



ドイツ語テレビニュース番組の語彙使用 : コレスポ ンデンス分析とランダムフォレストによる比較研究 —tagesschauとheuteの差異を中心に—

今道, 晴彦

(Citation)

統計数理研究所共同研究リポート, 484:145-159

(Issue Date)

2026-03-25

(Resource Type)

departmental bulletin paper

(Version)

Version of Record

(JaLCD0I)

<https://doi.org/10.24546/0100502568>

(URL)

<https://hdl.handle.net/20.500.14094/0100502568>



ドイツ語テレビニュース番組の語彙使用:

コレスポンデンス分析とランダムフォレストによる比較研究

—*tagesschau*と*heute*の差異を中心に—

今道 晴彦(広島大学)

Vocabulary usage in German TV News Programs:

A Comparative Study Using Correspondence Analysis and Random Forest

—Focusing on the Differences between *tagesschau* and *heute*—

IMAMICHI, Haruhiko (Hiroshima University)

概要

本研究は、ドイツ語のリスニング学習のための共通基盤の構築に繋げるべく、ドイツの公共放送のテレビニュース番組である *tagesschau*(ARD)と *heute*(ZDF)における語彙使用の差異を実用的観点から検証した。その際、両番組の字幕をコーパス化した上で、頻度上位 100 語の分布をコレスポンデンス分析で可視化し(RQ1), ランダムフォレストで識別項目の特定(RQ2)を行った。その結果、RQ1 では、両番組がほぼ二つに大別され、*tagesschau* は出来事の進展を描写することで、客観的叙述に徹し、*heute* は視聴者目線の視点を重視するという報道スタイルの違いあることが示唆された。また、RQ2 では、両番組を 95%の正解率で識別する上位 20 語が特定され、そのすべてが *heute* において顕著に現れ、ニュースのトピックを担う内容語よりも多義的用法を持つ機能語であることなどが確認された。このことから、どちらの番組を視聴するかによって、報道スタイルが変わり、とくに *heute* においては機能語の使われ方に特段の注意を払う必要があることが示唆された。

キーワード

テレビニュース, 語彙使用, コレスポンデンス分析, ランダムフォレスト

1. はじめに

テレビニュース番組では、書き言葉の正確さと話し言葉の伝わりやすさを併せ持つ言語使用が特徴的で、一般に中立的で、客観的描写に注意が払われている。とくにドイツでは、公共放送で用されるドイツ語が最も標準的であるとみなされることが多く、番組自体も高い評価を得てきた。公共放送の看板ニュース番組である *tagesschau*(ARD)と *heute*(ZDF)はその代表格で、365 日 24 時間体制でニュースを配信しており、番組やウェブ記事で発生した事実誤認や映像の取り違え

といったいわば負の側面も番組 HP で公開しており, Reuter Institute によるニュース消費行動に関するアンケート調査では, ドイツで最も信頼性の高いニュースブランドであることが示されている (Behre et al., 2025)。tagesschau は, 1952 年に放送を開始し, 1961 年からは現在と同じ 20 時から 15 分間の時間枠でニュース放送を行なっている。番組の公式情報によると, 2024 年に 1 日平均で 1 日平均 954 万 9 千人の視聴者数を記録し (<https://www.ard.de/die-ard/organisation-der-ard/ausgaben-fuer-nachrichten-100.html>), ドイツのニュース番組の草分け的存在となっている。また, heute は 1963 年に ZDF の開局とともに放送を開始し, 1973 年からは現在と同じ 19 時から約 20 分の放送枠となっている。heute はすでにニュース番組として定着していた tagesschau に対抗すべく, 放送開始時間を早めるだけでなく, たとえば司会者による主観的意見の表明や日常会話的な語りの導入などの差別化も行なってきた (Muckenhaupt, 1994)。ともに高い評価を得ており, 多くの視聴者を抱える番組であるがゆえに, そこで使用されるドイツ語は極めて一般性の高いドイツ語と言える。しかし, その一方で, テレビニュース番組のドイツ語には, 独特の現象も見られる。たとえば, テレビニュース番組では, 最も重要な情報が冒頭に配置され, 徐々に情報の重要度が弱まるという傾向が見られる。いわゆる逆三角形 (Umgekehrte Pyramide) と呼ばれるこのスタイルは文学作品や学術論文には馴染まず, 新聞を含めたニュース報道に特徴的とも言える。また, ドイツのテレビニュース番組では, 情報源や出所の明示が法的に義務付けられている。そのため, 以下の例文 (1) (2) に見られるように, 情報源や出所が前置詞句 (*Nach Angaben der Polizei*) や発話動詞 (*sagte*) などの語彙的標識や, 接続法第 I 式 (*sei...festgenommen worden*) といった文法的標識によって言語化されることが多い。

(1) **Nach Angaben der Polizei sei der Täter gegen 6 Uhr festgenommen worden.**

(2) **Der Täter sei gegen 6 Uhr festgenommen worden, sagte die Polizei.**

(警察によりますと、犯人は 6 時ごろ逮捕されたということです。)

とくに接続法 I 式は, 発話内容の帰属を明示するとともに, 発話者の主張を抑制することで客観性を高めるという機能を担っており, 新聞やニュースにおいて頻繁に見られると言われている (Duden - Die Grammatik, 2022)。また, 例文中にある受動表現 (*festgenommen worden*) も, 出来事中心の叙述が好まれるテレビニュース番組では頻繁に登場する言語形式でもある。

こうしたテレビニュース番組の文体的特徴については他にも多数報告されており, 実証的観点からデータに基づく検証も行われてきた (Bruns & Marcinkowski, 1997; Wannenmacher, 2024)。しかし, 従来の研究は, たとえば受動表現や名詞文体といった個々の言語現象をミクロ的に取り上げてきたが, それらの個別の特徴が言語現象全体の中でどのように関連しているのかといったマクロ的な視点については, あまり検討されてこなかった。また従来の研究は, 分析対象となるデータが比較的小規模で, 頻度調査を中心とする記述にとどまり, 統計的検証はほとんどなされていない。そのため, テレビニュース番組の文体を多角的に捉え, その内部構造を体系的に把握することもテレビニュース番組の文体的特徴を捉える上では, 今後必要になると思われる。

さらに、テレビニュース番組はリスニング学習において重要な教材となるもので、CEFR(ヨーロッパ言語共通参照枠)の自己評価表では、「UNDERSTANDING(理解すること)」という項目の中の「Listening(聞くこと)」において、テレビニュース番組が学習到達度を測る指標のひとつとして取り上げられている(Council of Europe, 2020)。したがって、テレビニュース番組のリスニングはドイツ語学習者にとって目指すべき通過点のひとつになっている。また、国内の大学では、CEFRを学習到達度の参考指標としている教育機関も少なくない。とくにドイツの大学に正規留学するためには、CEFRの基準が適用されるため、日本のドイツ語教育においても、看過できる問題ではない。

しかしながら、日本の状況を考えた場合、テレビニュース番組のリスニングは、学習者にとっても、教員にとってもなかなか敷居が高いのが実情で、難度が高い上に、学習項目が体系化されていないという点で、何から取り組むべきか路頭に迷う学習者がいてもおかしくない。また、教員にとっても、テレビニュース番組を教材として利用する場合、共通基盤が構築されておらず、教員の経験的判断に委ねられる部分が多い。そのため、テレビニュース番組がどのような文体的特徴を有しているのか、また、個々の番組間でどのような差異が存在するのかといったといった問題を明らかにすることは、学習方法や教育方法の体系化にも繋がる。その際、語彙使用の解明はこうした問題にアプローチする上で取り組むべき重要な課題のひとつになり、実用面で有益であると考えられる。

2. リサーチデザインと手法

2.1 研究目的とRQ

以上の考えに立脚し、本研究では、ドイツで高い信頼性を有し、リスニング学習の生きた教材とも言うべき *tagesschau* と *heute* を対象として、頻度上位 100 語(レマ)を手がかりとし、番組間の差異を実用的観点から明らかにすることを目的とする。両番組はインターネット上の HP より無償公開されており、信頼性の高い公式の字幕が付されている。ただし、欠損する数本の番組については YouTube の公式チャンネルから入手した。これらの番組は、リスニング学習の生きた教材であるだけのみならず、日本人ドイツ語学習者にとっても、もっとも安定的に利用できるニュース番組と言える。他にも、民間放送の利用も考えられるが、主要な番組は日本からの利用に制限がある。また、YouTube に関しても、公共放送はテレビ放送の内容をほぼそのまま放映しているが、民間放送の多くはその限りではない。また、ライブ放送や、ドイツ語学習者向けに教材としての利用価値が高い豊富な学習用コンテンツを提供する *Deutsche Welle* がドイツ語関係者の間ではよく知られているが、一般向けのすべてのニュース放送をサポートしているわけではない。*tagesschau* と *heute* は学習教材として難度が高いが、ドイツの代表的テレビニュース番組であるこれらの番組で使われるドイツ語が文体的にどのような特徴を有しているのかを整理することは、学習方法や教育方法の体系化にも繋がる。その際、マクロ的視点に立脚して、音声、形態、語彙、統語、構文などの多層的な観点からテレビニュース番組の言語使用を捉えることが望ましいが、本研究では、語彙使用に焦点を当て、具体的には、頻度上位 100 語に限定した上で、以下のリサーチクエスション(RQ)を設定する。

RQ1: 両番組における語彙の分布はどのような特徴を示すか？

RQ2: 両番組の識別に寄与する語彙指標は何か？

2.2 データ

分析に際し、本研究では、*tagesschau*と*heute*の2025年9月1日から2025年12月29日に放映された120日分の字幕を文字起こしし、コーパスとして使用した。番組全体としての言語的特徴や編集方針の差を浮き彫りにするためには、放送されたすべての内容を分析対象とすべきであるとの考えに基づき、収集範囲は冒頭の挨拶からエンディングの挨拶までとした。字幕は自動生成されたものであるものの、聴覚障害者向けに番組のガイドラインに沿って制作されており、誤変換の可能性は極めて低いものと考えられる。

表1は、両番組のテキストサイズの概要をまとめたものである。両番組の総語数についてウェルチの独立二群の t 検定を行った結果、*heute* ($M=2549.19$, $SD=554.03$) は、*tagesschau* ($M=1767.26$, $SD=163.42$) に比べ、約1.4倍多く、統計的に有意に語数が多かった ($t(139.55)=-14.83$, $p<.001$, 95% $CI[-886.19, -677.68]$)。この結果は、*heute* が *tagesschau* よりも放送時間が5分長いため、結果として提供される情報量も多いという物理的背景を裏付けるものであると考えられる。さらに、両番組の標準偏差(SD)に注目すると、*tagesschau* (163.42語) よりも、*heute* (554.03語) のほうが3倍以上大きい。これは *tagesschau* の放送時間が厳格であるのに対して、*heute* は放送回によって語数の変動が大きいことを示唆しており、番組の内容に応じて柔軟に放送分量を調整している可能性が高いことが示唆される。

表1

両番組のテキストサイズ(語数)

番組名	平均(M)	標準偏差(SD)	最小	最大
<i>tagesschau</i>	1767.26	163.42	1360	2554
<i>heute</i>	2549.19	554.03	849	3213

2.3 手法

2.3.1 データの事前処理

本研究では、ニュース番組における語の出現傾向を把握するため、頻度上位100語の1000語あたりの相対頻度を算出した。なお、トークン化には、Pythonの自然言語処理ライブラリであるspaCyのドイツ語モデル(*de_core_news_lg*)を使用した。同モデルは、トークン化や品詞タグ付けといった形態素解析や、文構造を明らかにする係り受け解析、さらには固有表現抽出などの高度な自然言語処理タスクを得意としている。なお、以下のRQ1の調査では変数が多くなると視認性が低くなるため、頻度上位100語を分析対象とした。

以上のことから、本研究では、ニュース番組120日分の字幕(計240ファイル分)に対して、頻度上位100語の相対頻度を求め、表2に示す頻度行列表(240×100)を作成した。

表 2

分析に用いた頻度行列表の一部 (1000 語あたりの調整頻度)

	der	in	sein	und	ein	zu	es	...
H_20250901.txt	265	76	71	58	52	35	20	...
H_20250902.txt	287	107	89	90	51	37	27	...
...	...	...	...	...	...	...	...	...
T_20251228.txt	120	40	41	13	28	23	20	...
T_20251229.txt	176	40	48	25	43	17	22	...

2.3.2 分析の手順

RQ1 では、マクロ的観点から両番組の全体的な語彙使用を捉えるため、表 2 の頻度行列を元に、コレスポンデンス分析 (Correspondence Analysis) を用いて、240 日分の各放送回を行カテゴリー、頻度上位 100 語を列カテゴリーとして頻度上位 100 語の分布を可視化した。

RQ1 によって両番組が区別されるのであれば、それぞれの番組に特徴的な語彙を抽出することは可能であるが、RQ2 では、分類精度が高い機械学習アルゴリズムとして知られているランダムフォレスト (Random Forest) によるデュアルアプローチで、両番組の識別に寄与する語彙の特定を行った。すなわち、2 つのアプローチにより、信頼性の高い分析結果を得ることを目指す。具体的には、表 2 の行列表を特徴量とし、頻度上位 100 語を説明変数、240 日分の各放送回の番組を目的変数として分類モデルを構築した。分析結果の安定性と汎用性を担保するため、5 分割交差検証 (5-fold cross-validation) を実施し、各分割データにおける平均的な特徴量重要度 (Mean Decrease Accuracy) を算出した。これにより、各番組を特徴づける頻度上位 100 語の寄与度を可視化した。

3. 結果と考察

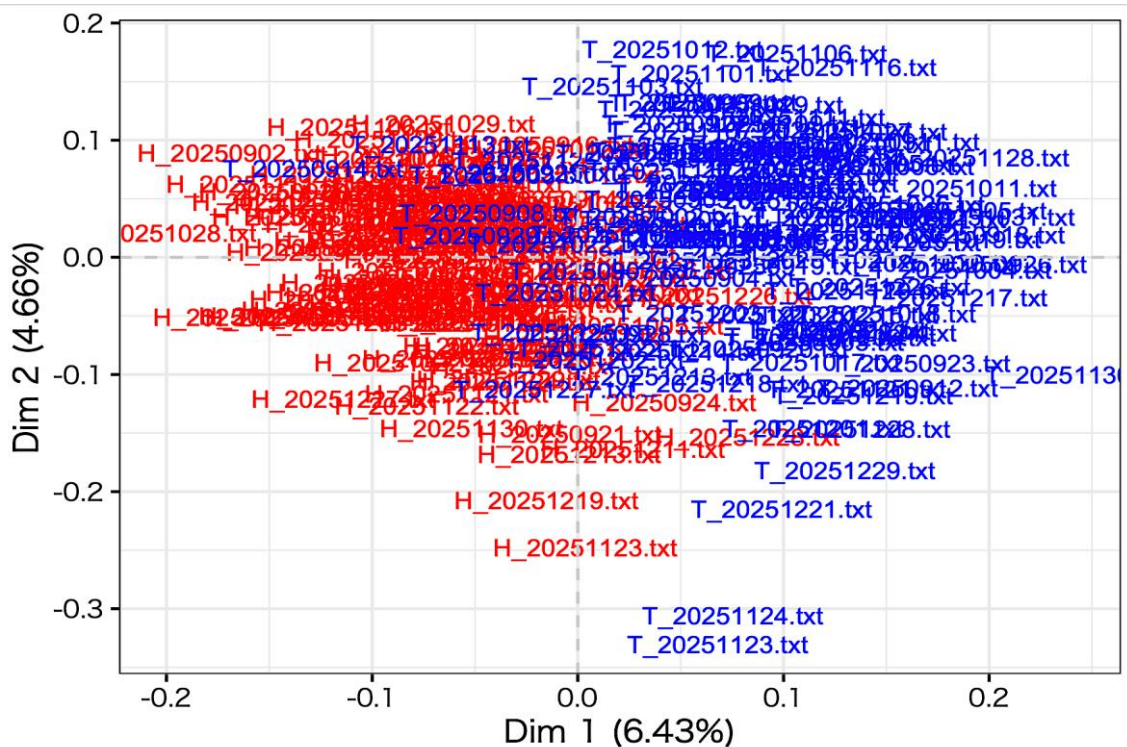
3.1 RQ1: 両番組における語彙の分布はどのような特徴を示すか？

以下の図 1 は、表 2 の頻度行列を用いてコレスポンデンス分析を行なった結果得られた出力図で、両番組の分布を示している。各番組を視覚的に識別できるようファイル名は、「番組名_年月日.txt」とし、*tagesschau* は「T_」、*heute* は「H_」ではじめており、たとえば *tagesschau* の 2025 年 12 月 1 日の放送分は「T_20251212.txt」となる。

240 日分のテレビニュース番組ということで第 1 次元の寄与率は 6.43% と低いが、プラス方向 (青色) に *tagesschau* が、マイナス方向 (赤色) に *heute* が布置され、明確に分離されるのがわかる。このことから、語彙選択の傾向が両番組では統計的に異なることが示唆される。また、*heute* については、プラス方向にもややプロットされていることが読み取れる。2.2 の表 1 のテキストサイズでは、*heute* の標準偏差が大きく、ばらつきがあることが確認されたが、語彙選択においても、ばらつきが見られることがわかる。

図 1

コレスポンデンス分析による出力図(番組)



次に、以下の図2は頻度上位語彙100語の分布を示す。番組間の差異を示す第1次元(横軸)に目を向けると、多くの語が中央に集中するものの、左右に広がっていることも確認できる。プラス方向が *tagesschau*、マイナス方向が *heute* に特徴的な語彙ということになり、*tagesschau*と比較すると、*heute* に特徴的な語彙が多いことが窺える。視認性を高めるため、それぞれに特徴的な語彙を取り出したものが表3である。

図 2

コレスポンデンス分析による出力図(頻度上位 100 語)

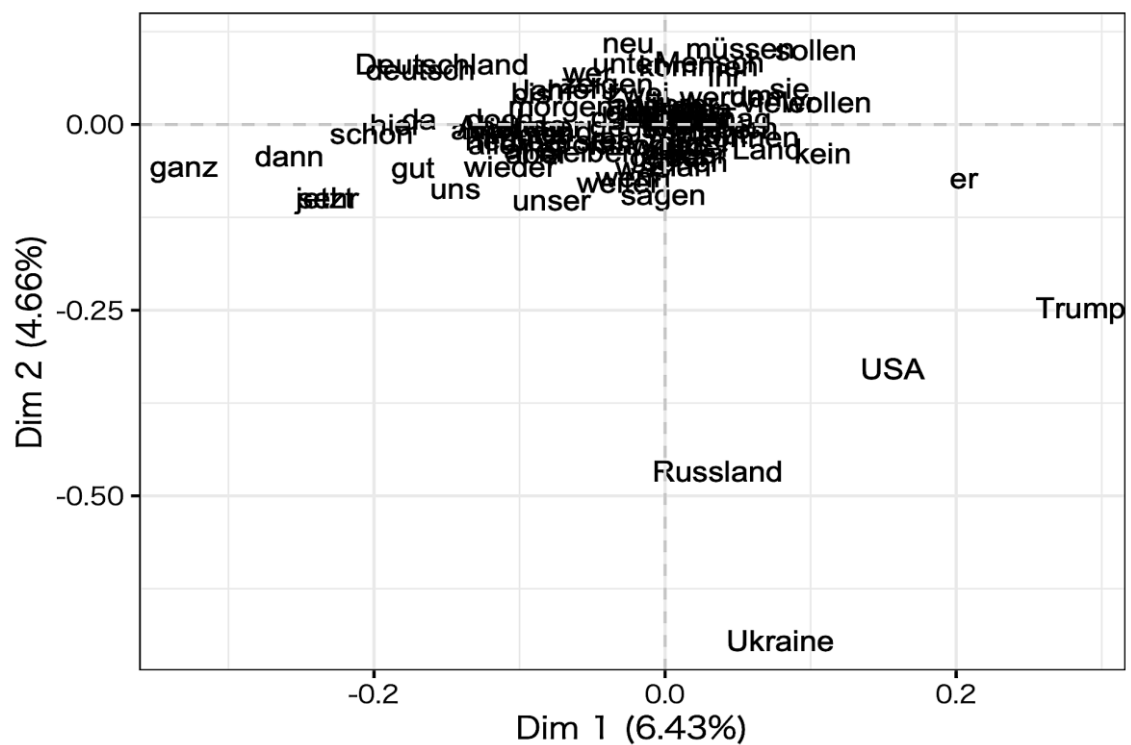


表 3

番組の識別に関わる語彙上位 20 語(第 1 次元)

	<i>tagesschau</i>	<i>heute</i>
1	Trump	ganz
2	er	dann
3	USA	jetzt
4	wollen	sehr
5	kein	schon
6	sollen	hier
7	sie	gut
8	Ukraine	da
9	viel	deutsch
10	Land	Deutschland
11	ich	uns
12	um	auch
13	können	viel
14	Tag	aller
15	müssen	doch
16	werden	Abend
17	ihr	heute
18	haben	wieder
19	an	immer
20	Russland	alle

表 3 を見ると両番組の違いが顕著に表れていることが読み取れる。まず、(頭文字が大文字で始

まる)名詞に着目すると, *tagesschau*では *Trump*(トランプ大統領), *USA*(アメリカ), *Ukraine*(ウクライナ), *Russland*(ロシア)が, *heute* では *Deutschland*(ドイツ)が確認できる。*Trump*(トランプ大統領)と *USA*(アメリカ)はとくに外れた位置にあり, *tagesschau* においてはアメリカ, とくにトランプ大統領を中心とする国外の報道に重きをおいていると言え, 国内ニュースと国際ニュースの対立のように見える。ただし, 名詞だけは, 第 1 次元(横軸)を国内ニュースと国際ニュースの対立とするには, 無理があるとともに, その他の語彙が果たしてこの対立の説明軸となるのかよく見極める必要がある。

そこで, 次に, 人称代名詞(主格)に着目すると, まず, *tagesschau* では, *er*(彼), *sie*(彼女), *ich*(私)が確認できる。3 人称と 1 人称が特徴的ということで, 一見すると矛盾するようにも見えるが, 実例に目を向けると, *ich*(私)のほとんどは, 「*Ich begrüße Sie zur tagesschau.* (ターゲスシャウです。)」というニュース冒頭で見られる挨拶表現であることがわかる。つまり, キャスターの個人的見解の表明ではなく, 儀礼的な発話を中心である。*tagesschau* においては, 例文(3)でメルツ首相を示す 3 人称の *er* が行為主体となるように, 登場人物が焦点化され, 出来事中心の客観的な叙述スタイルが優先されている。

(3) In der ARD Sendung "Arena" stellte sich Merz den Fragen. **Er** zeigte sich selbstkritisch, wie bei seinen "Stadtbild-Äußerungen", ...

ARD の「アレーナ」という番組でメルツ首相は質問に答えました。**彼**(首相)は自身の「都市景観発言」の時のように自己批判的な姿勢を見せました。

(*Tagesschau*:2025 年 12 月 9 日放送)

一方, *heute* においては 1 人称複数形が見られるが, 主格ではなく, 与格ないし対格の *uns* (我々に・我々を)が特徴的である。その用法は二つに大別され, ひとつは, 例文(4), (5)に見るように, *uns* がメディアを指す用法で, もうひとつは, 例文(6), (7)が示すように, 番組の中でメディアと視聴者の共通の視点を構築する用法である。*uns* を含む 2 語結合を調べてみると, 例文中にもあるような *bei uns*, *wir uns*, *für uns*, *von uns* などが高頻度で使用されていた。したがって, *heute* においては, 行為主体としての「我々」よりも, 番組の中で出来事を描写する際の視点としての「我々」を基準として, 情報提示がなされる傾向が強いことが示唆される。

(4) Das Finale gibt es bei **uns** ab 12:30 Uhr im Livestream.

決勝は 12 時 30 分から**我々**の(番組の)ライブストリームでご覧いただけます。

(*heute*:2025 年 9 月 12 日放送)

(5) Das war's von **uns** in aller Kürze.

以上, (我々から)手短にお伝えしました。

(*heute*:2025 年 12 月 5 日放送)

(6) Wir schauen **uns** die Details an und das sieht folgendermaßen aus.

詳細を見ていきましょう。次の通りです。

(*heute*:2025 年 9 月 22 日放送)

(7) Wir schauen **uns** die Details an und das sieht folgendermaßen aus.

詳細を見ていきましょう。次の通りです。

(*heute*:2025 年 9 月 22 日放送)

番組と視聴者との共通の視点ということと言うと、先にも言及した *Deutschland*(ドイツ)の他, *jetzt*(今), *hier*(ここ), そして自らと同じ *heute*(今日) といった語もこれに関連すると思われる。これらの語は, 時間的に, 空間的に視聴者と同じ位置に立つことを意味し, 視聴者と地続きの視点で報道していることが窺える。ちなみに, *heute*(今日)の事例に目を向けると, 特定のテーマを深掘りする同放送局の *heute journal*というニュースマガジンの番組名や, 専用アプリ *heute App*の紹介が 1 割強を占める。ただし, それを取り除いても, *heute* の事例数は多く, 「今夜」を意味する *heute Nacht* は共起語として次に多い。また, 興味深いことに, *heute* に *Nacht*(夜)が付されることで, 例文(8)にあるように, 用例の大半は天気予報によって占められる。19 時からのニュースということで, 視聴者にとっては, まず, 直近の今夜の天気(*heute Nacht*)は重要な関心事となることの表れだと考えることができる。以上で見たきたこれらの人称代名詞や, 時間, 場所を表す語は, *heute* においてはニュースが「今・ここ・我々」を基準とするコミュニケーション空間の中で情報を提示するスタイルである可能性が高いことを示している。

(8) Besonders in höheren Lagen wird es **heute Nacht** sehr frisch.

標高の高い地域では, **今夜**はかなり冷え込むでしょう。

(*heute*:2025 年 9 月 26 日放送)

また, *tagesschau*に着目すると, *wollen*, *sollen*, *können*, *müssen* といった話法の助動詞や, 受動態の助動詞 *werden* が特徴的に出現している。こうした助動詞は, 意図, 伝聞や予定, 可能性, 義務や必然性などのモダリティを表すとともに, 主観を排除し, 出来事を前景化することで客観的な距離感を生み出す機能があり, *tagesschau*が出来事の進行や状況を叙述するスタイルをとっていることが示唆される。

一方, *heute* に目を向けると, 会話でもよく登場する *schon*(すでに), *auch*(～もまた), *viel*(多くの), *aller/alle*(すべての), *wieder*(再び), *immer*(いつも)といった語も特徴的に見られる。これらの語は, 報じられる出来事の内容に関わるというよりは, 視聴者と共有された状況の中に, 出来事を位置付け, 出来事を一般的傾向として捉える機能を有している。たとえば, 例文(9)の *auch* は視聴者に加え, スポーツ担当者もいるという情報を追加するものであり, 視聴者との共有感を強

め、スポーツ担当者も含め今同じ位置にいることを示す働きがあると言える。また、例文(10)の *immer* は反復を意味し、雲がまだ残っているという状況を共有する機能を担っていると言える。

(9) Guten Abend, schön, dass Sie dabei sind, wie **auch** Katja Streso für den Sport.

こんばんは。ご覧いただきうれしく思います。スポーツ担当のカティア・シュトレーズさん**も**一緒です。

(*heute*:2025 年 10 月 22 日放送)

(10) Und da gibt es im Osten **immer** noch dichte Wolken.

そして東部では、**依然**として厚い雲に覆われています。

(*heute*:2025 年 10 月 09 日放送)

その他にも、会話などでより頻繁に使用される *ganz*(まったく)、*sehr*(非常に)といった強調語彙や、*gut*(良い)といった評価に関わる語彙、*doch*(話者の態度を示す)といった心態詞も、*heute* に特徴的に現れる。例文(11)はドイツ連邦教育研究省(Bundesforschungsministerium)の Dorothee Bär 大臣のインタビュー中の発言で、*Ich finde es ganz wichtig...*によって、自身の意見が強調されている。同番組は単に事実を述べるだけでなく、解説や中継などの対話的やりとりを取り入れるとともに、強調や評価をし、出来事の印象を視聴者に示しながらニュースを伝えていることが示唆される。

(11) ich finde es ***ganz*** wichtig, dass wir auch international auch so befreundet bleiben mit unseren verbündeten Ländern,

同盟諸国との間で、このように国際的にも友人であり続けることは**極めて**重要なことだと考えます。

(*heute*:2025 年 9 月 9 日放送)

なお、*ganz*(まったく)は一般に程度を強める副詞であるが、実例に目を向けると、天気予報での使用例が多く、例文(12)にあるような、地理的位置の端を強調する用法として定型的に用いられている。ドイツの天気予報では、日本のそれと異なり、行政区分よりも方位に基づく表現が使用され、とくにテレビニュースでは「一番南の端」のようなより誇張した表現が定型的に使われていることが示唆される。

(12) *Ganz* im Osten ist es zunächst trocken und auch ***ganz*** im Süden.

東部や南部の**境界**に近い地域では、はじめのうちは天気が持つでしょう。

(*heute*:2025 年 9 月 12 日放送)

以上の結果を総合すると、第1次元はニュースの主題の違いというよりも、ニュースの語り方や談

話スタイルの違いを反映したものであると考えられる。*tagesschau* はニュースの出来事の進展に重きを置きながら叙述するスタイルが相対的に多いのに対して、*heute* では視聴者との共有に重きを置いたスタイルを選択していることが推察される。先に、*heute* で *Deutschland*, *tagesschau* で *Trump*, *USA*, *Russland*, *Ukraine* などの語が対立的な分布を示すことを確認したが、これは政治主体が軸を形成するというよりも、出来事中心の報道の中で、結果的にこれらの語が集中したのではないかと考えられる。

また、第 2 次元(縦軸)では、第 1 次元と異なり、プラス方向に *Deutschland*, *deutsch*(ドイツの)などが、マイナス方向に *Ukraine*, *Russland*, *USA*, *Trump* などの国際政治に関する語が集中しており、こちらも国内ニュースと国際ニュースが対立する配置となっている。ただし、先の第 1 次元と異なり、マイナス側では上述の国名や固有名以外の語はまったく見られないのに対してプラス側には、*Deutschland*, *deutsch* をはじめ、多くの語が集中している。つまり、第 2 次元では *Deutschland* が話題になる時は、その話題が広範囲に及ぶことを示唆しているとも言える。一方、*Ukraine*, *Russland*, *USA*, *Trump* などが議論される際は、その限りではない。つまり、これらの国際政治に関わる語が突出して議論されていると言える。また、興味深いのは、第 1 次元では *Trump* や *USA* が *Russland* や *Ukraine* よりも突出しているのに対して、第 2 次元においては、*Ukraine* や *Russland* が極端な位置に現れている点である。*Trump* や *USA* は幅広く用いられるのに対して、*Ukraine* や *Russland* は、ウクライナ情勢といった特定の主題で、非常に顕著に現れたと考えられる。このことは、昨今のニュース報道においてウクライナ情勢が大きな比重を占めている状況とも合致する。なお、第 2 次元において、*Trump* や *USA* に関するニュースは、その他の語群に近い分、ドイツ国内の政治的議論にもつながっている可能性も考えられる。したがって、第 2 次元は、国内ニュースと国際ニュースの対立軸を示すものと考えられるが、国際ニュースといっても、とりわけウクライナ情勢に関する報道が突出していることが示唆される。

以上のコレスポンデンス分析の結果から、両番組における語彙の分布は報道スタイルの差異と、国内ニュースと国際ニュース、とりわけウクライナ情勢の差異という二つの側面によって説明できると考えられる。

3.2 RQ2: 両番組の識別に寄与する語彙指標は何か？

最後に、分類精度の高い機械学習モデルとして知られているランダムフォレストを用いて、両番組の識別に寄与する語彙の特定を行ったところ、表 4 の通り、*tagesschau* が 95.8%、*heute* が 94.2%、全体で 95.0%という極めて高い正解率が得られた。この結果は、両番組の語彙選択の違いを反映するものであり、単なる頻度差を意味するものではない。ドイツ語学習者にとっては、それぞれの番組をどのような目的で視聴するのが望ましいか、その戦略を練る上で、実用的な手がかりになるものと言える。

表 4

ランダムフォレストによる 5 分割交差検証の結果

対象番組 \ 結果	<i>tagesschau</i>	<i>heute</i>	識別率
<i>tagesschau</i>	115	5	95.8%
<i>heute</i>	7	113	94.2%
全体正解率			95.0%

図 3

番組識別に寄与する語彙指標(上位 20 語)

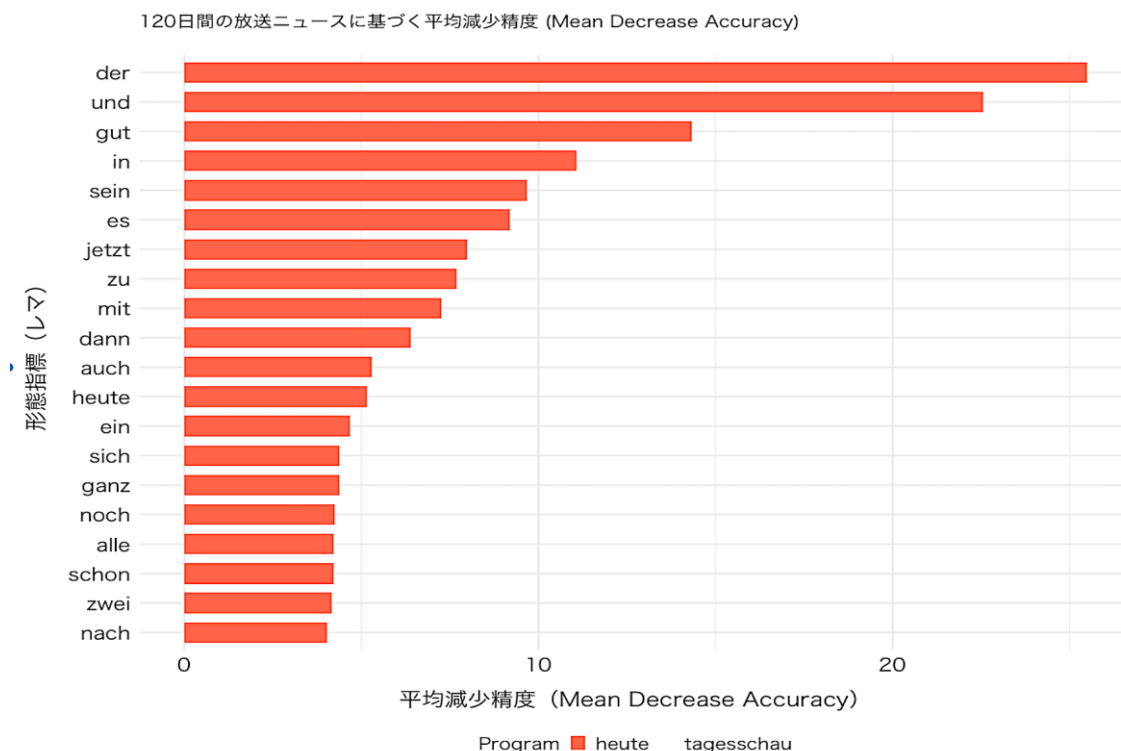


図 3 は、識別に寄与する語彙と評価された上位 20 語とその平均減少精度で、表 5 はその 20 語の出現頻度の平均と比率(%)を表している。表 5 の出現頻度の平均を確認すると、すべての語彙において *heute* が *tagesschau* を上回っており、図 3 においては *tagesschau* に特徴的と評価された語はない。先の表 1 においては、*heute*(2549.19 語)の総語数が *tagesschau*(1767.26 語)のそれと比べて、およそ 1.4 倍多いことを確認したが、とくに、*ganz*(3.57 倍)、*dann*(3.33 倍)、*jetzt*(2.98 倍)、*gut*(2.52 倍)は、それを大幅に上回ることがわかる。なお、コレスポンデンス分析によって *heute* に特徴的であるとされた 20 語(表 2)と、ランダムフォレストによって抽出された 20 語(表 5)を比べると、*gut*、*jetzt*、*dann*、*auch*、*heute*、*ganz*、*alle*、*schon* の 8 語が共通している。これらの 8 語は、たとえば、Regierung(政府)、Wirtschaft(経済)、Gesetz(法律)といったニュースのトピックに関わるような内容語ではない。いずれも基本語彙であり、出来事の展開や状況を説明する際に用いられる機能語であり、種々の談話機能を担う語と言える。

表 5

ランダムフォレストにより識別項目とされた 20 語における出現頻度の平均および比率 (%)

	語	heute	tagesschau	比率 (%)
1	der	245.83	184.09	1.34
2	und	61.47	35.60	1.73
3	gut	7.42	2.94	2.52
4	in	74.83	53.95	1.39
5	sein	61.32	47.33	1.30
6	es	30.61	22.43	1.36
7	jetzt	6.99	2.35	2.98
8	zu	30.47	23.28	1.31
9	mit	22.10	14.58	1.52
10	dann	6.53	1.96	3.33
11	auch	21.15	11.08	1.91
12	heute	9.19	4.57	2.01
13	ein	51.31	38.56	1.33
14	sich	17.40	13.20	1.32
15	ganz	4.11	1.15	3.57
16	noch	10.15	5.47	1.86
17	alle	5.11	3.00	1.70
18	schon	4.26	1.87	2.28
19	zwei	3.25	2.05	1.59
20	nach	11.25	7.62	1.48

3.3 リスニング学習への応用に向けて

以上の結果は、テレビニュースのリスニング学習を行う上においても、いくつかの示唆を与える。

まず、テレビニュースで使用される語彙は、両番組で異なる可能性があり、報道スタイルの違いに留意しながら学習する必要があると言える。とくに *heute* においては、高頻度の基本語彙が特徴的に現れるが、これらの語彙は種々の談話機能を担っており、辞書的意味の理解だけでは対応できるものではなく、出来事の展開や状況を説明する際に使用されるこれらの語彙の談話機能に注意を払うことが求められる。その際、コーパスの利用は効果的で、検索語の前後の文脈を表示する KWIC を通して実際の使用例を確認することも有効であると思われる。

また、実際の使用例に目を向けることで、思いがけない発見もある。たとえば、上記の二つの分析の結果、両番組の識別に寄与する語彙として *gut* が含まれていたが、実例を見ていくと、評価の意味だけでなく、「*gut* + 数字」といった強調の意味でも使用されていることがわかる (*gut 64 Prozent des Stroms* 優に 64 パーセントが)。こうした意味は辞書には記載されているものの、普段から目にするなり、意識しない限りはなかなか注意を向けることが難しい用例でもある。事例を丹念に見ていくことで、分析的観点からテレビニュースのドイツ語の構造を把握することもリスニングのための準備学習に効果的であると思われる。

また、今回は 120 分のテレビニュースの比較ということで、ニュースのトピックに偏りがあったが、テレビニュースの語彙使用が国内ニュースと国際ニュースというトピックの違いを反映している可能

性が高いことを踏まえると、トピック毎の語彙学習の必要性が示唆される。

4. まとめ

本研究では、ドイツの公共放送のテレビニュース番組である *tagesschau* と *heute* の 120 日分の字幕をコーパス化した上で、頻度上位 100 語の分布をコレスポネンシ分析で可視化 (RQ1) し、ランダムフォレストを用いて識別項目の特定 (RQ2) を行った。

その結果、RQ1 (語彙分布の可視化) では、第 1 次元で両番組が対立軸となることが判明し、*tagesschau* はアメリカやトランプ大統領の動向に関する報道が顕著で、話法の助動詞を用いて、主観を排し、事実をより客観的に報じるスタイルをとることがわかった。一方、*heute* では、強調や評価に関わる語彙や一人称複数形など、視聴者の視点で解説や対話を行うスタイルをとることが示唆された。また、第 2 次元では国内ニュースと国際ニュースが対立軸をなし、とりわけウクライナ情勢に関する報道が突出して多いことなどが確認された。

さらに、RQ2 (識別項目の特定) では、95% という極めて高い識別率で両番組の識別に寄与する上位 20 語が特定され、いずれも *heute* に特徴的で、ニュースのトピックに関わる内容語ではなく、機能語であることが判明した。また、そのうちの 8 語は、RQ1 のコレスポネンシ分析の結果においても、特徴的であると評価されるもので、多変量解析と機械学習という異なる手法によっても類似の傾向が示唆された。

なお、本研究では、両番組の 120 日分のデータを元に頻度上位 100 語を分析対象としたが、たとえば、より幅広くトピックの違いを拾い上げるためには、少なくとも 1 年分くらいのデータ量に拡充する必要がある。さらに、語彙使用のみならず、音声、形態、統語、構文など、他の観点から両番組の違いを検証することも重要であると考えられる。筆者の今後の課題としたい。

謝辞

本研究は JSPS 科研費 (JP25K04341) の助成を受けた。

研究インテグリティ宣誓

本稿著者は、本論文が著者自身によるオリジナルの研究成果であって、他者の著作物の剽窃や盗用、データの捏造や改竄、二重投稿、貢献度を反映しない著者名表記といった一切の不適切行為を行っていないことを宣誓する。

引用文献

- Behre, J., Hölig, S., Stöwing, E., & Möller, J. (2025). *Reuters Institute Digital News Report 2025: Ergebnisse für Deutschland*. (Arbeitspapiere des Hans-Bredow-Instituts, 77). Leibniz-Institut für Medienforschung | Hans-Bredow-Institut (HBI).
- Bruns, T., & Marcinkowski, F. (1997). Fallstudie zur Sprache der Fernsehnews.

- von ARD, ZDF, RTL und SAT.1. In O. Jarren (Ed.), *Politische Information im Fernsehen* (Schriftenreihe Medienforschung der Landesanstalt für Rundfunk Nordrhein-Westfalen, Vol. 22, pp. 211–263). VS Verlag für Sozialwissenschaften.
- Council of Europe. (2020). *Common European framework of reference for languages: Learning, teaching, assessment – Companion volume*. Council of Europe Publishing. <https://www.coe.int/en/web/portfolio/self-assessment-grid>
- Lange, B. P., Bayirli, A., & Schwab, F. (2020). Sprachniveau in Online- und TV-Nachrichten: Eine quantitative computergestützte Textanalyse der Online- und TV-Berichterstattung von ARD, ZDF, Sat.1 und RTL. *Studies in Communication Sciences*, 20(2), 135–158.
- Muckenhaupt, M. (1994). Von der Tagesschau zur Infoshow. In H. J. Heringer, G. Samson, M. Kaufmann, & W. Bader (éds.), *Tendenzen der deutschen Gegenwartssprache* (70–105). Presses Sorbonne Nouvelle.
- Wannenmacher, S. (2024). Über gewöhnliche Wörter und ungewöhnliche Dinge: Eine Analyse der Verständlichkeit deutscher TV-Nachrichten. *Journalistik. Zeitschrift für Journalismusforschung*, 73–4, 305–326.