https://doi.org/10.21248/idsopen.16.2026.67

Felicitas Kleber/Christoph Draxler/Jürgen Trouvain/Sven Grawunder

Rundfunknachrichten als linguistisch-phonetische Ressource für die Sprachwandelforschung

Ein halbautomatisiertes Verfahren und eine Demosprachdatenbank

Abstract Spoken language is subject to synchronous variation at the segmental and suprasegmental levels, which in turn can lead to diachronic change. This study presents a semi-automatically processed radio news corpus as a useful source for diachronic trend studies and provides an initial exploratory assessment of prosodic change in terms of speech rate. The corpus comprises radio news from 1956 to 2017, obtained from the public service broadcaster *Saarländischer Rundfunk*. An initial analysis revealed that during the 60-year period under investigation, the proportion of core news was reduced in favor of other news elements. Further signal-based and signal+text-based analyses of the core news items showed that speech rate increased over time, which was accompanied by fewer and shorter pauses. The articulation speed, on the other hand, did not change. Advantages and disadvantages of news corpora for prosodic change research are discussed.

Keywords radio news, prosodic change, speech and articulation rate, pauses

1. Einleitung

Sprachwandel findet auf allen linguistischen Ebenen geschriebener und gesprochener Sprache statt. Der vorliegende Beitrag thematisiert neue Untersuchungsmöglichkeiten im Bereich von Wandel gesprochener Sprache. Im Vergleich zu dessen schriftsprachlich basierter Rekonstruktion lassen sich rezente und aktuell stattfindende Wandelprozesse auf segmentaler (Einzellaute) aber auch suprasegmentaler (Prosodie) Ebene experimentalphonetisch erst seit dem Aufkommen von Tonträgern und dem Vorhandensein einer entsprechenden Datengrundlage untersuchen. Diese Datenlage sowie die technologischen Möglichkeiten ihrer Analyse haben sich in den vergangenen Jahrzehnten rapide verbessert. Das *ONZE Corpus* (Gordon/Maclagan/Hay 2007) oder das *Philadelphia Neighborhood Corpus* (Labov/Rosenfelder/Fruehwald 2013) sind Beispiele soziophonetischer Korpora, anhand derer sowohl Ursprung als auch Verbreitung komplexer Lautwandelprozesse (z. B. vokalische Kettenverschiebungen, diachrone Vokalnasalierung) signalphonetisch untersucht werden können (vgl. Kleber 2025). Nach nunmehr 100 Jahren vertonter deutscher Nachrichtensprache (Dussel 2022) bietet sich dieses Medium aufgrund von Verfügbarkeit (Rundfunkarchive) und Vergleichbarkeit (Lesesprache, relative Homogenität der Sprechenden) perspektivisch nicht nur für sprachtechnologische Untersuchungen (z. B. Hecht/Riedler/Backfried 2002) an, sondern auch für Sprachwandelstudien zum Deutschen (Bose/Grawunder/Schwarz 2018), z. B. in Form von *real-time* Trendstudien, in denen ältere und jüngere Daten verschiedener Personen vergleichbarer Altersgruppen verglichen werden (Bülow/Scheutz/Wallner 2019).

Die Grundlage für Studien dieser Art bieten Projekte wie die ARD-Nachrichtenarche (Grawunder 2011). In diesem Beitrag werden Vor- und Nachverarbeitung von Radionachrichten mittels der BAS WebServices (Kisler/Reichel/Schiel 2017) u. a. zur Analyse im EMU-SDMS-Format (Winkelmann/Harrington/Jänsch 2017) anhand einer Sprachdatenbank mit derzeit 49 Nachrichtensendungen vom Saarländischen Rundfunk aufgezeigt.

1.1 Nachrichtensprache

Seit etwa 100 Jahren – mit Aufkommen des Rundfunks als Massenmedium – werden Nachrichten vorgelesen. Sie stellen damit im Vergleich zu anderen Gattungen gesprochener Sprache wie beispielsweise Gedichtrezitationen, Predigten oder öffentlichen Reden ein relativ junges Genre dar.

Typische Merkmale von Radionachrichten sind die Monologform vorgelesener und vorbereiteter Sprache, eine recht komplexe Syntax mit korrekter Grammatik und das Fehlen bestimmter Satzmuster wie Fragen (vgl. Iivonen/Niemi/Paananen 1995). Neben diesen kompositorischen Differenzen unterscheidet sich Nachrichtensprache von anderen Formen gesprochener Sprache zudem in der Sprachproduktion. Normalerweise artikulieren Nachrichtensprechende sehr deutlich und verwenden einen „neutralen“ Tonfall. Ersteres dient einer besseren Verständlichkeit; letzteres soll eine neutrale Haltung der Sprechenden gegenüber den Inhalten zum Ausdruck bringen. Nachrichtensprache weist somit stilistische Merkmale auf, die in Kontrast zur spontanen Sprache in Gesprächen stehen.

Code	Beschreibung
COR	Kernnachricht, Meldung, klassische Nachricht
INT	Anmoderation/Einleitung
LOC	Ortsangabe oder Thema
OTO	O-Ton (Bericht oder Interviewausschnitt; oft mit Hintergrundgeräusch)
OUT	Abmoderation/Outro
PAC	Jingle/Senderkennung, akustische Trenner, Pausen- bzw. Zeitzeichen, akustische Verpackung
REP	Bericht (oft vorproduzierter Einspieler)
SPO	Sport
STO	Börsenmeldungen
TIM	Zeitansage
TOP	Themenübersicht
TRA	Verkehrsnachrichten
WET	Wetterbericht/-vorhersage/-interview

Tab. 1: Beschreibung der Nachrichtenelemente und ihre Kodierung (vgl. Schwiesau/Grawunder/Bose 2011)

Nachrichten, egal ob im Radio oder im Fernsehen vorgetragen, sind nicht nur eine Aneinanderreihung loser Nachrichtenmeldungen, sondern haben in der Regel eine festgelegte Struktur, die aus bestimmten Elementen wie Kernnachrichten, Berichten, Sportnachrichten, Wirtschaftsmeldungen, Wetterberichten und Verkehrsmeldungen besteht. Für eine Übersicht über die Elemente einer Radionachrichtensendung siehe Tabelle 1.

Die Mischung verschiedener textlicher Elemente erfordert eine Diskursstruktur. Die einzelnen Nachrichtenelemente erfüllen verschiedene Funktionen und unterscheiden sich entsprechend in prosodischen und anderen linguistischen Merkmalen. So ist z. B. die Ankündigung eines Ortes unmittelbar vor einer Kernnachricht im Vergleich zu dieser syntaktisch verkürzt und weist eine besondere Form fallender Intonation auf. Was den Unterschied in der Sprechgeschwindigkeit zwischen den Elementen betrifft, so fanden Grawunder/Kettel (2015) im Wesentlichen keinen Unterschied zwischen deutschen Nachrichtenlesungen von Kernnachrichten und Berichten (beide 4,3–4,4 Silben pro Sekunde, im Folgenden s/s), aber einen Rückgang auf 3,3 s/s für das dazwischenliegende überleitende Intro-Element. Insgesamt deuten diese Werte auf eine insgesamt langsamere Sprechgeschwindigkeit hin, insbesondere im Vergleich zu vorgelesener Sprache von nicht-professionellen Sprecherinnen und Sprechern (5,97 s/s in Pellegrino/Coupé/Marsico 2011; zur Variabilität von Sprechgeschwindigkeit siehe Kleber/Schmid im Dr.), aber auch den in Iivonen/Niemi/Paananen (1995) untersuchten Nachrichtensprechenden (5,8 s/s in).

Eine für das Genre möglicherweise generell langsamere Sprechgeschwindigkeit (siehe aber unten) stünde in Einklang mit der Beobachtung, dass der Sprechstil von Nachrichtensprechenden ein Paradebeispiel für *clear speech* ist, die sich eher durch Hyper- als durch Hypoartikulation (Lindblom 1990) auszeichnet. Insbesondere ältere Nachrichtensprechende haben – oft im Zusammenhang mit einer Schauspielausbildung – ein Sprechtraining erhalten. Diese Form der Sprechausbildung ist zwar für Nachrichtensprechende jüngerer Nachrichtensendungen nicht mehr obligatorisch, aber auch heute noch dürfen beim Saarländischen Rundfunk (SR) nach Angaben des Senders nur Sprecherinnen und Sprecher Nachrichten vorlesen, deren Sprechfähigkeit für dieses Medium zuvor geprüft wurde und die eine Sprechfreigabe erhalten haben. Dabei liegt der Schwerpunkt auf Sprechfluss, Verständlichkeit, Klarheit der Sprachleistung und einer standardnahen Aussprache. Dies führt wiederum dazu, dass die Sprache von Nachrichtensprechenden als vorbildliche Standardaussprache gelten kann. Sowohl im Britischen Englisch (BBC-Aussprache, siehe z. B. Ladefoged 2001) als auch im Deutschen (Lameli 2004) dient die Aussprache von Nachrichtensprechenden als Referenz für die Beschreibung von Standardsprache. Im Deutschen finden sich aber auch dort regionalsprachliche Merkmale (Spiekermann 2007). Die Vorbildfunktion von Nachrichtensprechenden zeigt sich besonders an der oft engen Wechselbeziehung zwischen der in Aussprachewörterbüchern kodifizierten Standardsprache und deren lautliche Umsetzung durch Nachrichtensprechende (z. B. Krech et al. 2009). Ihre Aussprache spiegelt damit den jeweils aktuellen Stand der Standardaussprache wider und sollte bei Untersuchungen zu Lautwandelprozessen (siehe 1.2) dieser Varietät mit berücksichtigt werden.

Ungeachtet dessen sollte die Aussprache von Nachrichtensprechenden nicht mit der Aussprache von nicht-professionellen Sprecherinnen und Sprechern der Standardaussprache gleichgesetzt werden, denn Unterschiede zwischen den Gruppen sind für eine Reihe von Sprachen belegt. Mok/Fung/Li (2014) verglichen beispielsweise Kantonesisch bei profes-

sionellen Fernsehsprechenden im Vergleich zu nicht-professionellen Sprechern und Sprecherinnen. Ihre Ergebnisse zeigten, dass die kantonesischen Nachrichtensprechenden nicht nur einen größeren Tonhöhenbereich und eine höhere Variabilität der Silbenlänge aufwiesen als die nicht-professionellen Sprecher und Sprecherinnen, sondern im Gegensatz zu den oben stehenden Beobachtungen zum Deutschen auch deutlich schneller sprachen. Barbosa/Madureira/de Mareüil (2017) zeigten anhand von Daten in vier Sprachen (europäisches und brasilianisches Portugiesisch, Französisch, Deutsch), dass sich professionelle und nicht-professionelle Sprecher und Sprecherinnen beim Lesen von Nachrichten und anderen Genres in mehreren prosodischen Parametern unterscheiden, darunter die Grundfrequenz (f0) und die Intensität. In Bezug auf Pausen stellten Trouvain/Werner/Möbius (2020) bei deutschen Radionachrichtensprechenden fest, dass sie durchschnittlich 16 Pausen pro Minute Sprechzeit produzierten und damit weniger im Vergleich zu anderen Genres. 80% dieser Pausen wurden zudem mit einem hörbaren Einatmungsgeräusch realisiert. Dieser Wert liegt damit deutlich über jenen, die für andere Genres gemessen wurden. Auch die Pausendauer war bei Radionachrichten viel kürzer im Vergleich zu jenen in vorgelesenen Erzählungen oder Gesprächen von nicht-professionellen Sprechern und Sprecherinnen. Im Tschechischen wiederum ist zwar die Zahl der Pausen beim Vorlesen von Nachrichten niedriger im Vergleich zu anderen Genres (hier dem Rezitieren von Gedichten), die Dauer der Pausen unterschied sich jedoch nicht (Šturm/Volín 2023). Nachrichtensprache wie auch Sprachaufnahmen professioneller Sprecherinnen und Sprecher unterscheiden sich also in vielerlei Hinsicht. Dieser Aspekt muss in Untersuchungen wie der vorliegenden berücksichtigt werden, obwohl hier weder das Genre Nachrichtensprache noch der Vergleich zwischen professionellen und nicht-professionellen Sprecherinnen und Sprechern im Vordergrund stehen, sondern Nachrichtensprache aufgrund der Verfügbarkeit für linguistische Fragestellungen – hier im Bereich von Lautwandel – genutzt wird.

1.2 Prosodischer Wandel

Sprache als inhärent dynamisches System unterliegt grundsätzlich Wandel im Laufe der Zeit. Die entsprechenden Lautwandelprozesse können nicht nur im Nachhinein, wenn sie abgeschlossen sind, beschrieben werden, sondern auch, wenn sie aktuell stattfinden (vgl. Kleber 2016). Lautwandel wird in der Regel als Abweichung der Aussprache auf segmentaler Ebene über längere Zeiträume untersucht, die sich über wenige (vgl. Zellou/Tamminga 2014) bis zu vielen Generationen (vgl. Hahn/Siebenhaar 2019, S. 232) erstrecken kann. Er kann jedoch auch prosodische Merkmale beeinflussen (manchmal auch als prosodischer Wandel bezeichnet, vgl. Page 1999; Bose/Grawunder/Schwarze 2018). Neben dem Einfluss systeminterner Faktoren auf Wandelprozesse (z. B. durch Koartikulation, Dekakzentuierung; vgl. Kleber 2016), sind extern motivierte Veränderungen innerhalb eines gesprochenen Genres zu erwarten. Letztere ergeben sich aus einem dynamischen Umfeld, wie ein sich schnell ausbreitendendes Medium in politisch, wirtschaftlich, und sprachtechnologisch sich verändernden Zeiten. Die Aufgabe bleibt jedoch bestehen, Informationen in bestmöglicher Form von einem Sprechenden an viele Zuhörende zu übertragen. Prosodische Veränderungen im Laufe der Zeit sind für andere Rundfunkgenres wie Fußballreportagen dokumentiert (Trouvain 2015). Daher ist auch im Genre der Radionachrichten über einen Zeitraum von mehreren Jahrzehnten mit Veränderungen der prosodischen Merkmale zu rechnen.

De Mareüil/Rilliard/Allauzen (2012) stellten in einer Untersuchung audiovisueller Nachrichten aus Frankreich von 1945 bis 1997 fest, dass in dem ca. 50-jährigen Untersuchungszeitraum eine Reihe prosodischer Merkmale zurückgegangen sind, darunter die durchschnittliche f0, der Tonhöhenanstieg in Verbindung mit der Anfangsbetonung, die Vokaldauer als

Merkmal einer emphatischen Anfangsbetonung sowie die Dauer sogenannter *Inter-Pause Units* (nachfolgend IPUs), d. h. zeitlichen, sprachlich gefüllten Einheiten zwischen Pausen. Die Sprechgeschwindigkeit zeigte in dieser Studie jedoch keine Änderung über die Zeit. Rilliard et al. (2023) untersuchten französische Fernseh- und Radiodaten aus vier, 60 Jahre umspannenden Zeiträumen (1955–56, 1975–76, 1995–96 und 2015–16) und fanden eine Tendenz zu tieferen Stimmen bei den Sprechenden, unabhängig von deren Geschlecht.

Savino/Wehrle/Grice (2024) verglichen in einer *real-time* Untersuchung den prosodischen Stil italienischer Nachrichtensprecher aus den späten 1960er Jahren mit den Produktionen zweier Nachrichtensprechender, die 2005 dieselben Texte vorlasen. Ihre Ergebnisse zeigen, dass die f0 hinsichtlich des Anteils der Anstiege und Abfälle über die Zeit (*wiggliness*) sowie der räumlichen Auslenkung (f0-range, hier *spaciousness*) in beiden Epochen ähnlich waren. In den neueren Aufnahmen war jedoch die durchschnittliche f0 höher und die Sprechgeschwindigkeit schneller, was hauptsächlich auf die Verringerung der Anzahl und Dauer der Pausen zurückzuführen war, wodurch viel längere IPUs entstanden.

Diese Studien liefern sprachübergreifend eher heterogene Beobachtungen und erschweren somit eindeutige Vorhersagen, wie z. B. kürzere Pausen und längere IPUs in jüngerer Zeit. Auch lässt sich daraus nicht ableiten, dass Entwicklungen über einen Zeitraum von mehreren Jahrzenten linear verlaufen, z. B. durch kontinuierlich kürzere Pausen und kontinuierlich längere IPUs. Ergebnisse zur Vokalnasalierung im Amerikanischen Englisch belegen, dass Änderungen zwischen zwei Generationen sich in der dritten Generation wieder umkehren können (Zellou/Tamminga 2014). Auch in dieser *real-time* Studie stehen neben einer Evaluation der Nachrichtenelemente, Sprechgeschwindigkeit und Pausen im Vordergrund, deren Entwicklung über einen vergleichbaren Zeitraum wie in Rilliard et al. (2023) nun für das Deutsche explorativ untersucht werden soll.

2. Methode

Sowohl die Daten als auch alle im folgenden vorgestellten Tools zur Datenverarbeitung stehen für wissenschaftliche Forschungszwecke zur freien Verfügung. Bei der Nachnutzung der Nachrichtendaten können wir uns auf das deutsche Urheberrechtsgesetz (UrhG) berufen, da es sich nicht nur um öffentlich ausgestrahlte, sondern auch um öffentlich finanzierte Daten handelt. Insbesondere §60d UrhG erlaubt die automatische und systematische Vervielfältigung von Quellenmaterial zur Erstellung eines Korpus für wissenschaftliche Forschungszwecke (European Commission/Senftleben 2022; Margoni/Kretschmer 2022).

2.1 Daten

Für diese Pilotstudie wurden 43 (von insgesamt 49) Radionachrichten aus den Jahren 1956 bis 2017 von einem deutschen öffentlich-rechtlichen Radiosender, dem SR, analysiert.[1] Die ersten 31 der analysierten Aufnahmen aus den Jahren 1956 bis 2002 wurden uns vom Archiv des Senders zur Verfügung gestellt. Das Korpus wurde um weitere Aufnahmen aus

1 Sechs im Korpus enthaltene Radionachrichten blieben unberücksichtigt, da es sich um weitere Nachrichtensendungen aus bereits repräsentierten Jahren handelte und diese Jahre dann überrepräsentiert gewesen wären. Standen zwei Nachrichtensendungen pro Jahr zur Verfügung wurde nur die der Europawelle bzw. von SR1 in die Analyse einbezogen.

den Jahren 2003 bis 2017 aus der Sammlung „Nachrichtenarche“[2] (Grawunder 2011) ergänzt, von denen 12 in die Analyse einbezogen worden sind. In der Regel handelt es sich dabei um Nachrichtensendungen der Europawelle bzw. des Nachfolgesenders SR1. In wenigen Fällen um Sendungen der Saarlandwelle, dem Vorgängersender von SR3 und in einem Fall um eine ARD-Nachrichtensendung. Die Vergleichbarkeit der Daten wird dadurch gewährleistet, dass alle Sendungen im SR in Saarbrücken aufgezeichnet worden sind. Wie die Zahlen bereits verraten, stehen uns nicht aus allen Jahren Daten zur Verfügung, auch weil gerade in den früheren Jahrzehnten Radionachrichten nur selten als archivierungswürdig angesehen wurden (Schwiesau/Grawunder/Bose 2011). Im Korpus nicht vertreten sind die Jahre 1958–1960, 1962–1968, 1970–1971, 1977, 1982, 1987, 1989, 2005, und 2014–2015. Für alle anderen Jahre fließt jeweils eine Nachrichtensendung pro Jahr in die Analyse ein (vgl. Fußn. 1). Die verschiedenen Jahre werden durch unterschiedliche Nachrichtensprechende repräsentiert, so dass das Korpus *real-time* Trendstudien ermöglicht, in denen ältere und jüngere Daten verschiedener Individuen verglichen werden (Bülow/Scheutz/Wallner 2019). Die Identifikation der Sprechenden ist derzeit noch nicht abgeschlossen: 76% der Kernnachrichten wurden aber bislang 14 Sprechern und 7 Sprecherinnen zugeordnet. In wenige Sendungen kann dieselbe Person die Kernnachrichten vorgetragen haben; von anderen Sprechenden liegt hingegen nur eine Sendung vor.

2.2 Vorverarbeitung

Ein Ziel der aktuellen Studie war die Spezifikation eines hochautomatisierten und effizienten Arbeitsablaufs, um anschließende Analysen gesprochener Sprachkorpora vorzubereiten. Die Gesamtdauer aller 49 Nachrichtensendungen der Demosprachdatenbank beläuft sich auf ca. 3,5 Stunden, die der hier untersuchten Sendungen auf 3 Stunden und 8 Minuten. Die Größe des Gesamtkorpus beträgt 1,22 GB. In einem ersten vorverarbeitenden Schritt wurden

1. die Audiodateien in WAV-Format mit dem Kommandozeilen-Tool ffmpeg automatisch konvertiert (Mono, Abtastrate mindestens 44,1 kHz, 16 Bit lineare Quantisierung; siehe ffmpeg-Codec unter https://ffmpeg.org/, Stand: 10.12.2025),
2. die Dateinamen angeglichen,
3. orthografische Transkripte mit satzähnlichen Segmenten (Chunks) mithilfe automatischer Spracherkennung (unter Verwendung eines Python-Skriptes) erstellt.

Die Dateinamenangleichung ist nicht nur im Sinne eines übersichtlicheren Korpus, sondern ermöglicht auch eine automatische Verarbeitung über Webdienste (siehe 2.3.1). Die automatische Spracherkennung (ASR) wurde mit einer lokalen Installation von WhisperX (Bain et al. 2023, Version 3.1.1) und mit den in (1) angegebenen Parametern durchgeführt:

2 Die „Nachrichtenarche“ wurde 2003 ins Leben gerufen mit dem Ziel, das Medium Nachrichtensprache u. a. korpuslinguistisch nutzbar zu machen und einer fehlenden Einheitlichkeit bei der Archivierung gesprochener Nachrichten entgegenzuwirken. Einmal jährlich, am 11. November, werden daher seit 2003 die Aufnahmen der 13-Uhr-Nachrichten von allen öffentlich-rechtlichen Radiosendern in Deutschland gesammelt.

(1) whisperx --model large-v3 --lang de
--compute_type float32 --output_dir results *.wav

Dies erzeugte (unter anderem) zeitlich mit dem Sprachsignal alignierte Transkriptdateien im .srt-Format (d. h. einem gängigen Textformat für Untertitel). Die Überprüfung, Korrektur und Weiterverarbeitung dieser Transkripte erfolgte anschließend im Transkriptionseditor Octra (Pömp/Draxler 2017) in Verbindung mit Octra Backend (Draxler/Pömp 2024), einer web-basierten Infrastruktur für Transkriptionsprojekte. Der genaue Arbeitsablauf besteht aus den folgenden Schritten (siehe Draxler et al. 2025 für weitere Details):

4. Manuelle Korrektur des in 3. generierten Transkripts inkl. Zusammenfassen der Chunks in Nachrichteneinheiten (d. h. Löschung von Chunk-Grenzen, siehe Abb. 1);
5. Manuelle Etikettierung der einzelnen in 4. erzeugten Nachrichteneinheiten nach ihrem Elementtyp (vgl. Tab. 1);
6. Automatisches Schneiden der Audiodateien entsprechend der in 4. korrigierten Chunk-Segmentgrenzen.

Der 4. und 5. Schritt in o. g. Aufzählung erfolgte parallel und nicht konsekutiv. Jede geschnittene Nachrichteneinheit besteht aus einem einzigen Nachrichtenelement-Etikett in spitzen Klammern (z. B. <COR>), gefolgt vom orthografischen ASR-Transkript der Nachricht (vgl. Abb. 1).

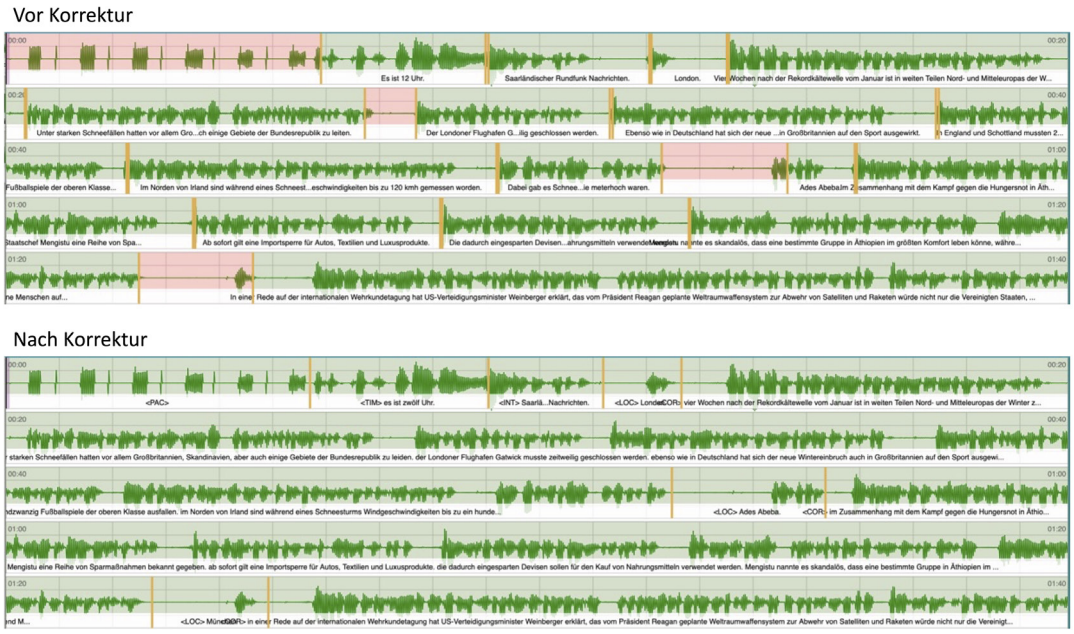

Abb. 1: Mithilfe ASR segmentierte und etikettierte Text- und Pausensegmente vor (oben) und nach (unten) der Korrektur in Octra (Pömp/Draxler 2017). Der Prüfprozess umfasst u. a. die Grenzkorrektur und die Etikettierung der Nachrichtenelemente (vgl. Tab. 1). Rot markiert sind Abschnitte, in denen die ASR aufgrund fehlender Spracherkennung kein Transkript geliefert hat. In der manuellen Korrektur wurden diese entweder mit <PAC> etikettiert (vgl. Tab. 1) oder – im Falle von Pausen – für deren in 2.3.1 beschriebene automatische Erkennung unetikettiert einem Snippet zugeordnet

Für die Berechnung der Wortfehlerrate (WER kurz für *Word Error Rate*), d. h. des Anteils von der ASR falsch erkannter Wörter, wurden die automatisch mit ASR generierten und die in Octra manuell korrigierten Transkripte normalisiert und anschließend mit der

Python-Bibliothek JIWER (Vaessen 2018) berechnet (siehe Draxler et al. 2025). Der WER lag bei 8,85% und war damit überraschend hoch für Sprachaufnahmen, die von professionellen Sprechern und Sprecherinnen unter sehr guten akustischen Bedingungen aufgenommen worden sind. Da die folgenden Arbeitsschritte auf den manuell wortkorrigierten Transkripten basieren, wirkt sich dies jedoch nicht auf die in 3. präsentierten Analysen aus.

Mit Blick auf die geplanten Analysen waren die Transkribierenden angewiesen, die rechte Grenze so nah wie möglich an das Ende des sichtbaren und hörbaren Sprachsignal zu setzen. Dies gewährleistet, dass Segmentgrenzen keine Pausen unterbrechen und somit eine zuverlässige Segmentierung (siehe 2.3.1) und Messung der Pausendauer (siehe 2.3.3; für Beispiele siehe Abb. 1 unten).

2.3 Annotation und Messung

2.3.1 Automatische Annotation

Das Schneiden der 43 analysierten Audiodateien in Octra resultierte in 832 Snippets mit je einem Nachrichtenelement. Mit einem Python-Skript wurde die BAS-Webservice-Pipeline ‚G2P → MAUS→ PHO2SYL'[3] (Kisler/Reichel/Schiel 2017) für jedes Dateipaar (d. h. wav-Datei plus die dazugehörige Textdatei mit orthographischer Transkription auf Phrasenebene) aufgerufen, um alle Snippets automatisch auf drei weiteren Ebenen in Wörter, Silben und Phoneme zu segmentieren und die Silben und Phoneme phonetisch zu etikettieren. Pausen werden in MAUS dabei auf allen drei Ebenen inklusive der Phonemebene segmentiert und durch das Etikett <p:> abgebildet. Das Output-Format entsprach dem EMU-Speech Database Management System (Winkelmann/Harrington/Jänsch 2017). Die 832 resultierenden Kleinstdatenbanken, bestehend aus je einer Audiodatei und der dazugehörigen Transkriptdatei im json-Format, wurden anschließend in einer einzigen EMU-Sprachdatenbank zusammengeführt (vgl. Abb. 2). Auf eine manuelle Korrektur der Segmentgrenzen und Etiketten wurde – im Gegensatz zur weniger zeitaufwendigen Korrektur der automatisch falsch erkannten Wörter (vgl. 2.2) – nicht nur aus Zeitgründen (vorerst) verzichtet: Diachrone Trends lassen sich anhand automatisch segmentierter Daten robust schätzen, da die Streuung innerhalb der Daten aufgrund fehlerhafter Segmentierungen größer ausfällt und Trends eher unter- als überschätzt werden (vgl. Kisler/Kleber 2019). Die Fehleranfälligkeit wurde zudem durch die zuvor erfolgte Wortkorrektur in Octra (vgl. 2.2) reduziert.[4]

3 G2P = Graphem zu Phonem, MAUS= Munich Automatic Segmentation, PHO2SYL = Silbentrennung basierend auf phonemischer Transkription.

4 Ein weiterer Vorteil der automatischen Segmentierung ist, dass diese nicht den natürlichen subjektiven Schwankungen bei einer manuellen Korrektur unterworfen ist. Im Gegensatz zu manuellen Eingriffen ist das Ergebnis automatischer Segmentierung replizierbar. Ein Vergleich zwischen automatischer und manueller Segmentierung ist ferner nur bedingt aussagekräftig, da es auch bei der manuellen Segmentierung keine eindeutig richtige Segmentierung gibt (siehe hierzu Kisler/Kleber 2019).

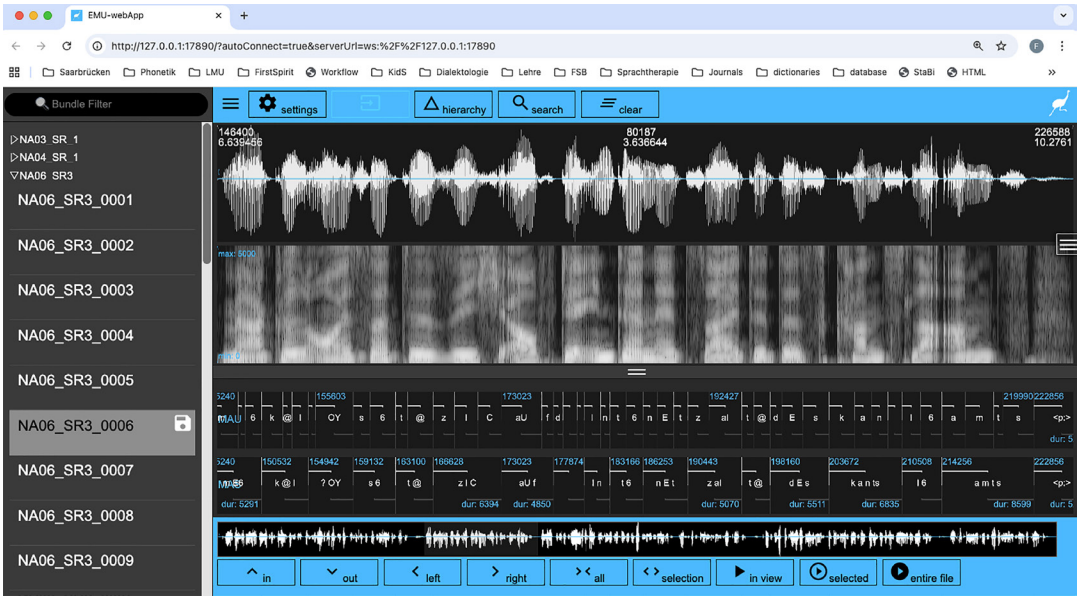

Abb. 2: Ausschnitt einer in der EMU-webApp geöffneten Kernnachricht (NA06_SR3_0006) einer Nachrichtensendung aus dem Jahr 2006 (NA06_SR3). Im Menü links sind weitere Nachrichtenelemente derselben Sendung sowie 2 weitere Nachrichtensendungen sichtbar. Unter dem Signalausschnitt sind die Phonem- und die Silbensegmentierung dargestellt

2.3.2 Signalbasierte Messung

Die Sprechgeschwindigkeit (Anzahl der Silben[5] pro Sprechzeit *inklusive* Pausen) und die Artikulationsgeschwindigkeit (Anzahl der Silben pro Artikulationszeit *exklusive* Pausen) sowie die Pausen pro Minute wurden zunächst auf der Grundlage des unsegmentierten akustischen Signals mit Hilfe des Praat-Skripts von de Jong/Wempe (2009)[6] gemessen. Das Skript erkennt Stimmhaftigkeit durch Periodizitätsanalyse und detektierter Intensitätsgipfel, denen Intensitätsminima vorausgehen, die als potenzielle Silbenkerne betrachtet werden. Das Skript löscht anschließend Peaks, die nicht stimmhaft sind. So werden manchmal Gipfel mit geringer Intensität übersehen, während Passagen mit hoher Intensität manchmal Doppelgipfel zugeordnet werden. Die Maße wurden mit linearen Regressionsmodellen in R statistisch ausgewertet (siehe 3.2 für Details).

2.3.3 MAUS-basierte Messung

Zur Validierung der rein signalbasierten Messung der Sprechgeschwindigkeit und des obigen Arbeitsablaufs wurden zusätzlich die Sprechgeschwindigkeit und die Pausen auf Grundlage einer Signal+Text-basierten MAUS-Segmentierung der Phoneme und Silben gemessen. MAUS segmentiert und etikettiert Sprachlaute auf Grundlage sowohl akustischer Merkmale, die aus dem Signal extrahiert werden, als auch lexikalischer Informationen und phonotaktischer Regeln, die aus Trainingsdaten gewonnen werden (vgl. Kisler/Reichel/Schiel 2017). Dadurch können phonologische Prozesse wie Assimilationen oder Elisionen, die u. a. zu den spontansprachlich häufigen Abweichungen von kanonischer Silbifizierung führen, Rechnung getragen werden. Mit MAUS segmentierte Daten bieten

5 Weder hier noch bei dem in 2.3.3 beschriebenen Ansatz entsprechen die Silben notwendigerweise den kanonischen, sondern den im Signal erkannten Silben. MAUS berücksichtigt die Wahrscheinlichkeit, mit der Silben realisiert werden und trägt Silbenelisionen Rechnung.

6 https://github.com/FieldDB/Praat-Scripts/blob/main/praat-script-syllable-nuclei-v2file.praat (Stand: 10.12.2025).

sich daher besonders für den Vergleich mit der in 2.3.2 beschriebenen Methode an. Zu diesem Zweck wurde die EMU-Sprachdatenbank in R unter Verwendung des Pakets emuR (Winkelmann/Harringont/Jänsch 2017) nach verschiedenen Labels (z. B. Nachrichtenelemente, Pausen, Silben) abgefragt. Zusätzlich zu den oben genannten Maßen wurde die Dauer der von MAUS segmentierten Pausen und die Dauer aller Segmente zwischen den Pausen, d. h. die IPUs, berechnet. Die Maße wurden mit linearen Regressionsmodellen in R geprüft (für Details siehe 3.2).

3. Ergebnisse

3.1 Verteilung der Nachrichtenelemente

Vor der Analyse der Sprechtempi und Pausen wurde das Korpus zunächst auf die Verteilung der Nachrichtenelemente hin explorativ-deskriptiv untersucht. Dabei stand einerseits der proportionale Anteil der Elemente *an* sowie deren Anordnung *innerhalb* einer Nachrichtensendung im Fokus. Nach der Sichtung aller Nachrichtensendungen und aller Nachrichtenelemente werden im Folgenden die Entwicklung der Kernnachrichten und Reportagen im Vergleich zu allen anderen Elementen anhand ausgewählter Jahre (= Sendungen) beschrieben. Letzeres dient der Übersichtlichkeit. Die Auswahl der Jahre erfolgte so, dass der Zeitraum breitestmöglich (1957–2017) und gleichmäßig (MW = 10 Jahre, SD = 1,1) abgebildet wird und jedes Jahrzehnt vertreten ist (vgl. 2.1).

Abbildung 3 zeigt zwei Maße für diese Stichprobe von Nachrichtensendungen. In der linken Grafik in Abbildung 3 ist zu sehen, dass der prozentuale Anteil der Kernnachrichten (COR) im Laufe der Zeit ab- und der der Berichte (REP) zunimmt. Der Anteil aller anderen Elemente (OTH) hat abgenommen.

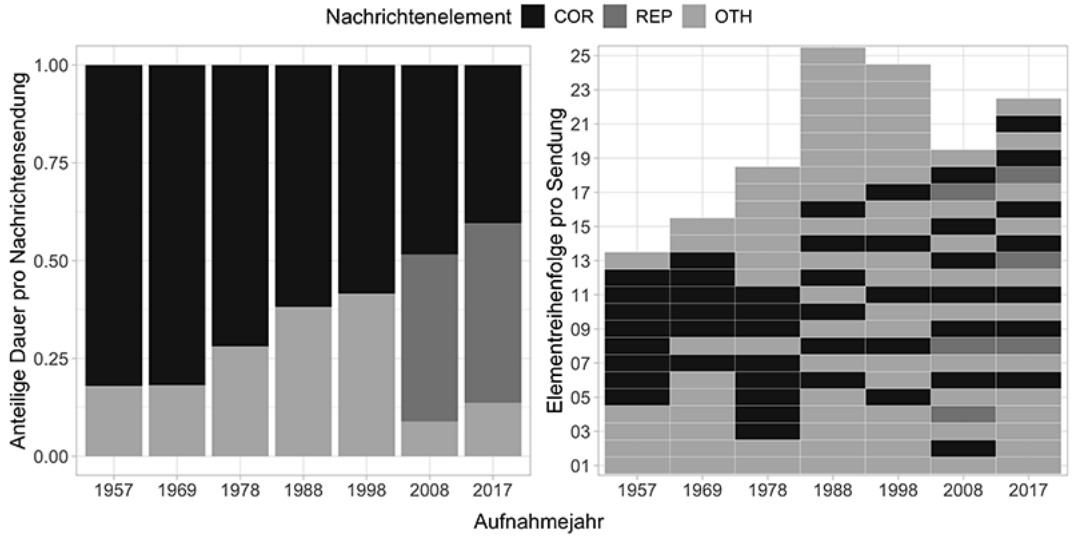

Abb. 3: An einer Nachrichtensendung anteilige Dauer ausgewählter Nachrichtenelemente (links) und deren Reihenfolge innerhalb einer Sendung (rechts; COR = Kernnachricht, REP = Bericht, OTH = andere)

Mit Fokus auf Kernnachrichten (COR) zeigt die rechte Grafik in Abbildung 3, dass diese Art von Elementen in den ältesten Sendungen früher einheitlicher am Ende einer Sendung vorgetragen wurden. Seit den 1980er-Jahren werden Kernnachrichten mit anderen Nachrichtenelementen durchsetzt präsentiert. Während die Zahl der Kernnachrichten innerhalb

einer Sendung ungefähr gleich geblieben ist, hat die Zahl anderer Nachrichtenelemente zudem zugenommen.

3.2 Tempo- und Pausenanalysen der Kernnachrichten

Die Analysen in diesem Abschnitt konzentrieren sich auf die Teilmenge der 251 Kernnachrichten aller 43 hier analysierten Sendungen. Mithilfe linearer Regressionen wurde der Zusammenhang zwischen Aufnahmejahr (= Sendung) und Sprechgeschwindigkeit bzw. Pausen pro Minute geprüft. Zuvor wurden die für jede Kernnachricht vorliegenden Werte der abhängigen Variable gemittelt, so dass pro Sendung (= Aufnahmejahr) nur ein Wert pro Sprecher bzw. Sprecherin vorlag, und die Verteilung der Residuen geprüft.

In Abbildung 4 ist die aus der signal- und der MAUS-basierten Analyse abgeleitete Sprechgeschwindigkeit dargestellt. Beide Analysen zeigen einen allmählichen Anstieg der Sprechgeschwindigkeit im Zeitraum von 1956 bis 2017, was sich in signifikanten Korrelationen widerspiegelt (signalbasierten Analyse: $R^2 = 0{,}29$, $F[1{,}41] = 16{,}56$, $p < 0{,}001$; MAUS-basierte Analyse: $R^2 = 0{,}17$, $F[1{,}41] = 8{,}68$, $p < 0{,}01$). Bei Betrachtung einzelner Zeiträume scheinen dabei Phasen der Sprechgeschwindigkeitsveränderung (1956–1957, 2006–2017) auf Phasen der Stabilität zu folgen (1986–1995). Eine statistische Analyse dieser Phasen ist erst bei Vorlage mehrerer Sprecherdaten sinnvoll. Die Veränderungen über die Zeit sind in beiden Analysen insgesamt ähnlich, obwohl die Sprechgeschwindigkeit bei der MAUS-basierten Analyse im Vergleich zur signalbasierten Analyse im Allgemeinen höher ist.

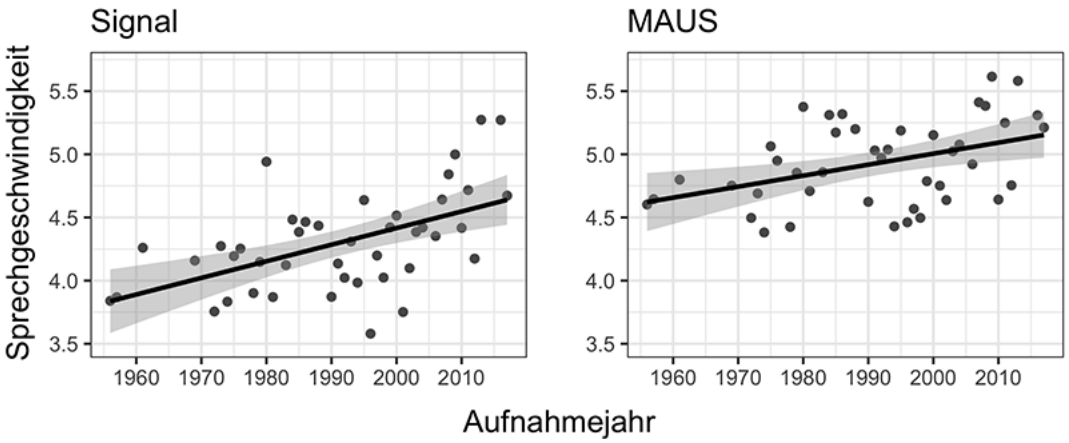

Abb. 4: Sprechgeschwindigkeit (s/s inklusive Pausen) in Abhängigkeit des Aufnahmejahres getrennt nach Messmethode (links: signalbasiert, rechts MAUS-basiert)

Die Anzahl der Pausen pro Minute ist hingegen beim signalbasierten Ansatz insgesamt geringer als beim MAUS-basierten Ansatz (siehe Abb. 5). Aber auch hier unterscheiden sich die Messungen nicht hinsichtlich des allgemeinen Trends hin zu weniger Pausen pro Minute in den jüngeren Aufnahmen (signalbasierte Analyse: $R^2 = 0{,}15$, $F[1{,}41] = 7{,}2$, $p < 0{,}05$; MAUS-basierte Analyse: $R^2 = 0{,}24$, $F[1{,}41] = 13{,}07$, $p < 0{,}001$).

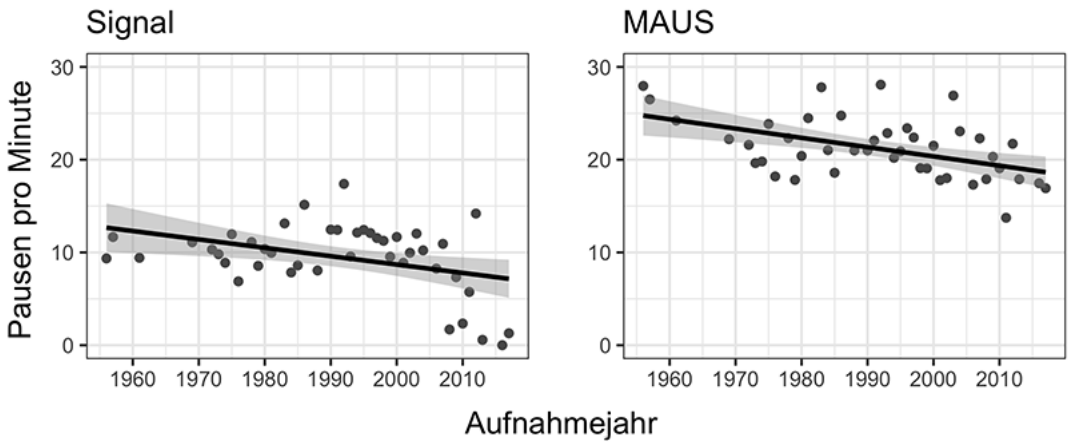

Abb. 5: Mittlere Anzahl der Pausen normalisiert für die mittlere Dauer der Kernnachrichten in Abhängigkeit des Aufnahmejahres getrennt nach Messmethode

Die jeweils ähnlichen Ergebnisse für beide Ansätze legitimieren die messmethodisch unabhängige Analyse der verbleibenden abhängigen Variablen, die entweder aus der signal- (Artikulationsrate, vgl. 2.3.2) oder der MAUS-basierten (Pausen- und IPU-Dauer, vgl. 2.3.3) Messung abgeleitet wurde.

Die linke Graphik in Abbildung 6 zeigt, dass es keine signifikante Korrelation zwischen Aufnahmejahr und Artikulationsrate im hier analysierten Korpus gibt (R^2 = 0,04, F[1,41] = 1,9, p = 1,18), die Artikulationsrate einzelner IPUs also im Gegensatz zur Sprechgeschwindigkeit nicht signifikant abgenommen hat.

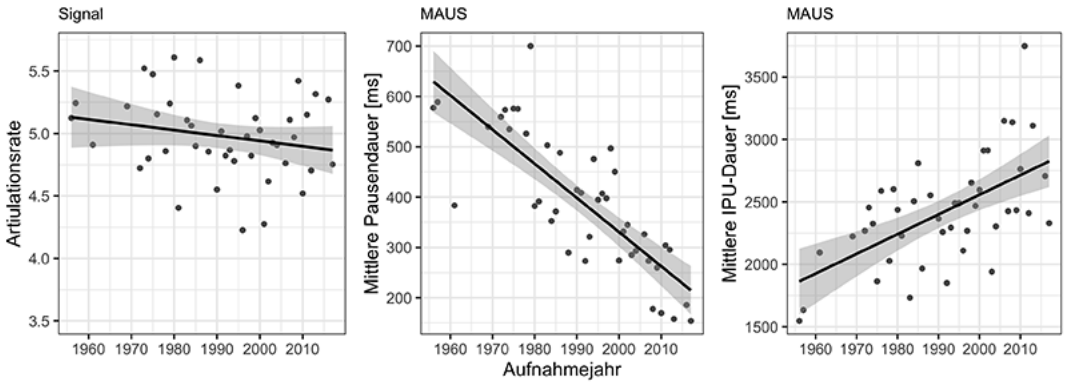

Abb. 6: Artikulationsrate (s/s exklusive Pausen; links), mittlere Pausen- (Mitte) und IPU-Dauer (rechts) in Abhängigkeit des Aufnahmejahres

Die mittlere Dauer der Pausen (Abb. 6 Mitte) hingegen hat über den 60-jährigen Zeitraum deutlich abgenommen (R^2 = 0,64, F[1,41] = 74,28, p < 0,001). Die Kürzung der Pausendauer geht einher mit einer über die Jahrzehnte signifikanten Zunahme der IPU-Dauer (Abb. 6 rechts; R^2 = 0,35, F[1,41] = 21,76, p < 0,001). Der beobachtete Trend in der Zunahme der Sprechgeschwindigkeit lässt sich somit auf weniger und kürzere Pausen zurückführen. Und obwohl die IPUs in den neueren Aufnahmen länger waren, hat sich die Artikulationsrate nicht erhöht, d. h. in jüngeren Sendungen wurden nicht mehr Silben pro Sekunde realisiert im Vergleich zu älteren.

4. Diskussion

Die Ziele der vorliegenden Studie lagen in der Vorstellung eines halbautomatisierten Verfahrens zur Aufbereitung eines Nachrichtenkorpus für Sprachwandelstudien im Allgemeinen und Lautwandelstudien im Besonderen sowie ersten Analysen zur Verteilung von Nachrichtenelementen innerhalb einzelner Sendungen und von Sprechgeschwindigkeit, Nachrichten- und Pausendauern in den Kernnachrichten.

Sowohl die Anordnung der Nachrichtenelemente als auch die Sprechgeschwindigkeit haben sich in dem hier untersuchten Korpus über den 60-jährigen Zeitraum verändert. Der Anteil der Kernnachrichten ist in den jüngeren Sendungen zurückgegangen und wird nicht länger *en bloc* präsentiert. Im Gegensatz zu französischen Radionachrichten (de Mareüil/Rilliard/Allauzen 2012), aber in Übereinstimmung mit früheren Berichten zum Italienischen (Savino/Wehrle/Grice 2024) zeigten die signal- und MAUS-basierten Analysen der Kernnachrichten, dass die Sprechgeschwindigkeit im Laufe der Zeit zunahm[7], nicht jedoch die Artikulationsgeschwindigkeit, was im hier untersuchten Korpus auf weniger und kürzere Pausen zurückzuführen ist. Inwiefern das Aufkommen privatrechtlicher Radiostationen und/oder die Anpassung an ein sich veränderndes Zielpublikum als mögliche Erklärungsansätze herangezogen werden können, wäre in weiteren Studien zu prüfen. Gleiches gilt für die Untersuchung einer möglichen Zunahme der Sprechgeschwindigkeit in der Standardsprache nicht-professioneller Sprecherinnen und Sprecher im selben Zeitraum, wobei dies aufgrund fehlender, vergleichbarer Daten kaum möglich ist. Gerade aufgrund ihrer Verfügbarkeit und der Repräsentativität für die Standardsprachen bieten sich die Daten für Sprachwandelstudien zur Standardaussprache an (vgl. 1.1 und Bose/Grawunder/Schwarz 2018). Dennoch ist anzumerken, dass die Stichprobe durch einzelne Sprechende, die jeweils ein bis wenige Jahr(e) repräsentieren, beeinflusst worden sein kann. Es handelt sich nicht um eine Langzeitstudie einzelner Nachrichtensprechenden. Die aktuelle Stichprobe von 43 Fernsehnachrichten mit 251 Kernnachrichtenelementen aus fast 60 Jahren muss also erweitert werden – möglichst mit vergleichbaren Daten anderer öffentlich-rechtlicher Sender – um die hier beschriebenen Trends zur Sprech- und Pausenrate vertiefend zu untersuchen. Auch eine weiterführende Analyse von Pausen, deren Bedeutung hier schon deutlich wurde, sowie weiterer prosodischer Merkmale (vgl. 1.2) ist wichtig und angedacht. Die Bewertung der Qualität der Pausenerkennung wird dabei zentral sein.

Die bei der Datenaufbereitung gewonnenen Erfahrungen haben zu einer Verbesserung des Octra-Backends geführt: Die lokale ASR ist nun als Hintergrundprozess innerhalb des Tools zugänglich, und das Schneiden des Audiosignals sowie die Erstellung paralleler Textdateien wurden als neue Funktionen integriert. Eine automatische Kennzeichnung der Nachrichtenelemente könnte von einer KI-basierten Lösung profitieren, sobald entsprechende Trainingsdaten mit manuellen Segmentierungen vorliegen. Ebenso wäre die automatische Erkennung, Reduzierung und Subtraktion von nichtsprachlichen Elementen (Atmung, Musikuntermalung) ein weiterer Schritt in der Datenvorverarbeitung, um bessere Ergebnisse bei der Erkennung von Pausen und Silben zu erzielen. Die Vorverarbeitung von Rundfunkaufnahmen für phonetisch-linguistische Korpora (die auch für andere Wissenschaftsbereiche relevant sind) ist aufwendig. Eine massive Reduzierung des manuellen Arbeitsaufwands, wie sie hier gezeigt wird, ist daher zu begrüßen.

7 Das erklärt möglicherweise auch die unterschiedlichen Sprechgeschwindigkeitswerte in Grawunder/Kettel (2015) und Iivonen/Niemi/Paananen (1995) (vgl. 1.1).

Literatur

Bain, Max/Huh, Jaesung/Han, Tengda/Zisserman, Andrew (2023): WhisperX: Time-accurate speech transcription of long-form audio. In: Proc. Interspeech 2023, 20–24 August 2023, Dublin, Ireland, S. 4489–4493.

Barbosa, Plinio A./Madureira, Sandra/de Mareüil, Philippe B. (2017): Cross-linguistic distinctions between professional and non-professional speaking styles. In: Proc. Interspeech 2017, August 20–24, 2017, Stockholm, Sweden, S. 3921–3925.

Bose, Ines/Schwiesau, Dietz (Hg.) (2011): Nachrichten Schreiben, sprechen, hören: Forschungen zur Hörverständlichkeit von Radionachrichten. Berlin: Frank & Timme.

Bose, Ines/Grawunder, Sven/Schwarze, Cordula (2018): Alles RElativ oder relaTIV? Sprachwandel, (Standard-)Aussprache und DaF-Unterricht. In: Moraldo, Sandro M. (Hg.): Sprachwandel: Perspektiven für den Unterricht Deutsch als Fremdsprache. (= Sprache – Literatur und Geschichte. Studien zur Linguistik/Germanistik 49). Heidelberg: Winter, S. 69–114.

Bülow, Lars/Scheutz, Hannes/Wallner, Dominik (2019): Variation and change of plural verbs in Salzburg's base dialects. In: Dammel, Antje/Schallert, Oliver (Hg.): Morphological variation. Theoretical and empirical perspectives. (= Studies in Language Companion Series 207). Amsterdam/Philadelphia: Benjamins, S. 95–134.

de Jong, Nivja H./Wempe, Ton (2009): Praat script to detect syllable nuclei and measure speech rate automatically. In: Behavior Research Methods 41, 2, S. 385–390.

De Mareüil, Philippe B./Rilliard, Albert/Allauzen, Alexandre (2012): A diachronic study of initial stress and other prosodic features in the French news announcer style: corpus-based measurements and perceptual experiments. In: Language and Speech 55, 2, S. 263–293.

Draxler, Christoph/Pömp, Julian (2024): Octra Backend – eine skalierbare Infrastruktur für Transkriptionsprojekte. In: Konferenz Elektronische Sprachsignalverarbeitung (ESSV) 2024, Regensburg, S. 102–107.

Draxler, Christoph/Kleber, Felicitas/Grawunder, Sven/Trouvain, Jürgen (2025): Teilautomatisierter Workflow zur Aufbereitung großer Audiodatenmengen für signalbasierte Analysen. In: Konferenz Elektronische Sprachsignalverarbeitung (ESSV) 2025, Halle/Saale, S. 138–145.

Dussel, Konrad (2022): Deutsche Rundfunkgeschichte. 4. Aufl. Köln: Halem.

European Commission: Directorate-General for Research and Innovation/Senftleben, Martin R. F. (2022): Study on EU copyright and related rights and access to and reuse of data. Luxemburg: Publications Office of the European Union. https://data.europa.eu/doi/10.2777/78973 (Stand: 10.12.2025).

Gordon, Elizabeth/Maclagan, Margaret/Hay, Jennifer (2007). The ONZE corpus. In: Beal, Joan C./Corrigan, Karen P./Moisl, Hermann L. (Hg.): Creating and digitizing language corpora. Bd. 2: Diachronic databases. London: Palgrave Macmillan, S. 82–104.

Grawunder, Sven (2011): Die Erforschung des Sprechens mittels Nachrichtenkorpora: Die Nachrichtenarche der ARD. In: Bose/Schwiesau (Hg.), S. 157–178.

Grawunder, Sven/Kettel, Sonja (2015): Anmoderieren / Überleiten / Antexten. In: SPIEL. Eine Zeitschrift für Medienkultur 1, 1, S. 151–170.

Hahn, Matthias/Siebenhaar, Beat (2019): Schwa unbreakable – Reduktion von Schwa im Gebrauchsstandard und die Sonderposition des ostoberdeutschen Sprachraums. In: Kürschner/Habermann/Müller (Hg.), S. 215–236.

Hecht, Robert/Riedler, Jürgen/Backfried, Gerhard (2002): German broadcast news transcription. In: 7th International Conference on Spoken Language Processing 2002, Denver, Colorado, USA, September 16–20, S. 1753–1756.

Iivonen, Antti/Niemi, Tuija/Paananen, Minna (1995): Comparison of prosodic characteristics in English, Finnish and German radio and TV newscasts. In: Proc. International Congress of Phonetic Sciences (ICPhS) 1995, Stockholm, S. 382–385.

Kisler, Thomas/Kleber, Felicitas (2019): Zur Validität automatisch segmentierter Daten: Eine akustische Analyse der mittelbairischen Lenisierung im Deutsch Heute-Korpus. In: Kürschner/Habermann/Müller (Hg.), S. 279–311.

Kisler, Thomas/Reichel, Uwe/Schiel, Florian (2017): Multilingual processing of speech via web services. In: Computer Speech & Language 45, S. 326–347.

Kleber, Felicitas (2016): Lautwandel. In: Domahs, Ulrike/Primus, Beate (Hg.): Handbuch Laut – Gebärde – Buchstabe. (= Handbücher Sprachwissen 2). Berlin/Boston: De Gruyter, S. 125–143.

Kleber, Felicitas (2025): Synchrone und diachrone Variation in der temporalen Struktur gesprochener Wörter in Varietäten der DACH-Region: Empirie, Simulation und ein interaktiv-phonetischer Lautwandelansatz. In: Proske, Nadine/Weber, Thilo/Dannerer, Monika/Deppermann, Arnulf (Hg.): Gesprochenes Deutsch: Struktur, Variation, Interaktion. (= Jahrbuch des Instituts für Deutsche Sprache 2024). Berlin/Boston: De Gruyter, S. 105–130.

Kleber, Felicitas/Schmid, Stephan (im Dr.): Towards a signal-based typology of consonant quantity in southern German varieties. In: Filipponio, Lorenzo/Dipino, Dalila/Garassino, Davide (Hg.): Exploring the relation between duration and length. Italo-Romance, Romance and beyond. (= Linguistik Aktuell). Amsterdam/Philadelphia: Benjamins.

Krech, Eva-Maria/Stock, Eberhard/Hirschfeld, Ursula/Anders, Lutz C. (2009): Deutsches Aussprachewörterbuch. Berlin/New York: De Gruyter.

Kürschner, Sebastian/Habermann, Mechthild/Müller, Peter O. (Hg.) (2019): Methodik moderner Dialektforschung: Erhebung, Aufbereitung und Auswertung von Daten am Beispiel des Oberdeutschen. (= Germanistische Linguistik 241–243). Hildesheim/Zürich/New York: Olms.

Labov, William/Rosenfelder, Ingrid/Fruehwald, Josef (2013): One hundred years of sound change in Philadelphia: Linear incrementation, reversal, and reanalysis. In: Language 89, 1, S. 30–65.

Ladefoged, Peter (2001): Vowels and consonants – An introduction to the sounds of languages. Maldon, MA/Oxford: Blackwell.

Lameli, Alfred (2004): Standard und Substandard: Regionalismen im diachronen Längsschnitt. (= Zeitschrift für Dialektologie und Linguistik. Beihefte 128). Stuttgart: Steiner.

Lindblom, Björn (1990): Explaining phonetic variation: A sketch of the H&H theory. In: Hardcastle, William J./Marchal, Alain (Hg.): Speech production and speech modelling. Dordrecht: Springer, S. 403–439.

Margoni, Thomas/Kretschmer, Martin (2022): A deeper Llok into the EU text and data mining exceptions: harmonisation, data ownership, and the future of technology. In: GRUR International 71, 8, S. 685–701.

Mok, Peggy P. K./Fung, Holly S. H./Li, Jingwen (2014): A preliminary study on the prosody of broadcast news in Hong Kong Cantonese. In: Proc. Speech Prosody 2014, Dublin, S. 1072–1075.

Page, Richard (1999): The Germanic Verschärfung and prosodic change. In: Diachronica 16, 2, S. 297–334.

Pellegrino, François/Coupé, Christophe/Marsico, Egidio (2011): A cross-language perspective on speech information rate. In: Language 87, 3, S. 539–558.

Pömp, Julian/Draxler, Christoph (2017): Octra – A configurable browser-based editor for orthographic transcription. In: Proc. Phonetik und Phonologie 2017, Berlin, S. 145–148.

Rilliard, Albert/Doukhan, David/Uro, Rémi/Devauchelle, Simon (2023): Evolution of voices in French audiovisual media across genders and age in a diachronic perspective. In: Proc. International Congress of Phonetic Sciences (ICPhS) 2023, Prag, S. 753–757.

Savino, Michelina/Wehrle, Simon/Grice, Martine (2024): The prosody of Italian newsreading: a diachronic analysis. In: Proc. Speech Prosody, 2–5 July 2024, Leiden, The Netherlands, S. 836–840.

Schwiesau, Dietz/Grawunder, Sven/Bose, Ines (2011): Die Nachrichtenarche der ARD. In: Bose/Schwiesau (Hg.), S. 147–155.

Spiekermann, Helmut (2007): Standardsprache im DaF-Unterricht: Normstandard – nationale Standardvarietäten – regionale Standardvarietäten. In: Linguistik Online 32, 3, S. 119–137.

Šturm, Pavel/Volín, Jan (2023): Occurrence and duration of pauses in relation to speech tempo and structural organization in two speech genres. In: Languages 8: 23. https://doi.org/10.3390/languages8010023 (Stand: 22.1.2026).

Trouvain, Jürgen (2015): Notes on the development of speaking styles over decades — the case of live football commentaries. In: Proc. First International Workshop on the History of Speech Communication Research 2015 (HSCR), Dresden, Germany, September 4—5, S. 160–166.

Trouvain, Jügen/Werner, Raphael/Möbius, Bernd (2020): An acoustic analysis of inbreath noises in read and spontaneous speech. In: Proc. Speech Prosody 2020, 25—28 May, Tokyo, Japan, S. 789–793.

Vaessen, Nik (2018): JiWER: Similarity measures for automatic speech recognition evaluation. Python Package Index. https://pypi.org/project/jiwer/ (Stand: 10.12.2025).

Winkelmann, Raphael/Harrington, Jonathan/Jänsch, Klaus (2017): EMU-SDMS: Advanced speech database management and analysis in R. In: Computer Speech and Language 45, S. 392–410.

Zellou, Georgia/Tamminga, Meredith (2014): Nasal coarticulation changes over time in Philadelphia English. In: Journal of Phonetics 47, S. 18–35.

Kontaktinformation

Felicitas Kleber
Fachrichtung Sprachwissenschaft und Sprachtechnologie
Universität des Saarlandes
Campus C7.2
66123 Saarbrücken
E-Mail: kleber@lst.uni-saarland.de

Christoph Draxler
Institut für Phonetik und Sprachverarbeitung
Ludwig-Maximilians-Universität München
Schellingstraße 3
80799 München
E-Mail: draxler@phonetik.uni-muenchen.de

Jürgen Trouvain
Fachrichtung Sprachwissenschaft und Sprachtechnologie
Universität des Saarlandes
Campus C7.2
66123 Saarbrücken
E-Mail: trouvain@lst.uni-saarland.de

Sven Grawunder
Institut für Musik-, Medien- und Sprechwissenschaften
Martin-Luther-Universität Halle-Wittenberg
Emil-Abderhalden-Straße 26/27
06108 Halle (Saale)
E-Mail: sven.grawunder@sprechwiss.uni-halle.de

Bibliografische Angaben

Dieser Text ist Teil der Publikation: Brunner, Annelen/Hansen, Sandra/Lang, Christian/Tu, Ngoc Duyen Tanja/Wolfer, Sascha (Hg.) (2026): Deutsch im Wandel – Tools, Methoden, Ressourcen. Beiträge zur Methodenmesse der IDS-Jahrestagung 1. (= IDS*open* 16). Mannheim: IDS-Verlag. https://doi.org/10.21248/idsopen.16.2026.63.