumentarien zum Errechnen des Stundenlohnes stehen auf Crowdsourcing-Plattformen zur Verfügung; es obliegt unserer Verantwortung als Auftraggeber, für ordentliche Arbeit auch eine ordentliche Bezahlung sicherzustellen. Insbesondere wenn es gelingen soll, mithilfe von Crowdsourcing hochqualitative linguistische Datenerhebungen durchzuführen, ist der Aufbau einer positiven Reputation in der Crowd-Community notwendige Voraussetzung für den Projekterfolg.

## 4 Korpusanalyse: Das World Wide Web als Datenquelle der Lexikografie

### 4.1 Webkorpora

Betrachtet man die inhaltlichen Arbeiten bei der Vorbereitung einer Online-Befragung und beim Crowdsourcing, so steht der Erstellung **eines** Fragebogens bei der Online-Befragung die (zum größten Teil automatisierte) Bereitstellung von **vielen** kleinen Aufgaben für die Worker gegenüber. Hier findet schon eine Verlagerung eines Teils der Arbeiten in das Vorfeld statt. Dies ist noch stärker der Fall bei der Korpusanalyse: Das später genutzte Korpus muss erstellt werden; außerdem müssen die der jeweiligen Fragestellung entsprechenden Analysetools bereitgestellt werden. Im günstigsten Falle stehen sowohl ein geeignetes Korpus wie auch die Analysetools zur Verfügung. Vergleichbar mit einem Baukastensystem können dann sowohl

Korpus wie Analysetools aus einem vorhandenen Reservoir entnommen werden und die notwendigen Vorarbeiten beschränken sich auf ein Minimum. Andernfalls umfassen Korpuserstellung und/oder Programmierung der Analysetools einen erheblichen Anteil an der gesamten Korpusarbeit. Da aber auch im günstigen Fall vorhandener Korpora und Tools diese Arbeiten von anderen erledigt wurden, bleibt die Aussage richtig, dass eine Korpusanalyse vergleichsweise viele Arbeiten im Vorfeld benötigt oder wenigstens nutzt.

Hier drängt sich die Idee geradezu auf, das Web als Korpus und eine Suchmaschine als Analysetool zu verwenden. Für jede Eingabe erhalten wir eine Trefferanzahl. Damit können wir vermutlich

- die Häufigkeit einzelner Wörter exakt bestimmen und
- die exakten Häufigkeiten für das gemeinsame Auftreten mehrerer Wörter (unmittelbar nebeneinander oder auch diskontinuierlich) bestimmen.

Leider ist diese Vorstellung falsch; die erhaltenen Zahlen lassen sich höchstens als grobe Näherungen betrachten und nur schwer vergleichen. Diese Probleme wurden bereits ausführlich diskutiert (Kilgarriff 2007) und sollen hier noch einmal kurz zusammengefasst werden:

- Suchmaschinen zählen Trefferdokumente, nicht einzelne Vorkommen. Es wird also nicht unterschieden, ob beispielsweise ein gesuchtes Wort einmal oder mehrfach vorkommt.
- Die Trefferanzahlen sind nicht stabil, bei einer späteren Suche können sich die Angaben ändern.
- Suchmaschinen berücksichtigen Duplikate von Texten im Internet und verfälschen so die Trefferzahlen.
- Trefferdokumente sind nicht unbedingt Texte, sondern auch Listen oder Quelltext in irgendeiner Programmiersprache.
- Die Suche lässt sich nicht zuverlässig auf Dokumente in einer Sprache beschränken; eine solche Beschränkung (egal wie gut) ist seitens der Suchmaschinen ohnehin nur für wenige Sprachen vorgesehen.
- Die Syntax für Anfragen an eine Suchmaschine ist nur schlecht dokumentiert. Manche Suchmaschinen benutzen eine eigene Form der Lemmatisierung, die sich nicht abschalten lässt. Die Wirkung von Operatoren wie + und –, Anführungszeichen sowie Groß-/Kleinschreibung ist nicht immer klar. Beispielsweise liefert (gefunden im Frühjahr 2012) die Suche bei Google nach
  - *Haus* –*Haus* Trefferdokumente mit den Wörtern *Häüs*, *Häus*, *håus* usw.
  - *Weltz* –*Weltz Leipzig* auch Trefferdokumente zum *Theater der Jungen Welt Leipzig*.
- Suchmaschinen erlauben nur eine gewisse Anzahl von Anfragen pro Nutzer. Auch für maschinelle Abfragen eingerichtete APIs sind auf einige Tausend Abfragen pro Tag beschränkt oder kostenpflichtig.
- Über die Ausgewogenheit des Webs als Korpus ist fast nichts bekannt.

Darüber hinaus gibt es Typen von Anfragen, die sich nicht mithilfe von Suchmaschinen beantworten lassen, weil dies eine zusätzliche Vorverarbeitung seitens der Suchmaschine erfordern würde:

- Welches sind die häufigsten Wortformen?
- Wie lang ist die durchschnittliche Wortlänge bzw. Satzlänge?
- Wie viele verschiedene Wortformen einer Sprache treten im Korpus auf?
- Welche Wörter treten in diesem Jahr deutlich häufiger auf als im Vorjahr?
- Welche Wörter sind veraltet?
- Welches sind die signifikantesten Nachbarn eines Wortes (zu Nachbarschaftskookkurrenzen siehe Heyer et al. 2005)?

Entscheidet man sich deswegen, ein eigenes Korpus zu erstellen, so bietet sich das Web als Datenquelle für die Texte eines eigenen Korpus immer noch an. Diese zweite Nutzungsmöglichkeit des Webs als Datenquelle ohne die Suchmaschinen als Analysetool soll daher im Folgenden untersucht werden. Die Grundvoraussetzung, das Vorhandensein von Texten vorzugsweise in einem einheitlichen und leicht zu verarbeitenden elektronischen Format wie HTML, ist damit erfüllt. Für das Korpus werden Texte benutzt, die von den Autoren für andere Zwecke verfasst wurden und die meist mit Urheberrechten versehen sind. Neben der Möglichkeit der technischen Nutzung sind also zusätzlich rechtliche Aspekte zu beachten.

Jedoch verfasst der Autor seinen Text nicht in dem Wissen oder der Annahme, dass sein Text für ein Korpus genutzt wird. In diesem Sinne sind die Textdaten völlig unverfälscht, da der Autor in keinem Zusammenhang mit dem späteren Analyseinteresse steht. Ob ein bestimmter Text jedoch für ein zu erstellendes Korpus genutzt wird, hängt vom Korpusersteller ab. Dieser hat durch seine Auswahl erheblichen Einfluss auf die Zusammensetzung und ist damit für die Ausgewogenheit selbst verantwortlich. Aus einer großen Dokumentensammlung wie dem Web kann sowohl zufällig wie auch nach vorgegebenen Kriterien ausgewählt werden.

#### **4.2 Der Prozess der Korpuserstellung**

In diesem Abschnitt sollen zwei unterschiedliche Herangehensweisen betrachtet werden. Die erste Methode dient dazu, ein (meist nicht sehr großes) Korpus zu erstellen, um damit anschließend spezielle Fragestellungen korpusbasiert zu bearbeiten. In der Regel ist der Korpusersteller auch der spätere Korpusnutzer. Diese Herangehensweise hat die folgenden zwei Vorteile: Der Korpusnutzer besitzt die volle Information über das Korpus, da er es selbst erstellt hat. Außerdem kann er bei der Erstellung pragmatisch vorgehen und einzelne der unten aufgeführten Arbeitsschritte weglassen, wenn er sich sicher ist, dass dadurch das Ergebnis seiner Untersuchungen nicht beeinträchtigt wird. Diesem Vorteil gegenüber steht der Nachteil, dass die Korpuserstellung (speziell für Nichtinformatiker) mit einem nicht zu unterschätzenden Zeitaufwand verbunden ist.

Bei der zweiten Methode wird das Korpus so erstellt, dass es sich später für möglichst viele Untersuchungen verwenden lässt. Hier sind die Qualitätsanforderungen höher, da die



Verwendungsmöglichkeiten nicht unnötig eingeschränkt werden sollen. Auch rechtfertigt die Trennung von Korpuserstellung und späterer vielfältiger Nutzung einen höheren Aufwand.

Die folgende Liste zeigt die Arbeitsschritte bei der Korpuserstellung (für Details siehe Quasthoff 2010a):

- Erstellen der URL-Liste;
- Download;
- HTML-Stripping;
- Sprachsortierung: Aussortieren von Texten in falschen Sprachen;
- optional: Identifikation von Sätzen (als Bestandteile von Text);
- optional: Musterbasierte Auswahl der wohlgeformten Sätze;
- optional: Löschen von Dubletten und ggf. auch Quasi-Dubletten;
- optional: Qualitätssicherung mit statistischen Messungen;
- optional: Bearbeitung mit linguistischen Tools (z. B. Lemmatisierung, NER, POS-Tagging, Chunking, Parsing);
- optional: Vorberechnung von Daten (z. B. Wortfrequenzen, Kookkurrenzen, Wortähnlichkeiten).

Immer wieder diskutiert wird die Frage nach der notwendigen Größe eines Korpus. Größere Korpora verlangen bei der Erstellung mehr Aufwand und erlauben weniger Kontrolle, da der Benutzer nicht mehr alle Texte des Korpus persönlich inspizieren kann. Auch erfordert die Suche in größeren Korpora mehr Aufwand und dauert deshalb oft länger. Im Gegenzug enthalten größere Korpora mehr Material, die Anzahl der Treffer bei einer Suche ist also größer. Mehr Treffer erhöhen die Sicherheit statistischer Aussagen; bei der Suche nach seltenen Objekten wird man überhaupt erst in größeren Korpora fündig. Diese Hinweise erlauben natürlich noch keine konkrete Abschätzung der benötigten Korpusgröße. Eine zahlenmäßige Abschätzung ergibt sich aus der Beobachtung, dass für viele Fragestellungen 10–20 Vorkommen entsprechender Objekte benötigt werden, um sie sicher zu identifizieren oder zu klassifizieren. Dies betrifft beispielsweise

- die Erkennung von Eigennamen (NER),
- die Ermittlung signifikanter Wortkookkurrenzen,
- die Erkennung semantischer Ähnlichkeit von Wörtern und
- die Erkennung von Phraseologismen.

Damit lassen sich verschiedene Fragestellungen danach klassifizieren, ob sie ein großes Korpus erfordern.

Ein kleines Korpus ist ausreichend für

- die Bestimmung der häufigsten Wörter für ein Lerner-Wörterbuch,
- Zeichenstatistiken verschiedener Art sowie

- die Bestimmung der durchschnittlichen Wort- und Satzlänge.

Hingegen ist ein großes Korpus nötig für die Suche nach

- stark unterrepräsentierten Bereichen der Sprache (z. B. gesprochene Sprache, Mundart, veraltete Wörter/Wendungen) und
- selteneren Phraseologismen oder seltenen Bedeutungen.

Ein Korpus kann dann als ausreichend groß betrachtet werden, wenn exemplarische Tests ergeben, dass die entsprechenden Wörter oder Wortgruppen im Korpus 10- bis 20-mal vorkommen. Für spezielle Fragestellungen können natürlich auch kleinere Anzahlen ausreichen oder größere Anzahlen nötig sein.

### 4.3 Beispielprojekt: Datenbeschaffung für ein Kollokationswörterbuch

Im Jahr 2005 stand einer der Autoren dieses Artikels vor der Frage, ob die damals im *Projekt Deutscher Wortschatz* ([wortschatz.uni-leipzig.de](http://wortschatz.uni-leipzig.de)) vorliegenden Korpusdaten ausreichend seien, um ein Kollokationswörterbuch für die deutsche Sprache zu erstellen. Das Korpus bestand aus 50 Millionen Sätzen mit rund 850 Millionen laufenden Wörtern, hauptsächlich von Online-Ausgaben von Zeitungen und anderen Internet-Medien. Neben den einfachen Wortfrequenzen wurden hier auch Kookkurrenzinformationen ausgewertet: Für jedes Stichwort sind alle Wörter verzeichnet, die statistisch auffällig mit dem Stichwort zusammen auftreten: entweder als unmittelbare linke oder rechte Nachbarn (Nachbarschaftskookkurrenzen) oder gemeinsam im Satz (Satzkookkurrenzen) (siehe auch Heyer et al. 2005).

Entsprechend dem Kollokationsbegriff von Hausmann (1985) handelt es sich bei Kollokationen um Paare von Wörtern, deren gemeinsames Auftreten typisch ist und die gewissen syntaktischen Mustern folgen, beispielsweise *Milch + holen* oder *erheblich(e) + Bedenken*. Es zeigt sich, dass solche Kollokationen stets auch Nachbarschaftskookkurrenzen sind, umgekehrt sind aber nicht alle Nachbarschaftskookkurrenzen auch Kollokationen. Da sich in einem größeren Korpus mehr Nachbarschaftskookkurrenzen identifizieren lassen, stellte sich die Frage, ob die Größe des bereitstehenden Korpus ausreichend war. Dies ließ sich folgendermaßen testen: Für einige Wörter wurden die wichtigsten Kollokationen zunächst von Hand zusammengestellt. Danach wurde geprüft, ob diese sich unter den automatisch ermittelten Nachbarschaftskookkurrenzen befinden. Dieser Test fiel auch für seltenere Stichwörter positiv aus; deshalb wurde das vorgeschlagene Verfahren als praktikabel und die Korpusgröße als ausreichend eingeschätzt. Der weiter gehende Versuch, zu einem Stichwort aus der Liste seiner Nachbarschaftskookkurrenzen mit automatischen Mitteln die tatsächlichen Kollokationen zu ermitteln, schlug allerdings fehl. Zwar lassen sich die Nachbarschaftskookkurrenzen zu einem Stichwort entsprechend verschiedenen Signifikanzmaßen ordnen, aber keine der Ordnungen bevorzugt Kollokationen. Damit waren die Daten für das Kollokationswörterbuch (Quasthoff 2010b) bereitgestellt. In einem letzten Schritt wurden die Nachbarschaftskollokationen zu den Stichwörtern nur entsprechend ihrer Wortart ausgewählt und anschließend manuell begutachtet. Dabei wurden die Kollokationen identifiziert und für die Wörterbuchartikel sortiert.

Verglichen mit einem traditionellen Vorgehen ergeben sich folgende Veränderungen: Die intellektuelle Suche nach den Kollokationen eines Stichwortes entfällt. Stattdessen liefert die Korpusanalyse eine Kandidatenliste, welche außer den gewünschten Kollokationen weitere ungewünschte Wörter enthält. Die Tätigkeit des Bearbeiters reduziert sich darauf, eine Kandidatenliste durchzusehen und geeignete Kandidaten zu entnehmen. Diese Tätigkeit ist einfacher, da es nur um das Wiedererkennen allgemein bekannter Kollokationen geht und nicht mehr darum, die Kollokationen zu einem Stichwort ohne weitere Hilfsmittel möglichst vollständig aufzuzählen.

#### **4.4 Kritische Würdigung der Korpusanalyse**

Der Einsatz von Korpora in der Lexikografie erlaubt es, zumindest einen Teil der vorher manuellen Arbeit durch andere, einfachere und weniger zeitaufwendige Arbeiten zu ersetzen. Typischerweise werden Daten aus dem Korpus generiert und dem Lexikografen als Vorschläge angeboten. Nach erfolgreicher Korpusarbeit kann der Lexikograf sicher sein, dass alle auszuwählenden Objekte in der Vorschlagsliste enthalten sind, also keine eigenen Vorschläge mehr erarbeitet werden müssen. Selbst wenn bei dieser Tätigkeit später bis zu 90% der automatisch generierten Vorschläge verworfen werden, bleibt eine enorme Zeitersparnis. Zum gegenwärtigen Zeitpunkt lassen sich solche Vorschlagslisten natürlich nur aus den Objekten des Korpus bilden, also Vorschläge für Wörter (als Lemma für das Wörterbuch oder als Wörter, die einem Lemma zugeordnet werden, wie Kollokationen oder Synonyme) oder Sätze (z. B. als Belegstellen). Weitere Tätigkeiten des Lexikografen wie die Organisation der Wörterbucheinträge, das Ergänzen von Definitionen, Verweisen usw. bleiben davon unberührt und wurden hier deshalb nicht betrachtet.

### **5 Möglichkeiten und Beschränkungen webbasierter linguistischer Forschung**

Aus den vorangehenden Ausführungen ist deutlich geworden, dass die drei hier vorgestellten Formen webbasierter linguistischer Datenerhebung im Hinblick auf die Anforderungen an die Forschenden folgende Gemeinsamkeiten aufweisen:

Die Aufgabe des Linguisten besteht in jedem Fall darin, aus dem jeweiligen komplexen Forschungsinteresse solche elementaren Fragestellungen zu extrahieren, die sich entweder von linguistisch interessierten Internetnutzern oder gar mit automatischen Verfahren bearbeiten lassen. Zu den so ermittelten Fragestellungen werden im nächsten Schritt Aufgaben (entweder direkt vom Forscher oder in Form automatisch erstellter und zu bewertender Datensätze) zur Bearbeitung bereitgestellt. Das Problem der Qualitätssicherung lässt sich in den meisten Fällen mittels Redundanz lösen: Mehr Worker und mehr Daten können eventuelle Schwächen bei der Bearbeitung durch bestimmte Worker oder Algorithmen aufdecken und ausgleichen.

In der Regel linguistisch nicht geschulte Sprachbenutzer liefern freiwillig, gegen Bezahlung oder ohne ihr Wissen umfangreiches Material, aus dem die Forschenden möglichst linguistische Allgemeingültigkeiten ableiten möchten. Insofern taucht immer wieder die Frage nach der Zuverlässigkeit oder Reliabilität dieser Datengrundlage auf. Im Text haben wir verschiedene Möglichkeiten der Qualitätssicherung angesprochen: Im Falle der am Sprachbenutzer orientierten Methoden der Online-Befragung und des Crowdsourcings wird

nicht den Angaben einzelner Probanden bzw. Crowdarbeiter vertraut, sondern werden Daten von mehreren Beteiligten aggregiert und die Gestaltung der Fragen bzw. Aufgaben wird auf dieses Format abgestimmt (z. B. durch einen beschränkten Umfang mit relativ kurzer Bearbeitungszeit, möglichst wenig repetitive Aufgaben, Einbau von Testaufgaben). Im Falle der am Sprachprodukt orientierten Methode Webkorpus besteht ein Trade-off zwischen Korpusgröße und den Möglichkeiten der Nachbearbeitung und Kontrolle der Rohdaten (z. B. Löschen von Dubletten, Qualitätssicherung durch statistische Messungen). Die Größe wird daher möglichst gut auf die Fragestellung bzw. den Verwendungszweck des Korpus abgestimmt.

Der Kritikpunkt der mangelnden Repräsentativität von Befragten, Crowdarbeitern bzw. im Webkorpus enthaltenen Texten kann nur durch eine möglichst breite Erfassung der für eine Fragestellung relevanten Einflussfaktoren dokumentiert werden, sodass bei einer statistischen Auswertung die Möglichkeit besteht, für diese Faktoren zu korrigieren.

Die Validität dieser Methoden kann prinzipiell mittels exemplarischer Vergleiche mit den Ergebnissen traditioneller Methoden, d. h. eines postalischen Versands, eines Tests von Probanden oder eines ausgewogenen linguistischen Korpus, überprüft werden. Allerdings stellen diese konventionellen Methoden wegen ihres vergleichsweise beschränkten Datenumfanges häufig keine echte Alternative dar. Vor dem Hintergrund begrenzter Ressourcen stehen wir nicht selten vor der Möglichkeit, webbasierte Methoden zu nutzen oder keine ausreichende Datengrundlage zu haben, um die entsprechende Frage zu beantworten. Unter einer sorgfältigen Berücksichtigung der in diesem Artikel angeführten Punkte, die für eine Qualitätssicherung zur Verfügung stehen, scheint uns die webbasierte Gewinnung von Massendaten eine große Chance zu eröffnen, um bisher nicht oder nicht zufriedenstellend beantwortete Fragen in Angriff nehmen zu können. Über die in diesem Artikel exemplarisch dargestellten Bereiche der Dialektologie, der Semantik und der Lexikografie hinausgehend sind die vorgestellten webbasierten Methoden für die verschiedensten gegenwartssprachlichen Fragestellungen der allgemeinen, der angewandten und der vergleichenden Sprachwissenschaft denkbar, die bisher mithilfe der klassischen Methoden Befragung, Experiment und Korpusanalyse untersucht wurden. Ihren Nutzen zeigen sie besonders im Bereich quantitativer Fragestellungen.

### Literaturverzeichnis

- Bandilla, Wolfgang (1999): "WWW-Umfragen – Eine alternative Datenerhebungstechnik für die empirische Sozialforschung?". In: Batinic, Bernad et al. (eds.) (1999): *Online Research. Methoden, Anwendungen und Ergebnisse*. Göttingen usw., Hogrefe: 9–19.
- Biemann, Chris (2012): "Creating a system for lexical substitutions from scratch using crowdsourcing". *Language Resources and Evaluation* 47/1: 97–122.
- Bloodgood, Michael/Callison-Burch, Chris (2010): "Using Mechanical Turk to Build Machine Translation Evaluation Sets". *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk (CSLDAMT '10)*, Los Angeles, California: 208–211.
- Chamberlain, Jon/Kruschwitz, Udo/Poesio, Massimo (2009): "Constructing An Anaphorically Annotated Corpus With Non-Experts: Assessing The Quality Of Collaborative Annotations". *Proc. ACL-IJCNLP 09, Singapore*: 57–62.

- Chen, David L./Dolan, William B. (2011): "Collecting Highly Parallel Data for Paraphrase Evaluation". *Proceedings of the 49<sup>th</sup> Annual Meeting of the Association for Computational Linguistics*, Portland, Oregon: 190–200.
- Christen, Helen (1988): *Sprachliche Variation in der deutschsprachigen Schweiz: dargestellt am Beispiel der I-Vokalisierung in der Gemeinde Knutwil und in der Stadt Luzern*. Wiesbaden: Steiner.
- Eichhoff, Jürgen (1982): "Erhebung von Sprachdaten durch schriftliche Befragung". In: Besch, Werner/Knoop, Wolfgang/Putschke, Wolfgang (eds.) (1982): *Dialektologie. Ein Handbuch zur deutschen und allgemeinen Dialektforschung*. Berlin usw., de Gruyter: 549–554.
- Eickhoff, Carsten/de Vries, Arjen P. (2011): "How Crowdsourcable is Your Task?". *Proceedings of the Workshop on Crowdsourcing for Search and Data Mining (CSDM) at the Fourth ACM International Conference*.
- Elspaß, Stephan/Möller, Robert (2006): "Internet-Exploration: Zu den Chancen, die eine Online-Erhebung regional gefärbter Alltagssprache bietet". *Osnabrücker Beiträge zur Sprachtheorie* 71: 141–156.
- Fort, Karén/Adda, Gilles/Cohen, K. Bretonnel (2011): "Amazon mechanical turk: Gold mine or coal mine?". *Computational Linguistics* 37/2: 413–420.
- Giles, Jim (2005): "Internet encyclopaedias go head to head". *Nature* 438: 900–901.
- Hausmann, Franz Josef (1985): "Kollokationen im deutschen Wörterbuch. Ein Beitrag zur Theorie des lexikographischen Beispiels". In: Bergenholtz, Henning/Mugdan, Joachim (eds.) (1985): *Lexikographie und Grammatik. Akten des Essener Kolloquiums zur Grammatik im Wörterbuch vom 28. bis 30.06.1984*. Tübingen, Niemeyer: 118–129.
- Heyer, Gerhard/Quasthoff, Uwe/Wittig, Thomas (2005): *Text Mining: Wissensrohstoff Text: Konzepte, Algorithmen, Ergebnisse*. Bochum: w3l-Verlag.
- Hong, Jisup/Baker, Collin F. (2011): "How Good is the Crowd at 'real' WSD?". *Proceedings of the Fifth Law Workshop (LAW V), Portland, Oregon*: 30–37.
- Hotzenköcherle, Robert et al. (1962–1997): *Sprachatlas der deutschen Schweiz*. Bern/Basel: Francke.
- Howe, Jeff (2006): "The Rise of Crowdsourcing". *Wired* 14. <http://www.wired.com/wired/archive/14.06/crowds.html>, Stand: 17. Oktober 2013.
- Ipeirotis, Panos (2010): "Demographics of Mechanical Turk". *NYU Working Papers: CeDER-10-01*.
- Jha, Mukund et al. (2011): "Corpus Creation for New Genres: A Crowdsourced Approach to PP Attachment". *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk (CSLDAMT '10), Los Angeles, California*: 13–20.
- Juska-Bacher, Britta (2009): *Empirisch-kontrastive Phraseologie. Am Beispiel der Bekanntheit von Bruegels Niederländischen "Sprichwörtern" im Niederländischen, Deutschen und Schwedischen*. Baltmannsweiler: Schneider Verlag Hohengehren.
- Juska-Bacher, Britta (2010a): "Wortgeografischer Wandel im Schweizerdeutschen. Sommersprossen, Küchenzwiebel und Schmetterling 70 Jahre nach dem SDS". *Linguistik online* 41: 19–43.
- Juska-Bacher, Britta (2010b): "SDS-Exploratoren und Online-Befragung – Lässt sich im Methodenmix ein Wandel in der Schweizer Dialektlandschaft nachweisen?". In: Christen,

- Helen et al. (eds.) (2010b): *Dialektologie: Wege in die Zukunft*. Stuttgart, Steiner: 279–293.
- Juska-Bacher, Britta (2011): "Helvetismen: nationale und areale Varianten? Kodifizierung und sprachliche Realität". *Yearbook of Phraseology* 2: 71–108.
- Kilgarriff, Adam (1999): "I don't believe in word senses". *Computers and the Humanities* 31/2: 91–113.
- Kilgarriff, Adam (2007): "Googleology is Bad Science". *Computational Linguistics* 33/1: 147–151.
- Klosa, Annette/Koplenig, Alexander/Töpel, Antje (2011): "Benutzerwünsche und Meinungen zu einer optimierten Wörterbuchpräsentation – Ergebnisse einer Onlinebefragung zu lexiKo". *OPAL* 3: 1–35.
- Kunath, Stephen A./Weinberger, Steven H. (2010): "The Wisdom of the Crowds Ear: Speech Accent Rating and Annotation with Amazon Mechanical Turk". *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk (CSLDAMT '10)*, Los Angeles, California: 168–171.
- Lane, Ian et al. (2010): "Tools for Collecting Speech Corpora via Mechanical-Turk". *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk (CSLDAMT '10)*, Los Angeles, California: 184–187.
- Little, Greg et al. (2009): "TurKit: Tools for Iterative Tasks on Mechanical Turk". *Proceedings of the ACM SIGKDD Workshop on Human Computation (HComp2009)*: 29–30.
- Madhani, Nitin et al. (2011): "They Can Help: Using Crowdsourcing to Improve the Evaluation of Grammatical Error Detection Systems". *Proceedings of the 49<sup>th</sup> Annual Meeting of the Association for Computational Linguistics: Shortpapers, Portland, Oregon, June 19–24, 2011*: 508–513.
- Maurer, Marcus/Jandura, Olaf (2009): "Zehn Jahre Sozialforschung mit dem Internet – eine Analyse zur Nutzung von Online-Befragungen in den Sozialwissenschaften". In: Jakob, Nikolaus/Schoen, Harald/Zerback, Thomas (eds.) (2009): *Sozialforschung im Internet. Methodologie und Praxis der Online-Befragung*. Wiesbaden, VS Verlag für Sozialwissenschaften: 61–73.
- Mitzka, Walter/Schmitt, Ludwig Erich (1951–1980): *Deutscher Wortatlas*. Gießen: Wilhelm Schmitz.
- Mohammad, Saif M./Turney, Peter D. (2011): "Crowdsourcing a Word-Emotion Association Lexicon". *Computational Intelligence* 59: 1–24.
- Niebaum, Hermann/Macha, Jürgen (2006): *Einführung in die Dialektologie des Deutschen*. Tübingen: Niemeyer.
- Novotney, Scott/Callison-Burch, Chris (2010): "Cheap, Fast and Good Enough: Automatic Speech Recognition with Non-Expert Transcription". *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk (CSLDAMT '10)*, Los Angeles, California: 207–215.
- Parent, Gabriel/Eskenazi, Maxine (2010): "Clustering dictionary definitions using Amazon Mechanical Turk". *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk (CSLDAMT '10)*, Los Angeles, California: 21–29.

- Quasthoff, Uwe (2010a): "Automatisierte Rohdatengewinnung für die Lexikographie". *Jahrbuch Lexicographica* 26: 47–64.
- Quasthoff, Uwe (2010b): *Wörterbuch der Kollokationen im Deutschen*. Berlin/New York: De Gruyter.
- Ross, Joel et al. (2010): "Who are the Crowdworkers? Shifting Demographics in Amazon Mechanical Turk". *CHI EA 2010*: 2863–2872.
- Rouhizadeh, Masoud et al. (2011): "Collecting semantic data by Mechanical Turk for the lexical knowledge resource of a text-to-picture generating system". *Proceedings of the Ninth International Conference on Computational Semantics (IWCS '11)*: 380–384.
- Schifferle, Hans-Peter (1995): *Dialektstrukturen in Grenzlandschaften. Untersuchungen zum Mundartwandel im nordöstlichen Aargau und im benachbarten südbadischen Raum Waldshut*. Bern usw.: Lang.
- Siebenhaar, Beat (2000): *Sprachvariation, Sprachwandel und Einstellung: Der Dialekt der Stadt Aarau in der Labilitätszone zwischen Zürcher und Berner Mundart*. Stuttgart: Steiner.
- Snow, Rion et al. (2008): "Cheap and Fast – But is it Good? Evaluating Non-Expert Annotations for Natural Language Tasks". *EMNLP '08 Proceedings of the Conference on Empirical Methods in Natural Language*: 254–263.
- Stone, Maria et al. (2011): "A Comparison of On-Demand Workforce with Trained Judges for Web Search Relevance Evaluation". *Proceedings of the ACM SIGIR Workshop on Crowdsourcing for Information Retrieval*: 15–20.
- Taddicken, Monika (2008): *Methodeneffekte bei Web-Befragungen. Einschränkungen der Datengüte durch ein ‚reduziertes Kommunikationsmedium‘?* Köln: Halem.
- Tai, Li et al. (2011): "Quality Control of Crowdsourcing through Workers Experience". *Proceedings of the ACM SIGIR 2011 Workshop on Crowdsourcing for Information Retrieval (CIR 2011), Beijing, China*: 28–35.
- Tourangeau, Roger (2004): "Survey Research and Social Change". *Annual Review of Psychology* 55: 775–801.
- Voyer, Robert et al. (2010): "A hybrid model for annotating named entity training corpora". *Proceedings of the Fourth Linguistic Annotation Workshop (LAW IV '10), Uppsala, Sweden*: 243–246.
- Wenker, Georg et al. (1926–1956): *Deutscher Sprachatlas*. Marburg: Elwert.
- Wolfensberger, Heinz (1967): *Mundartwandel im 20. Jahrhundert. Dargestellt an Ausschnitten aus dem Sprachleben der Gemeinde Stäfa*. Frauenfeld: Huber.
- Zaidan, Omar F./Callison-Burch, Chris (2010): "Predicting human-targeted translation edit rate via untrained human annotators". *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics (HLT '10), Los Angeles, California*: 369–372.
- Zerback, Thomas et al. (2009): "Zehn Jahre Sozialforschung mit dem Internet – eine Analyse zur Nutzung von Online-Befragungen in den Sozialwissenschaften". In: Jakob, Nikolaus/Schoen, Harald/Zerback, Thomas (eds.) (2009): *Sozialforschung im Internet. Methodologie und Praxis der Online-Befragung*. Wiesbaden, VS Verlag für Sozialwissenschaften: 15–31.