

Vera Bachmann, Michael Fleig, Christiane Heibach,
Solveig Ottmann, Silke Roesler-Keilholz (Hg.)
Staunen – Rechnen – Rätseln

Vera Bachmann (Dr. phil.), geb. 1979, lehrt Medienwissenschaft an der Universität Regensburg. Sie promovierte an der LMU München mit einer Arbeit zur Ästhetik der Oberfläche in der Literatur des 19. Jahrhunderts. Ihre Forschungsinteressen betreffen Raummetaphoriken des Medialen, weibliche Kollektive, Geld und Digitalität sowie Narrative der Inflation.

Michael Fleig (Dr. phil.), geb. 1979, ist wissenschaftlicher Mitarbeiter am Lehrstuhl für Medienwissenschaft an der Universität Regensburg. 2018 promovierte er mit einer Arbeit über den französischen Filmemacher Michel Gondry. Seine Schwerpunkte beziehen sich auf den Kurzfilm, digitale Bilder, Bild-Ton-Kopplungen, Early Cinema und mediale Zeitstrukturen.

Christiane Heibach (Prof. Dr. phil. habil.), geb. 1967, ist Professorin für Medienästhetik an der Universität Regensburg. Sie habilitierte sich 2007 an der Universität Erfurt mit einer Arbeit zu Multimedialer Aufführungskunst. Ihre Forschungsschwerpunkte sind Medienästhetik, Epistemologie der Multi- und Intermedialität, Theorie und Ästhetik Digitaler Medien sowie Medienökologien.

Solveig Ottmann (Dr. phil.), geb. 1981, ist akademische Rätin am Lehrstuhl für Medienwissenschaft der Universität Regensburg. Sie promovierte an der Ruhr-Universität Bochum mit einer Dissertation zum Weimarer Radio bei Hans Flesch und Ernst Schoen. Die Schwerpunkte ihrer Forschung und Lehre umfassen die Theorie und Geschichte akustischer Medien, Entwicklungen (in) der digitalen Kultur sowie Extremismus im Netz.

Silke Roesler-Keilholz (Dr. phil.), geb. 1979, ist wissenschaftliche Assistentin am Lehrstuhl für Medienwissenschaft der Universität Regensburg. In ihrem Habilitationsprojekt zu Überwachungsarchitekturen erweitert sie die raumtheoretische Reflexion ihrer Dissertation um eine topologische Dimension. Ihre Arbeitsfelder sind Medien der Über-/Unterwachung, Ästhetiken des Postfeminismus, Raumtheorien sowie Fashion Media Studies.

Vera Bachmann, Michael Fleig, Christiane Heibach,
Solveig Ottmann, Silke Roesler-Keilholz (Hg.)

Staunen – Rechnen – Rätself

Explorationen des Medialen

[transcript]

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

© 2023 transcript Verlag, Bielefeld

Alle Rechte vorbehalten. Die Verwertung der Texte und Bilder ist ohne Zustimmung des Verlages urheberrechtswidrig und strafbar. Das gilt auch für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und für die Verarbeitung mit elektronischen Systemen.

Umschlaggestaltung: Maria Arndt, Bielefeld

Umschlagabbildung: Skulptur »Sker« des Künstlers Peter Lang, fotografiert von Kiron Patka

Satz: Christine Fraunhofer M.A.

Druck: Majuskel Medienproduktion GmbH, Wetzlar

<https://doi.org/10.14361/9783839462621>

Print-ISBN 978-3-8376-6262-7

PDF-ISBN 978-3-8394-6262-1

Buchreihen-ISSN: 2569-2240

Buchreihen-eISSN: 2702-8984

Gedruckt auf alterungsbeständigem Papier mit chlorfrei gebleichtem Zellstoff.

Besuchen Sie uns im Internet: <https://www.transcript-verlag.de>

Unsere aktuelle Vorschau finden Sie unter www.transcript-verlag.de/vorschau-download

Inhalt

Einleitung

Vera Bachmann, Michael Fleig, Christiane Heibach,
Solveig Ottmann & Silke Roesler-Keilholz | 9

STAUNEN

Die dunkle Seite des Mondes

Eine medienphilosophische Miniatur
Stefan Münker | 19

Il est tard – Es ist worden spät

Goethe, Byron, Baudelaire in einem Gedicht Stefan Georges
Eckart Goebel | 27

Europäische Galaxien

Lunare Projektionsflächen von Venedig über Paris nach Berlin und London
und zurück: Die Diskrepanz von Paul Linckes *Frau Luna* und Georges Méliès'
Le Voyage dans la Lune
Doris Gerstl | 45

Der Mond, das Geld und die Nazis

Zur Reflexivität von Medien während des Nationalsozialismus und danach
Hendrik Buhl | 51

Kinematisches Staunen

The Spielberg Face (Kevin B. Lee, USA 2011)
Stefanie Diekmann | 59

Das Rauschen der Räume

Die Krankenhäuser und Arthur Schnitzler
Fatima Naqvi | 67

Das wunderbare Rechnen zum Tode

Paläontologische Kontingenzen, Konvergenzen und Katastrophen
in *The Difference Engine*
Geoffrey Winthrop-Young | 79

Staunen über Foucault

Diskurs – Medium – Fiktion
Achim Geisenhanslüke | 93

RECHNEN

Kann man soziale Medien anschreiben?

Dirk Baecker | 101

Maschinen, die Gedichte schreiben

Künstliche Intelligenz als epistemische und anthropologische Zäsur
Hans-Christian von Herrmann | 109

„And isn't the same as plus.“

J. M. Coetzees Erkundungen des Verhältnisses von Wort und Zahl
Robert Stockhammer | 117

Wahrnehmen und Erinnern, Denken und Vergessen

Jorge Luis Borges' Erzählung *Funes el memorioso*
Wolf Kittler | 137

Techno-Logie als Schauplatz operativen Wissens

Wolfgang Ernst | 157

Der Aufstieg des Kinos zur legitimen Kultur

Ein Beitrag zur Mediengeschichte Frankreichs
Jochen Mecke | 169

Interoperable Medien zwischen Sender and Receiver in der digitalen Medizin

Laura Niebling & Cornelius Borck | 183

Turing-Unvollständigkeit

Moritz Hiller | 197

RÄTSELN

Anfang und Ende der Spekulation

Zur Philosophie und Ökonomie von Technologien der Vorhersage an der Schnittstelle von Medienwissenschaft und Wissenschaftsgeschichte
Andreas Sudmann | 209

Rätsel der Rezeption

Versuche mit Augenbewegungen

Ursula Regener | 219

Gnostisches Wissen und Kosmisches Erbrechen in Nabokovs Roman

Приглашение на казнь (Einladung zur Enthauptung)

Sven Spieker | 239

„Only dead men have sunk below this.“

Zu heterotopen Unterwasserwelten

Solveig Ottmann & Silke Roesler-Keilholz | 249

Distant Listening für Hörfunkspots

Sandra Reimann & Christian Wolff | 261

Die „Druckerei La Bohème“ als Sphinx

Rätsel in der Amberger Zeitung *Baracke!*

Isabella von Treskow | 279

Das Lachen Foucaults

Henning Schmidgen | 289

Chiffren des Ungewussten

Zur Konjunktur der Jahreszahlenbücher

Vera Bachmann | 303

Wünsche und ihre Medien

Stefan Willer | 315

Autor:innenverzeichnis | 325

Dank | 335

Distant Listening für Hörfunkspots

Sandra Reimann & Christian Wolff

Der vorliegende Beitrag versteht sich u. a. als Übertragung einer bereits seit längerer Zeit geführten Methodendiskussion in den digitalen Geisteswissenschaften (*digital humanities*) auf das Forschungsfeld der Hörfunkwerbung. Der Literaturwissenschaftler Franco Moretti hat das Konzept des *distant reading* eingeführt,¹ das seitdem vielfältig aufgegriffen, variiert, aber auch kritisch diskutiert worden ist. Im Kern handelt es sich um zwei Überlegungen: Literarische oder künstlerische Artefakte stehen als Gegenstandsbereich geisteswissenschaftlicher Betrachtung mittlerweile in vielen Fällen digital zur Verfügung. Sie können daher einer rechnergestützten Analyse zugänglich gemacht werden. Gleichzeitig liegen sie in den digitalen Repräsentationsformaten in großen Mengen vor. Daraus entsteht die Denkfigur des *distant reading*: Der Computer mag nicht in der Lage sein, im engeren Sinne zu verstehen, was der Inhalt eines Textes oder eines Bildes ist, aber er kann mit den zur Verfügung stehenden Methoden (zum Beispiel automatische Indexierung, Erkennung von Mustern und visuellen Strukturen, allgemein der Einsatz maschinellen Lernens) andere Mengengerüste verarbeiten und aus ihnen Informationen extrahieren, als dies dem Menschen bzw. der Wissenschaftlerin möglich wäre, die durch *close reading* vergleichsweise weniger Texte intensiver („näher“) liest und studiert. Der Computer kann viele tausend oder gar alle literarischen Werke einer Epoche, Kunstwerke einer Gattung etc. analysieren. Neben dem *distant reading* spricht man auch von *distant viewing*² und *distant watching*³ mit Bezug zur Analyse großer Mengen von Bildern oder Filmen. *Distant listening* einer großen Menge von Radiobeiträgen ist ebenfalls bereits thematisiert worden.⁴

Es liegt auf der Hand, dass auch das *Regensburger Archiv für Werbeforschung* (RAW) mit seinen umfangreichen Beständen – u. a. betrifft das die Sammlung *Historisches Werbefunkarchiv* (HWA) mit rund 50.000 historischen Werbefunksendungen – Gegenstand dieses methodischen Trends in den digitalen Geisteswissenschaften werden könnte. Dabei ist näher zu betrachten, in welcher Form die ursprünglich analog auf Tonbändern oder Schallplatten gespeicherten Werbespots digitalisiert wurden:

1 Vgl. Moretti (2013).

2 Vgl. Arnold & Tilton (2019).

3 Vgl. Burghardt & Wolff (2017).

4 Vgl. Have & Enevoldsen (2021).

In einem ersten Schritt erfolgte die Umwandlung analoger Signale in eine geeignete digitale Repräsentation, zum Beispiel als MP3-Dateien. Eine weitergehende (automatische) Digitalisierung im Sinne der Erkennung von Inhalten ist bisher nicht durchgeführt worden. Dieser Frage, also inwiefern mittlerweile Rechentechnik in der Lage ist, hier einen Beitrag zu leisten, widmet sich der Beitrag. Es geht darum, anhand einer kleinen Stichprobe zu überprüfen, welches Potenzial im Einsatz digitaler Methoden liegt.

Der Ausgangspunkt hierfür ist die gewachsene Leistungsfähigkeit von Verfahren der Spracherkennung, an denen seit vielen Jahrzehnten geforscht wird⁵ und die in vielen Einsatzbereichen mittlerweile erfolgreich im praktischen Einsatz ist.⁶ Die Besonderheit im vorliegenden Fall ist die Heterogenität des Materials: Die Spots sind typischerweise kurz, enthalten neben Sprache auch andere akustische Phänomene (z. B. Musik, Geräusche), werden in vielen Fällen von mehreren Sprecher*innen gesprochen und weitere Effekte (Dialekt, Neologismen, besonders schnelle/langsame Sprache, besondere Akzentuierung) sind für sie typisch. Insgesamt lässt sich das Material mit bisherigen Anwendungsfeldern für die Spracherkennung kaum vergleichen. Zwar gibt es erste Überlegungen und Erfahrungsberichte, wie Spracherkennung wissenschaftliche Prozesse beschleunigen kann, etwa beim Einsatz in der qualitativen empirischen Forschung zur (teil-)automatischen Transkription von Interviews und weiteren sprachlichen Äußerungen.⁷ Für das Deutsche liegen zudem Erkenntnisse zu *oral history*-Interviews vor;⁸ dies unterscheidet sich aber substantiell von der sprachlichen und akustischen Vielfalt der hier vorliegenden Werbespots.

Ziel war es, unterschiedliche, aber leicht verfügbare Spracherkennungssoftware und -dienste auf ihre grundsätzliche Anwendbarkeit für eine derartige Transkription zu untersuchen. Neben drei Varianten, die das akustische Ausgangsmaterial automatisch in Text überführen, wurde auch ein Abschriftendienst getestet, bei dem die Transkription durch Fachkräfte erfolgt und bei dem auch in gewissem Umfang ein Lektorat Bestandteil des Prozesses ist. Dieser wurde deswegen hinzugenommen, weil derartige Dienste mittlerweile vergleichsweise kostengünstig online zugänglich sind und eine weitere Option für die Texterschließung der Aufnahmen im HWA darstellen.

Im Sinne des *distant reading* oder *distant listening* können so unterschiedliche Ziele verfolgt werden: Im Kleinen geht es darum zu prüfen, ob sich die typischerweise sehr

5 Vgl. Furui (2010).

6 Der Koautor Wolff nutzt seit mittlerweile mehr als sechs Jahren Spracherkennung produktiv und hat mittlerweile mehr als 2.000 Sprachaufnahmen in Texte verwandelt, überwiegend Gutachten zu wissenschaftlichen Arbeiten (Abschlussarbeiten, Aufsätze), aber auch Korrespondenz und eigene wissenschaftliche Texte. Auch Teile dieses Aufsatzes sind zunächst als Diktat entstanden. Bei dieser Nutzungsform ist relativ viel Aufwand in das Trainieren der Anwendungssoftware (*Nuance Dragon Dictate V. 15*) auf der Basis verfügbarer Aufnahmen und korrigierter Texte geflossen, u. a. wurde so das Erkennungsvokabular wesentlich erweitert und an die Fachdomäne des Autors angepasst. Diese Möglichkeit steht für die kurzen und heterogenen Hörfunkspots nicht offen.

7 Vgl. dazu bereits Fogg & Wightman (2000).

8 Vgl. Gref, Köhler & Leh (2018).

aufwändige und zeitintensive Transkription von Hörfunkspots durch (Teil-)Automatisierung effizienter gestalten lässt. Daneben ist aber auch zu fragen, ob sich mit vertretbarem Mitteleinsatz die Bestände des *Historischen Werbefunkarchiv* im Sinne der Extraktion gesprochener Sprache als Text vollständig erschließen lassen und damit einer Analyse zugänglich gemacht werden können, bei der jeweils das gesamte Korpus untersucht werden kann, ähnlich wie das insbesondere in der Korpuslinguistik mit ihren mittlerweile digital verfügbaren Korpora üblich geworden ist.⁹

Das Ursprungsrätsel unseres Beitrags ist die mediale, formale, historische und inhaltliche Vielschichtigkeit der Hörfunkspots im *Regensburger Archiv für Werbeforschung*, das sich nur durch die Zusammenarbeit verschiedener Disziplinen lösen lässt. Davon ausgehend rätseln wir, welche Potentiale eine weitergehende Digitalisierung mit sich bringen könnte. Und rätselhaft ist aus Rezipient*innenperspektive – das zeigen unsere Versuche – manch automatische Transkription. Ganz besonders rätseln wir aber, mit welchen Verfahren wir die beste Version der digitalisierten Verschriftung erreichen können und welche Chancen das *distant x-ing-Paradigma* (*distant reading, viewing, listening*) der digitalen Geisteswissenschaften dafür bietet.

ZUR TRANSKRIPTION AUS LINGUISTISCH-WERBEKOMMUNIKATIVER SICHT

Das *Regensburger Archiv für Werbeforschung* beherbergt in der Universitätsbibliothek Regensburg mehrere Sammlungen unterschiedlicher Datenträger mit Werbematerialien; die erste und größte ist das *Historische Werbefunkarchiv* mit rund 50.000 Radiospots überwiegend sehr bekannter Marken aus den Jahren 1948 bis 1987. Es handelt sich dabei um weltweit eines der größten Werbefunkarchive; seit 2003 ist die Sammlung an der Universität Regensburg untergebracht. Eine zeitlich daran anschließende Sammlung mit weiteren rund 50.000 Spots bis zum Jahr 2000 liegt dort ebenfalls vor. Zu den akustischen Kommunikaten zählen auch mehrere Sammlungen von Werbeschallplatten, die als Medium vor allem in den 1950er bis 1970er Jahren zum Einsatz kamen. Weitere Sammlungen mit audiovisuellen Kommunikaten sind vorhanden. Die ältere Hörfunkspotsammlung (HWA) und ein Teil der Werbeschallplatten können über eine Datenbank im Internet recherchiert werden.¹⁰ Weiterhin liegen Metadaten in Form gedruckter Materialien vor, die teilweise wertvolle Einblicke in den Entstehungsprozess der Kampagnen geben.

Die auditiven Bestände des *Historischen Werbefunkarchivs* und weiterer auditiver Sammlungen des RAW müssen bisher händisch transkribiert werden, um sie für weitere Analysen nutzbar zu machen. Diese Aufgabe wurde aus Zeit- und Kostengründen nicht zentral für den ganzen Bestand durchgeführt, sondern musste punktuell nach Bedarf seitens der Forschung selbst geleistet werden.¹¹ Einige Hintergründe seien dazu skizziert: Im Vorfeld der Untersuchung akustischer Werbung sind Herausforderungen zu berücksichtigen, die (nicht nur) die medial gesprochene Sprache

9 Vgl. Hirschmann (2019, S. 226 f.).

10 <https://raw.uni-regensburg.de/>.

11 Vgl. Gerber (2006).

betreffen: Die Flüchtigkeit des Mediums Hörfunk oder – hinsichtlich der verwendeten Darstellungsmittel – vergleichbarer Werbeformen (z. B. die Werbeschallplatte) erschweren eine intensive Beschäftigung mit dem Kommunikat und machen eine Transkription, die zusätzlichen zeitlichen Aufwand und Expertise verlangt, notwendig.

Gemeinhin werden deshalb Zugeständnisse an die Verschriftung gemacht, indem man sich an der Fragestellung bzw. der zu bearbeitenden Aufgabe orientiert, das heißt beispielsweise: Will man den Umgang mit Dialekten untersuchen und eine entsprechende Zuordnung vornehmen, muss der Fokus auf diesen sprachlichen Aspekt gelegt werden und phonetisch-phonologisch transkribiert werden. Geht es nicht um diese lautliche Ebene, ist eine standardnahe Verschriftlichung angemessen, um die Lesbarkeit zu erleichtern, z. B. bei textgrammatischen Untersuchungen zur Kohärenz. Die Einteilung des Sprachtextes in „Minimale textgrammatische Einheiten“ (MTE) bezieht sich dabei auf Sätze oder Setzungen.¹² Da wir es allerdings mit medial gesprochener Sprache zu tun haben, ist eine solche Klassifikation bereits häufig Interpretation und sollte so zurückhaltend wie möglich vorgenommen werden. Die Orientierung an der Verständlichkeit für die Rezipierenden ist dabei hilfreich. So spricht nichts dagegen, eine Entscheidungsfrage (Finitum an erster Stelle) mit einem Fragezeichen zu versehen: *Willst du heute Abend mit mir ins Kino gehen?*. Auch ein Imperativsatz kann mit einem Ausrufezeichen enden: *Gib mir bitte die Butter!*. Aus syntaktischer Sicht sind diese Beispiele eindeutig und die Interpunktion lässt keinen Spielraum für (falsche) Interpretation. Anders dagegen ist es bei Formulierungsmustern, die lediglich kommunikativ-semantic eine Aufforderung oder Frage zum Ausdruck bringen: Die Äußerung *Mir ist kalt!* kann implizit eine Aufforderung zum Schließen eines Fensters sein; syntaktisch ist es ein Aussagesatz und das Ausrufezeichen schon Interpretation. Der Satz *Du gehst heute nicht mit uns ins Kino* wird nur als Frage verstanden, wenn die Intonation eine solche anzeigt. Die oben schon angesprochenen Dialekte stellen weiterführende Herausforderungen für die Transkribierenden dar, zumal die Funktion der Textsorte (Produkt-)Werbung und die Intention des dahinterstehenden Senders, nämlich den Kaufakt hervorzurufen, auch beinhaltet, Verstehensprobleme bei den Rezipierenden zu vermeiden.¹³

Ein frühes Beispiel für die Transkription von Radiowerbespots liefert Hartmut Stöckl. Er nutzt die Spaltenschreibung und folgende Rubriken: Zeit/Sek. – Sprache – Stimme – Musik – Geräusch. Dabei greift er auf die geschriebene Standardsprache zurück und setzt auch Interpunktion ein, die allerdings bereits eine Interpretation voraussetzt: Beispielsweise endet der Satz „Entdecken Sie Biofresh für die ganze Vielfalt gesunder Ernährung.“ mit einem Punkt, wo doch syntaktisch ein Ausrufezeichen im Imperativsatz zu erwarten gewesen wäre. Noch mehr Interpretation verlangen die weiteren Rubriken (mit Ausnahme von Zeit/Sek.), z. B. Stimme: „sonore Männerstimme (stellenweise ungläubig fragend)“, „sonore Männerstimme (Enttäuschung überspielend)“ oder „attraktive Frauenstimme (belehrend, akkurat)“. Bei den Geräuschen nennt Stöckl in seinen Beispielen „Zischen“ oder „Kühlschranktürklappen“ – parallel zum gesprochensprachlichen Text „Wieso steht an unserem neuen

12 Greule & Reimann (2015, S. 80–81).

13 Vgl. Reimann (2009).

Kühlschrank Biofresh“.¹⁴ Der Musikpädagoge Bernhard Hofmann führt in seinem Beitrag über „Musik in der Hörfunkwerbung“ anhand der Transkription eines Hörfunkspots der *Diebels Brauerei* folgende Kategorien im Rahmen paramusikalischer Instrumentalklänge an: „Reifen“: „Inzidenzklang“, „Mjamjamjam“: „Parasprachlicher Stimmlaut“ und „Glockenspiel“: „Paramusikalischer Klang“. Der Sprachtext wurde auch in Buchstabenschrift transkribiert inklusive Interpunktion, die Raum für Interpretation lässt: *„Da gibt's nur Eines: Issumer Alkoholfrei. Das schmeckt. Issumer Alkoholfrei. Man trinke..... und staune.“*¹⁵

Albrecht Greule führt eine Kommentar-Spalte in seine Hörfunkspot-Transkription – in Anlehnung an eine von einer Studierendengruppe angefertigte Verschriftlichung – ein, die die Kategorien Zeile¹⁶, gesprochene Sprache und Musik im vorgeführten Beispiel ergänzt und Platz für subjektiv notwendig erscheinende Angaben macht. Zur Bedeutung des Transkripts schreibt er:

Um methodisch vertretbare Aussagen zur Gestaltung des Spots machen zu können, ist die Transformation des Gehörten vom akustischen Kanal in den visuellen Kanal notwendig, mit anderen Worten: Das Gehörte muss in ein lesbares Transkript formuliert werden. [...] Im Zentrum steht dabei die möglichst genaue schriftliche Wiedergabe der gehörten Sprache (Spalte 2). Sie muss nicht getreu der Orthographie erfolgen. Denkbar wären auch andere Codierungen, z. B. eine in phonetischer Transkription oder eine, die Großbuchstaben bzw. Unterstreichungen einsetzt, um Betonungen zum Ausdruck zu bringen. Es gilt jedenfalls der Grundsatz: Aus der Transkription kann bei der Interpretation nur herausgeholt werden, was vorher an Informationen in sie hineingesteckt wurde.¹⁷

Die Kategorie Kommentar problematisiert er dabei und plädiert für einen sorgsamsten Umgang:

Nicht minder wichtig, aber auch problematisch ist Spalte 4 ‚Kommentar‘. In sie wurden mehr oder weniger subjektive Eindrücke der Analysatorinnen aufgenommen, z. B. ‚verschmitzt, altklug‘. Es handelt sich dabei um metasprachliche Äußerungen, also um Einschätzungen von Sprache bzw. Sprecher/in, für die wir keine objektivierbaren Beschreibungen zur Verfügung haben, mit Hilfe der Sprache.¹⁸

Bei der konventionellen Verschriftung kann man sich auch an den bewährten Transkriptionssystemen der gesprächs-/konversationsanalytischen Ausrichtung der Sprachwissenschaft orientieren.¹⁹ Populär ist die Umschrift des gesprächsanalyti-

14 Stöckl (2007, S. 188–189).

15 Hofmann (2008, S. 53). Weitere Transkriptionsbeispiele siehe bei Kaunzner (2008) und Reimann (2008a).

16 Die Einteilung in Zeilen orientiert sich an der schon erwähnten Klassifikation in Minimale Textgrammatische Einheiten (MTE), die natürlich in der medial gesprochenen Sprache auch einer gewissen Interpretation unterliegt.

17 Greule (2008, S. 90f.).

18 Greule (2008, S. 92).

19 Zur Transkription siehe auch die Beiträge von Redder, Selting sowie Sager in Brinker et al. (2001).

schen Transkriptionssystem (GAT, in der Weiterentwicklung GAT 2), die als Minimal-, Basis- und Feintranskript unterschiedlich aufwendig und detailliert ist und somit gut auf die jeweilige Fragestellung ausgerichtet werden kann:

Berücksichtigt werden dabei – je nach Transkriptionsverfahren – auch paraverbale Komponenten wie Stimme, Intonation, Sprechverhalten (Lautstärke, Sprechtempo) und nonverbale Merkmale wie Mimik, Gestik sowie weitere kontextuelle Informationen (Situation und ‚Rollen‘ der Gesprächspartner). [...] Bei GAT handelt es sich um eine sogenannte ‚literarische Umschrift‘: Es werden nur Besonderheiten von Merkmalen, die nicht dem Standard entsprechen, berücksichtigt, z. B. Assimilationen wie *ists*. Ferner werden u. a. Pausen (nach Längen gestaffelt: (.), (-) bis (---) oder mit Angabe der Dauer, z. B. (3.0)), Überlappungen (in eckige Klammern [] und untereinander geschrieben) und schnelle Anschlüsse (=) sowie para- und nonverbale Auffälligkeiten – z. B. ((*seufzt*, 5 Sek.)) unter Hinzufügung der Dauer, interpretierende Kommentare (z. B. <<*empört*> [inclusive Angabe der Reichweite] >) und die Tonhöhenbewegung am Ende der jeweiligen Einheit (vor dem Sprecherwechsel) – festgehalten. Generell erfolgt Kleinschreibung (Großschreibung nur zur Betonung). Transkriptzeilen werden nummeriert. Nach der Zeilennummer wird die Sprecherkennung notiert. Eine aufwendige phonetische Umschrift (z. B. IPA; vgl. Kap. VII.2.1 Die Lautschrift) ist dagegen nur notwendig, wenn entsprechende (phonetische) Analysekriterien berücksichtigt werden sollen (z. B. bei der Untersuchung von Dialekten).²⁰

Das *Leibniz-Institut für deutsche Sprache* (IDS) in Mannheim hat speziell für Audiodateien im Rahmen des Projekts „Forschungs- und Lehrkorpus Gesprochenes Deutsch (FOLK)“ den Transkriptionseditor FOLKER entwickelt,²¹ der das Transkribieren nach GAT 2 unterstützt. Wichtig ist zu beachten, dass es sich bei dieser Software nicht um einen Spracherkenner handelt. Die Transkription muss also weiterhin händisch eingegeben werden. Eine Weiterentwicklung ist in der besseren Organisation von akustischer Datei einerseits und Transkriptbereich andererseits zu sehen. Alle Vorteile der Transkription mit FOLKER fassen Selting et al. zusammen.²² Beispiele zur Transkription von Hörfunkspots aus dem *Regensburger Archiv für Werbeforschung* mittels FOLKER finden sich bei Ulrike Kaunzner.²³

Im vorliegenden Beitrag gehen wir in zwei Stufen vor. Für eine kleine Auswahl an Hörfunkspots erfolgt zunächst eine Vorbewertung auf einer einfachen Skala, die eine erste Abschätzung der Eignung der Spracherkennung oder vergleichbarer Dienstleistungen darstellt. Im Anschluss wird eine detailliertere Analyse der Umsetzungsleistung für einen ausgewählten Spot (Fallbeispiel) vorgenommen.

20 Kessel & Reimann (2017, S. 243–244).

21 http://agd.ids-mannheim.de/FOLK_extern.shtml.

22 Vgl. Selting et al. (2009, S. 401).

23 Vgl. Kaunzner (2021).

AUSWAHL DER BETRACHTETEN DIENSTE

Neben einer seit langem bekannten Anwendungssoftware, die auch als Marktführer gilt, *Dragon Dictate* der amerikanischen Firma *Nuance*²⁴, wurden zwei Online-Dienste ausgewählt, die direkt über den Browser verfügbar gemacht werden können.

Dabei handelt es sich zum einen um die Plattform *convertspeech.com*, die für sehr viele Sprachen eine Spracherkennung anbietet, unter anderem auch für das Deutsche. Sie kann in kleineren Umfängen kostenfrei genutzt werden, bei intensiverer Nutzung stehen Volumen- bzw. Abonnementmodelle mit entsprechenden Kostenätzen zur Verfügung. Hinzugenommen wurde zum anderen auch das Dienstleistungsangebot der Plattform *abtipper.de*, das von sich behauptet, im deutschsprachigen Bereich führend zu sein.²⁵ Die Besonderheit dieser Plattform ist das differenzierte Angebot: Neben einer kostengünstigen automatischen Spracherkennung („KI-basiert“) gibt es Dienste, bei denen die eigentliche Transkription und gegebenenfalls ein anschließendes Lektorat von Fachkräften durchgeführt wird. Die vollautomatische Transkription wird dabei explizit für größere Datenmengen angeboten: „Hochvolumige Datenmengen, z. B. Digitalisierung von Archiven“.²⁶ Das Ziel war, sowohl hinsichtlich des tatsächlichen Leistungsspektrums als auch hinsichtlich der Angebotstypen (zu lizenzierende Software versus Online-Angebot versus Umsetzung durch Fachkräfte) typische Dienste abzudecken. Die Kosten bewegen sich zwischen 0,29€ pro Audiomminute für die vollautomatische Umsetzung bis hin zu 1,99€ pro Audiomminute für Transkriptionen mit 4-Augen-Prinzip und erweitertem Lektorat, die insbesondere für sprachwissenschaftliche Forschung und Abschlussarbeiten gedacht sind.²⁷

Unbrauchbar – die Erkennungsleistung ist in einem Maße inkorrekt und/oder unvollständig, dass sie insgesamt unbrauchbar ist, da der Nachbearbeitungsaufwand vom Umfang her einer vollständigen manuellen Transkription gleichkommt oder diese gar übersteigt.

Größerer Nachbearbeitungsaufwand erforderlich – hier liegt zwar eine in weiten Teilen korrekte Umsetzung bzw. Texterkennung vor, weitere Nachbearbeitung ist aber durchgängig erforderlich.

Geringer Nachbearbeitungsbedarf – hier ist die Texterkennung annähernd vollständig und fehlerfrei, in gewissem Umfang ist eine Nachbearbeitung nötig.

Keine Nachbearbeitung erforderlich – in diesem Fall ist die Umsetzung korrekt und vollständig; Nachbearbeitung ergibt sich hier (wie auch in anderen Fällen) aus weiteren Repräsentationsanforderungen oder Erkenntnisinteressen, die sich mit den Funktionen der Texterkennung alleine nicht abdecken lassen.

24 *Nuance Dragon Professional Individual*, Version 2015, die derzeit aktuelle Fassung, vgl. https://shop.nuance.de/store/nuanceeu/de_DE/Content/pbPage.dragon-professional-individual – Zugriff: 26.4.2022.

25 „Der Marktführer für Transkriptionen“, <https://www.abtipper.de> – Zugriff: 26.4.2022.

26 <https://www.abtipper.de/leistungen-preise/> – Zugriff: 26.4.2022.

27 Dabei handelt es sich um Nettopreise, Stand April 2022.

Transkription mit				
Spot	<i>Dragon Dictate</i>	<i>Convertspeech</i>	<i>abtipper.de</i> automatisch	<i>abtipper.de</i> von Hand/Lektorat
Blendax (1957a)	Darf ich mich vorstellen. Ich bin Karies Kariespezialist für Zahnverfall. Bitte machen Sie mal den Mund auf und lassen Sie mich hinein sehen aber günstig sehr günstig, da kann ich jetzt sehr schnell bei Patienten zerstören vorausgesetzt, dass ihre Zähne niemals putzen mit Blendax, da enthält jede Tube das Kampfmittel gegen Karies, was an die Enzym BX, das an die Enzym Basics gegen Karies	Darf ich mich vorstellen Ich bin Karies? Karies Spezialist für zo Fall Bitte machen Sie mal den Mund auf und lassen Sie mich hineinschauen A günstig, sehr günstig da gar nicht jetzt sehr schnell, weil ein paar Zähne zerstören, vorausgesetzt, dass ihre Zähne niemals putzen. Mit Blaine Dax Da enthält jede Tube das Kampfmittel gegen Karies, das an die Enzym BX, das Anti Enzym BX gegen Karies.	I: [00:00:02] Hey, darf ich mich vorstellen? Ich bin Carries Carries Spezialist für Zahn Zerfall. Bitte machen Sie mal den Mund auf und lassen Sie mich hineinschauen. Ah, günstig. Sehr günstig, ne? Da kann ich jetzt sehr schnell mal ein paar Zähne zerstören. Vorausgesetzt, dass Sie Ihre Zähne niemals putzen. Mit. B: [00:00:23] Blender Wachs. Da enthält jede Tube das Kampfmittel gegen Karies, das Anti Enzym BX, das Anti Enzym BX gegen Karies.	B: Darf ich mich vorstellen? Ich bin Karies. Karies, Spezialist für Zahnerfall. Bitte machen Sie mal den Mund auf und lassen Sie mich hineinschauen. Ah, günstig. (lacht) Sehr günstig. (lacht) Da kann ich jetzt sehr schnell mal ein paar Zähne zerstören. Vorausgesetzt, dass Ihre Zähne niemals putzen mit . #00:00:22# B2: Blendax. Da enthält jede Tube das Kampfmittel gegen Karies. Das Antienzym BX. Das Antienzym BX gegen Karies.
Bernbacher²⁸	Die besten die wichtigsten Nazi bestenfalls nah wichtigsten die besten die besten Bernbach auf Deutschlands besten deutschen und internationalen Mode Hartweizennudeln mit bis	die besten die bissigsten nahe, die besten sonst. Na die bissigsten, die besten, na gut, die Besten so ist das bis ich machen sie ja schon immer aber jetzt ist Bernbacher auch Deutschlands beste Nudel. Das hat den Test der Zeitschrift Feinschmecker unter 36 deutschen und internationalen Nudeln bernbacher Hart, weil es nur den mit Biss.	I: [00:00:00] Die Besten. Die Basics. Na, die besten Songs. Na, die bis Kingston. Die Besten. Na gut, die Besten. B: [00:00:07] Tja, so ist es, bis ich. Machen Sie ja schon immer. Aber jetzt ist Birnbacher auch Deutschlands beste Nudel. Das sagt der Test der Zeitschrift Feinschmecker unter 36 deutschen und internationalen Nudeln. Birnbacher hart. Weizen Nudeln mit bis.	B1: Die Besten. #00:00:01# B2: Die Bissigsten. #00:00:02# B1: Nein, die Besten sind es. #00:00:03# B2: Nein, die Bissigsten. #00:00:05# B1: Die Besten. #00:00:06# B2: Na gut, die Besten. #00:00:07# B3: Tja, so ist es. Bissig waren sie ja schon immer. Aber jetzt ist Bernbacher auch Deutschlands beste Nudel. Das sagt ein Test der Zeitschrift Feinschmecker unter 36 deutschen und internationalen Nudeln. Bernbacher. Hartweizennudeln mit Biss.
Flensburger	Moin Moin Moin Folie Prinzipien Flensburger Pilsener	Moin, Moin, heute was wollt ihr jetzt? Glänzt Lenz Pletz Pletz klären. Es gibt hier nichts was, es gibt es hier nicht. Flensburger Pilsen hat das Letzt.	Name der Datei: Flensburger_2 I: [00:00:03] Ein Moin, moin, moin. Nä, das wollt ihr? Viets Flint Clans gib's hier nich, was das heißt? Pläne gibts hier nicht. Tschüss. Tschüss. Tschüss. Tschüss. Tschüss. Tschüss. Flensburger Pilsener. Das lernst.	B1: Moin. #00:00:03# B2: Moin. #00:00:03# B3: Moin. #00:00:04# B4: Moin. #00:00:04# B5: Moin. #00:00:05# B6: Na, was wollt ihr? #00:00:05# B1: Flens. #00:00:06# B2: Flens. #00:00:06# B3: Flens. #00:00:07# B4: Flens. #00:00:08# B5: Flens. #00:00:09# B6: Flens gibt es hier nicht. #00:00:10# B1: Was? #00:00:10# B2: Was? #00:00:10#

28 Die Spots der Marken *Bernbacher*, *Flensburger* und *Vodafone* wurden vor allem wegen der dialektalen Umsetzung, die mutmaßlich Schwierigkeiten bei der automatischen Verschriftung verursachen dürfte, ausgewählt. Die Spots sind auf die 2000er Jahre zu datieren (vgl. Reimann 2008a und 2008b; Reimann 2009). Sie liegen beispielsweise auf Hörfunkkassetten und CDs vor.

				<p>B3: Was? #00:00:11# B4: Was? #00:00:11# B5: Was? #00:00:12# B6: Flens gibt es hier nicht. #00:00:13# B1: Tschüss. #00:00:13# B2: Tschüss. #00:00:14# B3: Tschüss. #00:00:14# B4: Tschüss. #00:00:15# B5: Tschüss. #00:00:15# B6: Tschüss. (Plopp) #00:00:17# B7: Flensburger Pilsener. (Plopp) Das flenst.</p>
Vodafone	<p>Ich sollte später Foto, dass Großwater Konrad nipponischer in Wales liquidiert von Vodafone verbindet mobile Menschen weltweit. In Deutschland ist dabei</p>	<p>so das später davor doch das Sagen ob der ließ geschenk zwar davon über unser der Konzentration der polnische in Gewalt ist nix für dir, der Euro wohl davon ist auch das Auge. Der Hund ist macht Herr C fix, wo immer sie sind, wurde von verbindet mobile Menschen weltweit in Deutschland ist der zwei live dabei.</p>	<p>I: [00:00:00] Ich sog das Boot davon, zog das Zahngold. Ob Stalins Geschenk wohl davon gekommen seid? Der Kunst Rache. Gegengewalt ist nichts für dir. Der oder wo davon ist doktor saugut die und die Sportamt hörte fix.</p> <p>I: [00:00:12] Wo immer sie sind. Vodafone verbindet mobile Menschen weltweit. In Deutschland ist D2 live dabei.</p>	<p>B1: Ich sag dir, Spezi, Vodafone sagt er, saugut. Habe es der Lies geschenkt. Vodafone. Ich ... #00:00:04# -da kann sie ratschen. ... #00:00:06# - wäre das nichts für dich und deine Alte? Vodafone. Ich sage dir, saugut, die wo das gemacht haben. Ja, Zefix. #00:00:11# B2: Wo immer Sie sind, Vodafone verbindet mobile Menschen weltweit. In Deutschland ist D zwei live dabei.</p>

ERSTE BEWERTUNG DER UMSETZUNGSERGEBNISSE

	Transkription mit			
Spot	<i>Dragon Dictate</i>	<i>Convertspeech</i>	<i>abtipper.de automatisch</i>	<i>abtipper.de von Hand/Lektorat</i>
Blendax (1957a)	Geringer Nachbearbeitungsbedarf	Geringer Nachbearbeitungsbedarf	Geringer Nachbearbeitungsbedarf	Keine Nachbearbeitung erforderlich
Bernbacher	Unbrauchbar	Größerer Nachbearbeitungsaufwand erforderlich	Größerer Nachbearbeitungsaufwand erforderlich	Geringer Nachbearbeitungsbedarf
Flensburger	Unbrauchbar	Unbrauchbar	Größerer Nachbearbeitungsaufwand erforderlich / nahezu unbrauchbar	Keine Nachbearbeitung erforderlich
Vodafone	Unbrauchbar	Unbrauchbar	Unbrauchbar	Keine Nachbearbeitung erforderlich

Probleme in der Verschriftung bereiten dialektale und umgangssprachliche Äußerungen (*Bernbacher*, *Flensburger*, *Vodafone*). Der ‚alte‘ *Blendax*-Spot (1957a) dagegen lässt sich gut mit den Spracherkennern transkribieren. Das liegt an der standardnahen und tendenziell deutlichen Aussprache und dem vergleichsweise geringen Sprechtempo in den Spots der frühen Zeit.²⁹

29 Siehe dazu auch Kaunzner (2021, S. 92–93).

AUTOMATISIERTE TRANSKRIPTIONEN IM LINGUISTISCHEN VERGLEICH: DAS FALLBEISPIEL *BLENDAX*

Das folgende Experiment beinhaltet die Transkription einer Audio-Datei mit dem Titel „Meine Damen und Herren – Ich habe eine sehr wichtige Nachricht“ der Marke *Blendax* aus dem *Regensburger Archiv für Werbeforschung*.³⁰ Zunächst wird die händische Transkription von Kathrin Unverdorben gezeigt,³¹ dann folgen die Verschriftungen der Spracherkennung *Dragon Dictate*, *Convertspeech* und *abtipper.de*. Es kann vorab festgehalten werden, dass die automatischen Spracherkennung zu diesem Spotbeispiel nicht brauchbar sind. Wo die Probleme liegen, wird folgend detailliert dargestellt.

HÄNDISCHE TRANSKRIPTION (UNVERDORBEN)

ZZ	Sprecher1 (Karies)	Sprecher 2
1	Meine Damen und Herren – Ich habe eine <u>sehr</u> wichtige Nachricht –	
2	Wichtig für Ihren <u>Mund</u> und Ihre <u>Zähne</u>	
3	Selbstverständlich wird <u>Sie</u> das <u>besonders</u> interessieren	
4	Für eine wirklich <u>gute</u> Pflege ihres <u>Mundes</u> und ihrer <u>Zähne</u> brauchen Sie wohl das <u>Allerbeste!</u>	
5	Das heißt eine <u>äußerst</u> wirksame Zahnpasta	
6	/ <u>Natürlich</u> nehmen Sie <u>Blendax</u> denn <u>Blendax</u> ist ja so <u>gut</u> und <u>preiswert</u>	
7		Hören Sie doch nicht auf ihn!
8		<u>Ich</u> bin <u>Karies</u> he he mich kennen Sie doch <u>Spezialist</u> für <u>Zahnverfall</u> he he
9		Ich will von <u>Blendax</u> nichts <u>wissen</u> / weil dieses <u>Blendax</u> für <u>mich</u> das <u>Ende</u> bedeutet und /
10		Äh was ist denn das? Aach: <u>Blendax!</u> – Hilfe!
11	Naaa also –	

30 Blendax-Werke (1957b). Entgegen der Einordnung von Kathrin Unverdorben (2008) als Hörfunkspot sei an dieser Stelle lediglich von Audio-Datei gesprochen. Im RAW wird sie der Kategorie ‚Film‘ zugeordnet.

31 Unverdorben (2008, S. 149–150). Unverdorben setzt ebenfalls eine Kommentarspalte in der Tradition der Regensburger Werbelinguistik-Forschung für ihre Transkriptionen von *Blendax*-Werbespots der 1950er bis 1980er Jahre ein, die allerdings aus Gründen der Vergleichbarkeit hier nicht gezeigt wird. Zudem kennzeichnet sie paraverbale Komponenten. Einige Hinweise zu ihrer Verschriftung seien hier genannt: „Zudem werden im gesprochenen Text Auffälligkeiten gekennzeichnet, um schon beim ersten Lesen einen nahezu originalgetreuen Eindruck des Gesprochenen zu vermitteln: Besondere Betonungen werden unterstrichen, Atempausen werden mit ‚Slash‘ gekennzeichnet (/), stille Pausen mit einem waagrechten Strich (–), Ausrufe mit einem Ausrufezeichen (!), Fragen mit einem Fragezeichen (?), und Aufzählungen werden mit einem Doppelpunkt (:) verdeutlicht.“ (Ebd., S. 106–107).

12	/Übrigens: Sie kennen doch alle die normale Blendaxtube für <u>fünfzich</u> Pfennich und die <u>große</u> für <u>fünfundsiebzich</u> Pfennich	
13	Und jetzt: Jetzt gibt es die neue Blendax-Doppeltube	
14	Und – <u>die</u> hat ihre Vorteile /	
15	Sehn Sie mal: / die normale Tube kostet – <u>fünfzich</u> Pfennich – die <u>große</u> Tube <u>fünfundsiebzich</u> Pfennich / und die <u>neue</u> Blendax- <u>Doppeltube</u> kostet nicht etwa eine Mark <u>fünfundzwanzich</u> sondern nur – eine Mark!	
16	Und noch ein kleiner Vorteil: Alle Blendaxtuben haben den neuen handlichen <u>Tubenverschluss</u> / den – kann man einfach nicht verlieren	
17	Aber das Wichtigste ist und bleibt: Blendax enthält BX – BX gegen Karies	
18	Blendax macht die Zähne <u>blendend</u> weiß	
19	Blendax hält die Zähne und den Mund – <u>gesund</u> !	

Die nachfolgende Tabelle zeigt die drei automatisch erzeugten Transkriptionen dieses Spots:

Spot	Dragon Dictate	Convertspeech	Abtipper automatisch
Blendax (1957b)	Meine Damen und Herren, ich habe eine sehr wichtige Nachricht wichtig für ihren Mund und ihre Zähne. Selbstverständlich wird sie das besonders interessieren, wie eine wirklich gute Pflege ihres Mundes und ihrer Zähne brauchen Sie wohl das allerbeste, d. h. eine äußerst wirksame Zahnpasta. Natürlich nehmen sie Blendax, wenn Blendax ist ja so gut und preiswert wären, sieht auch nicht auf ihn kennen sich auch Spezialist jetzt ich will wann Blendax nichts wissen also. Übrigens, sie kennen doch alle, die normale Blendbacktube für 0,50 DM und die große für 0,75 DM und jetzt, jetzt gibt es die neue Blendax Doppeltube und die hat ihre Vorteile sehen Sie mal die normale Tube kostet 0,50 DM. Die große Tube 0,75 DM und die neue Pentax Doppeltube kostet nicht etwa 1,25 DM, sondern nur eine Mark und noch ein kleiner Vorteil. Alle Blendwachstum haben den neuen handlichen Tubenverschluss, den kann man einfach nicht verlieren. Aber das wichtigste ist und bleibt Blendax enthält BX BX gegen Karies Blendax macht die Zähne blendend weiß. Blendax hält die Zähne und den Mund gesund	meine Damen und Herren, ich habe eine sehr wichtige Nachricht Wichtig für ihren Mund und ihre Zähne. Selbstverständlich wird sie das besonders interessieren. Für eine wirklich gute Pflege ihres Mundes und ihrer Zähne brauchen sie wohl das Allerbeste das heißt eine äußerst wirksame Zahnpasta. Natürlich nehmen sie Blender ax, denn Blendax ist ja so gut und preiswert, hören Sie doch nicht auf ihn ich bin Karies, mich kennen sie doch spezialist wird seinen Verfall ich will von Blendax nichts wissen, weil dieses Blendax für mich das Ende bedeutet. Und was ist denn das? Ach blinder über? Na also, übrigens Sie kennen doch alle die normale Blendax Tube für 50 Pfennig und die große für 75, wenn ich und jetzt jetzt gibt es die neue Blendax doppelt Tube und die hat ihre Vorteile Sehen Sie mal die normale Tube kostet 50 Pfennig, die große Tube 75 Pfennig und die neue Blendax doppelt Tube kostet nicht etwa eine Mark 25, sondern nur eine Mark und noch ein kleiner Vorteil Alle Blendax Tuben haben den neuen, handlichen Tuben Verschluss. Den kann man einfach nicht verlieren, aber das Wichtigste ist und bleibt Blendax enthält WX BX gegen Karies, Blendax macht die Zähne blendend, weiß Blendax hält die Zähne und den Mund gesund	I: [00:00:01] Meine Damen und Herren, ich habe eine sehr wichtige Nachricht. Wichtig für Ihren Mund und Ihre Zähne. Selbstverständlich wird Sie das besonders interessieren. Für eine wirklich gute Pflege Ihres Mundes und Ihrer Zähne brauchen Sie wohl das Allerbeste. Das heißt eine äußerst wirksame Zahnpasta. Natürlich. Nehmen Sie Blendung. Denn Blendung ist ja so gut und preiswert. B: [00:00:24] Hören Sie doch nicht auf ihn. Ich bin Karies. Mich kennen Sie doch. Spezialist für Zahn Verfall. Ich will von Blender X nichts wissen, weil dieses Blender X für mich das Ende bedeutet. Und was ist denn das? Ach, Blender. Hilfe! I: [00:00:41] Na also. Übrigens, Sie kennen doch alle die normale Blendax Tube für 50 Pfennig und die große für 75 Pfennig. Und jetzt? Jetzt gibt es die neue Blender X Doppel Tube und die hat ihre Vorteile. Sehen Sie mal, die normale Tube kostet 50 Pfennig, die große Tube 75 Pfennig. Und die neue Pentax Doppel Tube kostet nicht etwa eine Mark 25, sondern nur eine Mark. Und noch ein kleiner Vorteil Alle Blender X Tuben haben den neuen handlichen Tuben Verschluss. Den kann man einfach nicht verlieren. Aber das Wichtigste ist und bleibt Blendung. S enthält BX, bx gegen Karies Blendung. Es macht die Zähne blendend weiß. Blendung. Es hält die Zähne und den Mund gesund.
Bewertung	Größerer Nachbearbeitungsaufwand erforderlich	Größerer Nachbearbeitungsaufwand erforderlich	Größerer Nachbearbeitungsaufwand erforderlich

STRUKTUR

Besonders auffallend ist, dass *Dragon Dictate* und *Convertspeech* den Text wie einen Monolog wiedergeben. Verschiedene Sprecheranteile werden also nicht unterschieden. Das ist besonders problematisch, da *Convertspeech* sogar Sätze konstruiert, die der eine Sprecher beginnt und vom anderen dann weitergeführt werden: *Natürlich nehmen sie Blender ax, denn Blendax ist ja so gut und preiswert, hören Sie doch nicht auf ihn ich bin Karies, [...]*. Der Imperativ (*hören Sie doch nicht...*) ist bereits Sprecher 2 zuzuordnen. *abtipper.de* und Unverdorben setzen je zwei Sprechende an: Sprecher 1 – Sprecher 2 – Sprecher 1 bzw. I – B – I. Zudem werden die verschiedenen Sprecher bei Unverdorben auch graphisch durch die Spaltenschreibung – jeder Sprechende bekommt eine eigene Spalte – hervorgehoben. Eine dritte Spalte – „Kommentar“ und siehe die Ausführungen oben – wird hier weggelassen, weil es folgend nicht um subjektive Einschätzung und Interpretation gehen kann. Solche Informationen sind von den Spracherkennern ohnehin nicht zu erwarten, so dass die Grundlage für einen Vergleich dahingehend fehlt.

INTERPUNKTION UND SYNTAX

Alle vier Transkriptionen weisen Interpunktion auf, und zwar auch in uneindeutigen Fällen. Bei Unverdorben ist zu erwähnen, dass auch die Zeilenzählung Interpunktion ‚ersetzt‘: Am Übergang von einer zur nächsten Zeile verzichtet sie auf Interpunktion. Die Zeilenzählung hier ist allerdings nicht in allen Fällen nachvollziehbar und wird vage erklärt: „nach dem Zusammenhang des Gesprochenen und nach sinnvollen Abschnitten im Aufbau des Werbespots“.³² Naheliegender ist das Setzen eines Punkts nach einem vollständigen Satz, z. B. *Selbstverständlich wird sie das besonders interessieren.* (*Convertspeech, abtipper.de*). Unverdorben setzt hier gar keine Interpunktion. Eindeutige Beispiele sind auch syntaktisch ‚echte‘ Fragesätze wie *Und was ist denn das?* (*Convertspeech, abtipper.de*); bei Unverdorben wird das Beispiel so transkribiert: *Äh was ist denn das? abtipper.de* setzt ein weiteres Fragezeichen nach der Äußerung *Und jetzt?*, während Unverdorben keine Frage hört, sondern eine Rechtsverweisung (Kataphorik) annimmt, vermutlich auch mit Spannungsaufbau und Auflösung im folgenden Satz. Den Doppelpunkt setzt sie übrigens fast durchgehend in dieser Bedeutung – rechtsverweisend – ein: *Übrigens: Sie kennen doch alle die normale Blendaxtube* (ZZ 12), *Und jetzt: Jetzt gibt es die neue Blendax-Doppeltube* (ZZ 13), *Sehn Sie mal: / die normale Tube kostet* (ZZ 15), *Und noch ein kleiner Vorteil: Alle Blendaxtuben haben den neuen handlichen Tubenverschluss* (ZZ 16). Das Ausrufezeichen nach einem Imperativsatz ist syntaktisch erwartbar und wird von Unverdorben eingesetzt: *Hören Sie doch nicht auf ihn!*; *abtipper.de* verwendet an der Stelle lediglich einen Punkt. Auch das Komma vor einem mit Subjunktion eingeleiteten Nebensatz (*weil*) bei *abtipper.de* fördert die Lesbarkeit, ist syntaktisch zu erwarten und hat nichts mit ‚gewagter‘ Interpretation zu tun: *nichts wissen, weil dieses Blendax für mich das Ende bedeutet*. Unverdorben setzt dagegen – zu-

32 Unverdorben (2008, S. 106).

rückhaltend – nur ein Pausenzeichen (/). Dagegen sind andere Interpunktioneingriffe problematisch, weil sie unnötig weitgehende Interpretation anzeigen. Das wäre besonders bei der händischen Transkription von Unverdorben vermeidbar: Der Einsatz des Halbgeviertstrichs (Gedankenstrichs) beispielsweise ist aus syntaktischer Perspektive nicht klar. In einem Fall wird er vermutlich auch an der Stelle einer Pause verwendet: *die normale Tube kostet – fünfzig Pfennich. abtipper.de* und *Convertspeech* setzen Interpunktion sowohl bei syntaktisch vollständigen Sätzen, wie oben mit Beispielen erläutert, als auch bei unvollständigen Sätzen: *Meine Damen und Herren, ich habe eine sehr wichtige Nachricht. Wichtig für Ihren Mund und Ihre Zähne. (abtipper.de)*. Auffallend ist bei *Convertspeech*, dass mehrfach auf Interpunktion verzichtet wird, obwohl die nächste Äußerung mit Großschreibung beginnt: *meine Damen und Herren, ich habe eine sehr wichtige Nachricht Wichtig für ihren Mund und ihre Zähne, [...] und die neue Blendax doppelt Tube kostet nicht etwa eine Mark 25, sondern nur eine Mark und noch ein kleiner Vorteil Alle Blendax Tuben haben den neuen, handlichen Tuben Verschluss*. Das kommt aber auch bei *abtipper.de* vor: *Und noch ein kleiner Vorteil Alle Blender X Tuben haben den neuen handlichen Tuben Verschluss*. *Convertspeech* verwendet – in diesem Beispiel – das kleinste Spektrum an unterschiedlichen Interpunktionszeichen (Punkt, Komma, Fragezeichen).

MARKEN- UND PRODUKTNAMEN

Es ist klar, dass die händisch vorgenommene Transkription von Unverdorben den Marken-/Produktnamen korrekt wiedergibt. Hier gibt es bei den Spracherkennern – über das hier behandelte *Blendax*-Beispiel hinaus – teils große Probleme. Und es tritt anscheinend auch kein ‚Gewöhnungseffekt‘ auf. *abtipper.de* schafft eine korrekte Schreibung kein einziges Mal. Hier findet man die Schreibungen *Blendung*, *Blender X*, *Pentax*. *Convertspeech* schreibt den Namen *Blendax* (auch im Kompositum, z. B. *Blendax-Doppeltube*) erstaunlicherweise zehn Mal korrekt, nur ein einziges Beispiel passt nicht: *Blender ax*. Auch bei *Dragon Dictate* findet sich oft die richtige Schreibung.

SEMANTIK, LEXIK, WORTBILDUNG – ZUR BEDEUTUNG VON WÖRTERN UND SÄTZEN

Die Spracherkener zeigen vielfach Probleme bei der Wiedergabe inhaltlich sinnvoll erscheinender Sätze bzw. Äußerungen. Hier wird also deutlich, dass eine händische Nachbearbeitung alternativlos ist, z. B. *Ach blinder über? (Convertspeech)*, *Ach, Blender. Hilfe! (abtipper.de)*: Das händisch transkribierte Beispiel lautet: *Aach: Blendax! – Hiilfe!* (Unverdorben). Bei *Dragon Dictate* macht die gesamte Äußerung, die Sprecher 2 zuzuordnen ist, keinen Sinn und ist zudem unvollständig. Mehrdeutigkeit kann den Spracherkennern Schwierigkeiten bereiten: *Blendax macht die Zähne blendend, weiß Blendax hält die Zähne und den Mund gesund (Convertspeech)*. Bei diesem Beispiel tritt das interessante Phänomen der Homonymie (oder in anderen Fällen der Polysemie) auf; *weiß* wird im vorliegenden Fall als Verb verstanden und nicht in der intendierten Be-

deutung des Farbadjektivs *weiß*. Komposita (*Zahnverfall*, *Zahnzerfall*, *Doppeltube*, *Tubenverschluss*) werden von den Spracherkennern fast ausschließlich jeweils als Simplizia wiedergegeben, falls die einzelnen Komponenten überhaupt erkannt werden (*abtipper.de*: *Doppel Tube*, *Tuben Verschluss*, *Convertspeech: doppelt Tube*, *Tuben Verschluss*). *Dragon Dictate* kreiert unpassende Komposita, wie *Blendbackstube* (statt *Blendaxtube*), *Blendwachstum* (statt *Blendaxtuben*). Das Kurzwort *BX* wird von den Spracherkennern nicht korrekt wiedergegeben, was kein Wunder ist: Es soll eine Wortneubildung der Marke *Blendax* sein.³³

LAUTUNG UND SCHREIBUNG

Kathrin Unverdorben verwendet meist die Standardsprache für die Schreibung, weicht davon aber bei entsprechender Aussprache gelegentlich ab: *Zahncrem*, *Pfennich*, *wichtich* oder *Hiilfe*. Dieses Vorgehen findet sich bei *Convertspeech* und *abtipper.de* nicht. Interessant ist, dass *Dragon Dictate* als einziger Transkribent das Kurzwort *DM* (z. B. 0,75 DM) ansetzt und nicht *Pfennig* (oder ähnliche Möglichkeiten zur Wiedergabe der Aussprache), also die Lautung meist komplett außen vor lässt. Das passiert allerdings nicht konsequent, wie dieses Beispiel zeigt: *nicht etwa 1,25 DM, sondern nur eine Mark*. Zudem fallen bei *Dragon Dictate* ‚Verhörer‘ auf: *wenn* statt *denn* oder *wären* statt *hören*.

FAZIT

Abschließend kommen wir zur Ausgangsüberlegung zurück: Gezeigt werden konnte durch die stichprobenhaften Untersuchungen, dass wenigstens derzeit kein vollautomatisches Verfahren so gute Qualität liefert, dass eine für weitere linguistische Analysen erforderliche Transkriptionsqualität ohne Nachbearbeitung möglich ist. Zu fragen ist allerdings, ob die bisher erreichbare beste Qualität wenigstens eingeschränkte Analysen möglich macht. Zu erwarten wäre, dass bei einer Perspektive des *distant reading* Fehler im Einzelfall nicht ins Gewicht fallen. Lügen hinreichend viele Transkripte vor, so könnte man einfache quantitative Aspekte analysieren, selbst wenn Fehler noch enthalten sind. Im Kontext von Big Data wird davon gesprochen, dass *messy data* im Detail dann kein Problem sind, wenn große Datenmengen in ihrer Gesamtheit betrachtet werden.³⁴ Einfache sprachstatistische Betrachtungen (Umfang des sprachlichen Materials, Wortschatz ohne besondere Phänomene wie Markennamen etc., langfristige Entwicklungen) sind vorstellbar. Wo allerdings die Grenze zwischen einem akzeptablen Fehlerniveau und nicht gegebener Verwertbarkeit liegt, ist noch genauer zu prüfen. Letztlich überrascht es nicht, dass bei der extremen Heterogenität der Hörfunkspots des *Historischen Werbefunkarchivs* in formaler und sprachlicher Hinsicht selbst die besten automatischen Verfahren nicht fehlerfrei arbeiten.

33 Vgl. Greule & Reimann (2011, S. 239–240).

34 Vgl. zu dieser Diskussion Mayer-Schönberger & Cukier (2013, S. 32 ff.) und Schöch (2013).

Auf dem Weg zum *distant listening*? Eine vollständige Transkription der Bestände des Hörfunkarchivs und weiterer auditiver Bestände bleibt ein Ziel, das vielfältige Forschungsmöglichkeiten eröffnet. Geht man von einem Mengengerüst von derzeit bis zu etwa 100.000 Hörfunkspots im *Historischen Werbefunkarchiv* aus und betrachtet gleichzeitig die Kosten von bis zu zwei Euro pro Minute für eine professionelle und lektorierte Transkription, so gelangt man bei einer grob geschätzten mittleren Länge der Spots von 30 Sekunden in eine Größenordnung, bei der die Kosten für eine hochqualitative Transkription im Bereich von einem Euro pro Spot liegen, für das gesamte Korpus also im Bereich von 100.000 €. Dies ist als Kostenumfang für eine umfassende Transkription im Bereich des Möglichen. Eine vollständige automatische Verschriftlichung mit dem KI-Dienst bei *abtipper.de* wäre für das Gesamtkorpus nach dieser Überschlagsrechnung bereits für etwa 17.500 € möglich. Insofern wäre es interessant zu untersuchen und näher zu diskutieren, welche relevanten Erkenntnisinteressen mit einer nicht fehlerfreien Transkription bereits zu erreichen wären. Klar ist, dass viele interessante Fragen etwa im Bereich der Verwendung von Markennamen und Neologismen oder der Varietätenlinguistik so nicht beantwortet werden können, da hier typische Fehlerquellen der vollautomatischen Verschriftlichung liegen.

LITERATUR

- Arnold, Taylor & Tilton, Lauren (2019). Distant Viewing: Analyzing Large Visual Corpora. In: *Digital Scholarship in the Humanities*. 34, Supplement_1. S. i3–i16.
- Blendax-Werke (1957a). „Darf ich mich vorstellen?“. In: *Regensburger Archiv für Werbeforschung*. R-Nummer: 4530; Blendax vom 12.1.1957. [Werbung, Audio]. URL: <https://raw.uni-regensburg.de/details.php?r=4530> – Zugriff: 13.4.2022.
- Blendax-Werke (1957b). „Meine Damen und Herren – Ich habe eine sehr wichtige Nachricht“. In: *Regensburger Archiv für Werbeforschung*. R-Nummer: 8040; Blendax vom 17.8.1957. [Werbung, Audio bzw. Blendax-Film]. URL: <https://raw.uni-regensburg.de/details.php?r=8040> – Zugriff: 13.4.2022.
- Brinker, Klaus; Antos, Gerd; Heinemann, Wolfgang & Sager, Sven F. (Hg.) (2001). *Text- und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung*, Halbbd. 2. (= Handbücher zur Sprach- und Kommunikationswissenschaft, 16). Berlin/New York: De Gruyter.
- Burghardt, Manuel & Wolff, Christian (2016). Digital Humanities in Bewegung: Ansätze für die computergestützte Filmanalyse. In: *DHd 2016. Book of Abstracts*. Leipzig: Universität Leipzig. S. 108–112.
- Fogg, Terry & Wightman, Colin W. (2000). Improving Transcription of Qualitative Research Interviews with Speech Recognition Technology. In: *Proceedings of the Annual Meeting of the American Educational Research Association*, New Orleans, LA, 24. – 28. April 2000.
- Furui, Sadaoki (2010). History and Development of Speech Recognition. In: Cheng, Fang & Jokinen, Kristiina (Hg.). *Speech Technology*. New York: Springer. S. 1–18. URL: https://doi.org/10.1007/978-0-387-73819-2_1.

- Gerber, Gabriele (2006). *Das Historische Werbefunkarchiv – ein Digitalisierungsprojekt der Universitätsbibliothek Regensburg* (= Berliner Handreichungen zur Bibliotheks- und Informationswissenschaft, Band 185). Berlin: Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin.
- Gref, Michael; Köhler, Joachim & Leh, Almut (2018). Improved Transcription and Indexing of Oral History Interviews for Digital Humanities Research. In: *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 7. – 12. Mai 2018, Phoenix Seagaia Conference Center, Miyazaki, Japan. Paris: ELRA. S. 3124–3131.
- Greule, Albrecht (2008). Spot-Texte analysieren. Der Beitrag der Sprachwissenschaft zur Untersuchung der Hörfunkwerbung. In: Reimann, Sandra (Hg.). *Werbung hören. Beiträge zur interdisziplinären Erforschung der Werbung im Hörfunk* (= Medien: Forschung und Wissenschaft, 17). Münster: LIT. S. 89–96.
- Greule, Albrecht & Reimann, Sandra (2011). Von „Doppel- und Normaltuben“ und dem „Anti-Enzym BX“ – Wortbildung in der Hörfunkwerbung. In: Elsen, Hilke & Michel, Sascha (Hg.). *Wortbildung im Deutschen zwischen Sprachsystem und Sprachgebrauch. Perspektiven – Analysen – Anwendungen* (= Perspektiven Germanistischer Linguistik, 5). Stuttgart: ibidem. S. 225–263.
- Greule, Albrecht & Reimann, Sandra (2015). *Basiswissen Textgrammatik*. Tübingen: Gunter Narr Verlag (UTB).
- Have, Iben & Enevoldsen, Kenneth (2021). From Close Listening to Distant Listening: Developing Tools for Speech-Music Discrimination of Danish Music Radio. In: *DHQ: Digital Humanities Quarterly* 15, 1 (2021). URL: <http://www.digitalhumanities.org/dhq/vol/15/1/000522/000522.html> – Zugriff: 13.4.2022.
- Hirschmann, Hagen (2019). *Korpuslinguistik*. Stuttgart: Metzler.
- Hofmann, Bernhard (2008). Musik in der Hörfunkwerbung. In: Reimann, Sandra (Hg.). *Werbung hören. Beiträge zur interdisziplinären Erforschung der Werbung im Hörfunk* (= Medien: Forschung und Wissenschaft, 17). Münster: LIT. S. 37–58.
- Kaunzner, Ulrike A. (2008). Vom „Leiden“ der Pause am Lampenfieber. Sprechen im Hörfunk. In: Reimann, Sandra (Hg.). *Werbung hören. Beiträge zur interdisziplinären Erforschung der Werbung im Hörfunk* (= Medien: Forschung und Wissenschaft, 17). Münster: LIT. S. 75–88.
- Kaunzner, Ulrike A. (2021). *Die Stimme als Zeitzeugin. Werberhetorik im Hörfunk*. Tübingen: Gunter Narr Verlag.
- Kessel, Katja & Reimann, Sandra (2017). *Basiswissen Deutsche Gegenwartssprache*. 5. Auflage. Tübingen: Gunter Narr Verlag (UTB).
- Mayer-Schönberger, Viktor & Cukier, Kenneth (2013). *Big Data: A Revolution that will Transform how we Live, Work, and Think*. Boston: Houghton Mifflin Harcourt.
- Moretti, Franco (2013). *Distant Reading*. New York: Verso Books.
- Redder, Angelika (2001). Aufbau und Gestaltung von Transkriptionssystemen. In: Brinker, Klaus; Antos, Gerd; Heinemann, Wolfgang & Sager, Sven F. (Hg.). *Text- und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung*, Halbbd. 2 (= Handbücher zur Sprach- und Kommunikationswissenschaft, 16). Berlin/New York: De Gruyter. S. 1038–1058.
- Reimann, Sandra (2008a). *MEHRmedialität in der werblichen Kommunikation. Synchrone und diachrone Untersuchungen von Werbestrategien*. Tübingen: Gunter Narr Verlag.

- Reimann, Sandra (2008b): „Es gibt immer was zu tun“. Eine Krieriologie zur Bestimmung des Slogans im Hörfunk. In: Reimann, Sandra (Hg.). *Werbung hören. Beiträge zur interdisziplinären Erforschung der Werbung im Hörfunk* (= Medien: Forschung und Wissenschaft, 17). Münster: LIT. S. 167–187.
- Reimann, Sandra (2009). „Grüß Gott alle miteinander, hier spricht Luis Trenker“ – Dialekt in deutscher Hörfunkwerbung gestern und heute. In: Hochholzer, Rupert; Kanz, Ulrich; Wildfeuer, Alfred & Zehetner, Ludwig (Hg.). *Mundart und Medien. Beiträge zum 3. dialektologischen Symposium im Bayerischen Wald*, Walderbach, Mai 2008. (= Regensburger Dialektforum, 16). Regensburg: edition vulpes. S. 251–272.
- Sager, Sven F. (2001). Probleme der Transkription nonverbalen Verhaltens. In: Brinker, Klaus; Antos, Gerd; Heinemann, Wolfgang & Sager, Sven F. (Hg.). *Text- und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung*, Halbbd. 2 (= Handbücher zur Sprach- und Kommunikationswissenschaft, 16). Berlin/New York: De Gruyter. S. 1069–1085.
- Schöch, Christof (2013). Big? Smart? Clean? Messy? Data in the Humanities. In: *Journal of Digital Humanities*. 2, Nr. 3. URL: <https://hal.archives-ouvertes.fr/hal-00920254> – Zugriff: 26.4.2022. S. 2–13.
- Selting, Margret (2001). Probleme der Transkription verbalen und paraverbalen/prosodischen Verhaltens. In: Brinker, Klaus; Antos, Gerd; Heinemann, Wolfgang & Sager, Sven F. (Hg.). *Text- und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung*, Halbbd. 2. (= Handbücher zur Sprach- und Kommunikationswissenschaft, 16). Berlin/New York: De Gruyter. S. 1059–1068.
- Selting, Margret et al. (1998). Gesprächsanalytisches Transkriptionssystem (GAT). In: Grewendorf, Günther & von Stechow, Arnim (Hg.). *Linguistische Berichte*, 173. Opladen: Westdeutscher Verlag. S. 91–122.
- Selting, Margret et al. (2009). Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). In: *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion*, 10. S. 353–402. (<https://www.gespraechsforschung-ozs.de>).
- Stöckl, Hartmut (2007). Hörfunkwerbung – „Kino für das Ohr“. Medienspezifika, Kodeverknüpfungen und Textmuster einer vernachlässigten Werbeform. In: Roth, Kersten S. & Spitzmüller, Jürgen (Hg.). *Textdesign und Textwirkung in der massenmedialen Kommunikation*. Konstanz: UVK Verlagsgesellschaft mbH. S. 177–202.
- Unverdorben, Kathrin (2008). „Blendax hält die Zähne und den Mund – gesund“. Diachrone sprachwissenschaftliche Untersuchung zur Zahnpastawerbung im Hörfunk. In: Reimann, Sandra (Hg.). *Werbung hören. Beiträge zur interdisziplinären Erforschung der Werbung im Hörfunk* (= Medien: Forschung und Wissenschaft, 17). Münster: LIT. S. 101–160.

