

Zwischen Interpretation und Berechnung

Formale Entscheidbarkeit als Grundlage
erklärbarer Sequenzanalyse

Paul Koop

2026

Zusammenfassung

Die vorliegende Arbeit führt ein formales Entscheidungsverfahren für die Algorithmisch Rekursive Sequenzanalyse (ARS) ein. Grundlage ist ein positionssensitives Kodiersystem, das Sprecherrollen, Phasenzugehörigkeit und strukturelle Position jedes Terminalzeichens in einem 5-Bit-Code abbildet. Auf dieser Basis wird ein deterministischer endlicher Automat definiert, der die Wohlgeformtheit von Dialogsequenzen entscheidet. Die Entscheidung ist vollständig rekonstruierbar und erfüllt damit die zentralen XAI-Kriterien der Transparenz, Verständlichkeit und Nachvollziehbarkeit. Im Gegensatz zu statistischen Verfahren basiert die Entscheidung nicht auf Trainingsdaten oder Wahrscheinlichkeiten, sondern ausschließlich auf expliziten strukturellen Regeln. Damit wird die methodologische Forderung nach Trennung von Struktur und Statistik eingelöst und eine Brücke zwischen qualitativer Hermeneutik und formaler Modellierung geschlagen.

Inhaltsverzeichnis

1	Einleitung: Das Validitätsproblem sequenzieller Analyse	3
2	Das Kodiersystem: Struktur als Code	3
2.1	Anforderungen an ein strukturelles Kodiersystem	3
2.2	Das 5-Bit-Kodiersystem	4
2.3	Kodierung der Terminalzeichen	4
3	Formales Entscheidungsverfahren	4
3.1	Dialogphasen als Zustandsraum	4
3.2	Definition des Automaten	5
3.3	Die Übergangsfunktion	6
3.4	Entscheidbarkeit der Wohlgeformtheit	7
4	Erfüllung der XAI-Kriterien	7
4.1	Transparenz	7
4.2	Rekonstruierbarkeit	7
4.3	Trennung von Struktur und Statistik	8
4.4	Vergleich mit statistischen Verfahren	8
5	Anwendung auf empirische Daten	8
5.1	Die sieben Transkripte	8
5.2	Überführung in die Kodierung	9
5.3	Validierung durch den Automaten	9
6	Diskussion	9
6.1	Methodologische Bedeutung	9
6.2	Verhältnis zur XAI-Diskussion	10
6.3	Grenzen des Verfahrens	10
7	Fazit und Ausblick	11
A	Die sieben Transkripte in kodierter Form	13
A.1	Transkript 1	13
A.2	Transkript 2	13
A.3	Transkript 3	13
A.4	Transkript 4	13
A.5	Transkript 5	13
A.6	Transkript 6	13

A.7	Transkript 7	14
-----	------------------------	----

1 Einleitung: Das Validitätsproblem sequenzieller Analyse

Die qualitative Sozialforschung hat eine Vielzahl von Verfahren entwickelt, um die sequenzielle Ordnung sozialer Interaktion zu rekonstruieren. Objektive Hermeneutik (Oevermann et al., 1979) und Konversationsanalyse (Sacks et al., 1974) teilen die grundlegende Einsicht, dass Bedeutung in Interaktionen nicht punktuell, sondern sequenziell konstituiert wird. Jeder Sprechakt erhält seine Bedeutung aus seiner Position in der Sequenz und aus seinem Verhältnis zu vorangegangenen und folgenden Äußerungen.

Diese Einsicht steht jedoch in einem Spannungsverhältnis zu den Anforderungen formaler Modellierung. Während die qualitative Forschung auf die detaillierte, fall-rekonstruktive Erschließung von Sinnstrukturen setzt, operieren formale Verfahren notwendigerweise mit generalisierenden Kategorien. Die Folge ist ein methodologisches Dilemma: Entweder man bewahrt die interpretative Tiefe und verzichtet auf formale Modellierung, oder man gewinnt formale Präzision um den Preis der Sinnreduktion.

Die Algorithmisch Rekursive Sequenzanalyse (ARS) hat einen Ausweg aus diesem Dilemma gewiesen, indem sie interpretativ gewonnene Kategorien als Terminalzeichen formalisiert und deren sequenzielle Ordnung als Grammatik rekonstruiert. Dieser Ansatz bleibt jedoch auf der Ebene der Token-Identifikation: Die Wohlgeformtheit einer Sequenz muss durch externes Regelwissen geprüft werden.

Der vorliegende Beitrag geht einen Schritt weiter. Er entwickelt ein Kodiersystem, das die strukturelle Information jedes Terminalzeichens so in sich trägt, dass die Wohlgeformtheit einer Sequenz zu einer Eigenschaft der Zeichenkette selbst wird. Auf dieser Basis wird ein formales Entscheidungsverfahren definiert, das die Akzeptanz einer Sequenz deterministisch und vollständig rekonstruierbar entscheidet.

2 Das Kodiersystem: Struktur als Code

2.1 Anforderungen an ein strukturelles Kodiersystem

Ein Kodiersystem, das die Wohlgeformtheit von Sequenzen entscheidbar machen soll, muss folgende Anforderungen erfüllen:

1. **Sprecheridentifikation:** Die Rolle des Sprechers (Kunde/Verkäufer) muss

aus dem Code selbst erkennbar sein.

2. **Phasenzugehörigkeit:** Die Zugehörigkeit zu einer dialogischen Phase (Begrüßung, Bedarf, Abschluss, Verabschiedung) muss kodiert sein.
3. **Positionssensitivität:** Die Position innerhalb der Phase (Eröffnung, Fortführung, Abschluss) muss unterscheidbar sein.
4. **Monotonieprüfung:** Es muss entscheidbar sein, ob die Phasenabfolge regelkonform ist.
5. **Alternierungsprüfung:** Es muss entscheidbar sein, ob die Sprecherrollen korrekt alternieren.

2.2 Das 5-Bit-Kodiersystem

Aus diesen Anforderungen ergibt sich ein 5-stelliges Binärsystem:

$$\underbrace{S}_1 \underbrace{P_1 P_2}_2 \underbrace{U_1 U_2}_2$$

- **Bit 1 (Sprecher):** 0 = Kunde (K), 1 = Verkäufer (V)
- **Bits 2-3 (Hauptphase):** 00 = Begrüßung (BG), 01 = Bedarfsteil (B), 10 = Abschlussteil (A), 11 = Verabschiedung (AV)
- **Bits 4-5 (Unterphase):** 00 = Basisebene, 01 = Folgeebene

2.3 Kodierung der Terminalzeichen

Aus diesem System ergeben sich folgende Kodierungen:

3 Formales Entscheidungsverfahren

3.1 Dialogphasen als Zustandsraum

Die dialogische Struktur wird durch einen endlichen Zustandsraum abgebildet:

$$Q = \{q_0, q_{BG}, q_B, q_A, q_{AV}, q_\perp\}$$

- q_0 : Startzustand (leere Sequenz)

Tabelle 1: Kodierung der Terminalzeichen

Symbol	Bedeutung	Code	Interpretation
KBG	Kunden-Gruß	00000	Kunde, BG, Basis
VBG	Verkäufer-Gruß	10000	Verkäufer, BG, Basis
KBBd	Kunden-Bedarf	00100	Kunde, B, Basis
VBBd	Verkäufer-Nachfrage	10100	Verkäufer, B, Basis
KBA	Kunden-Antwort	00101	Kunde, B, Folge
VBA	Verkäufer-Reaktion	10101	Verkäufer, B, Folge
KAE	Kunden-Erkundigung	01000	Kunde, A, Basis
VAE	Verkäufer-Auskunft	11000	Verkäufer, A, Basis
KAA	Kunden-Abschluss	01001	Kunde, A, Folge
VAA	Verkäufer-Abschluss	11001	Verkäufer, A, Folge
KAV	Kunden-Verabschiedung	01100	Kunde, AV, Basis
VAV	Verkäufer-Verabschiedung	11100	Verkäufer, AV, Basis

- q_{BG} : Begrüßungsphase
- q_B : Bedarfsteil
- q_A : Abschlussteil
- q_{AV} : Verabschiedung
- q_{\perp} : Fehlerzustand

Die Menge der akzeptierenden Zustände ist:

$$F = \{q_{AV}\}$$

Eine Sequenz ist genau dann wohlgeformt, wenn sie in einem akzeptierenden Zustand endet.

3.2 Definition des Automaten

Wir definieren einen deterministischen endlichen Automaten

$$\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$$

mit:

- Q : Zustandsmenge

- $\Sigma \subseteq \{0, 1\}^5$: Terminalalphabet
- $\delta : Q \times \Sigma \rightarrow Q$: Übergangsfunktion
- q_0 : Startzustand
- F : akzeptierende Zustände

3.3 Die Übergangsfunktion

Die Übergangsfunktion δ realisiert folgende Regeln:

Begrüßungsphase:

$$\delta(q_0, 00000) = q_{BG} \quad (\text{KBG})$$

$$\delta(q_{BG}, 10000) = q_{BG} \quad (\text{VBG})$$

Bedarfsteil:

$$\delta(q_{BG}, 00100) = q_B \quad (\text{KBBd})$$

$$\delta(q_B, 10100) = q_B \quad (\text{VBBd})$$

$$\delta(q_B, 00101) = q_B \quad (\text{KBA})$$

$$\delta(q_B, 10101) = q_B \quad (\text{VBA})$$

Abschlusssteil:

$$\delta(q_B, 01000) = q_A \quad (\text{KAE})$$

$$\delta(q_A, 11000) = q_A \quad (\text{VAE})$$

$$\delta(q_A, 01001) = q_{AV} \quad (\text{KAA})$$

$$\delta(q_{AV}, 11001) = q_{AV} \quad (\text{VAA})$$

Verabschiedung:

$$\delta(q_{AV}, 01100) = q_{AV} \quad (\text{KAV})$$

$$\delta(q_{AV}, 11100) = q_{AV} \quad (\text{VAV})$$

Fehlerfälle: Alle nicht definierten Übergänge führen in den Fehlerzustand:

$$\delta(q, \sigma) = q_{\perp} \quad \text{falls keine Regel definiert}$$

3.4 Entscheidbarkeit der Wohlgeformtheit

Satz 1 (Entscheidbarkeit): Das Wohlgeformtheitsproblem ist für den Automaten \mathcal{A} entscheidbar.

Beweis: Der Automat \mathcal{A} ist endlich, deterministisch und vollständig definiert. Für jede Eingabe $w = \sigma_1 \dots \sigma_n \in \Sigma^*$ existiert genau ein Lauf

$$q_0 \xrightarrow{\sigma_1} q_1 \xrightarrow{\sigma_2} \dots \xrightarrow{\sigma_n} q_n.$$

Da Q endlich ist, ist dieser Lauf endlich berechenbar. w ist genau dann wohlgeformt, wenn $q_n \in F$. Damit ist das Problem entscheidbar. \square

4 Erfüllung der XAI-Kriterien

4.1 Transparenz

Die Entscheidung des Automaten ist vollständig transparent:

- Die Zustandsmenge Q ist explizit angegeben.
- Die Übergangsfunktion δ ist vollständig definiert.
- Jeder Schritt im Lauf ist dokumentierbar.

Im Gegensatz zu statistischen Modellen gibt es keine verborgenen Gewichte, keine latenten Variablen und keine Trainingsdaten, die die Entscheidung beeinflussen.

4.2 Rekonstruierbarkeit

Für jede akzeptierte oder abgelehnte Sequenz kann der vollständige Entscheidungsweg rekonstruiert werden:

$$q_0 \xrightarrow{\sigma_1} q_1 \xrightarrow{\sigma_2} \dots \xrightarrow{\sigma_n} q_n$$

Jeder Übergang ist durch die Definition von δ begründet. Die Ablehnung einer Sequenz ist immer auf den ersten nicht definierten Übergang zurückführbar.

4.3 Trennung von Struktur und Statistik

Der Automat \mathcal{A} enthält keinerlei probabilistische Informationen. Seine Entscheidungen sind:

- **deterministisch:** gleiche Eingabe \rightarrow gleiche Ausgabe
- **kontextfrei:** unabhängig von empirischen Häufigkeiten
- **strukturerhaltend:** abgeleitet aus der Grammatik

Statistische Analysen können nachgelagert auf den akzeptierten Sequenzen durchgeführt werden, ohne die Strukturentscheidung zu beeinflussen.

4.4 Vergleich mit statistischen Verfahren

Tabelle 2: Vergleich mit statistischen Verfahren

Kriterium	Statistische Verfahren	Automat \mathcal{A}
Entscheidungsgrundlagen	Tagungsdaten, Gewichte	Explizite Regeln
Transparenz	Gering (Black Box)	Vollständig
Rekonstruierbarkeit	Approximativ	Exakt
Datenabhängigkeit	Hoch	Keine
Erklärbarkeit	Post-hoc	Ad-hoc

5 Anwendung auf empirische Daten

5.1 Die sieben Transkripte

Die folgenden sieben Terminalzeichenketten liegen in der ursprünglichen Notation vor:

- 1: KBG, VBG, KBBd, VBBd, KBA, VBA, KBBd, VBBd, KBA, VAA, KAA, VAV, KAV
- 2: VBG, KBBd, VBBd, VAA, KAA, VBG, KBBd, VAA, KAA
- 3: KBBd, VBBd, VAA, KAA
- 4: KBBd, VBBd, KBA, VBA, KBBd, VBA, KAE, VAE, KAA, VAV, KAV
- 5: KBG, VBG, KBBd, VBBd, KAA
- 6: KBBd, VBBd, KBA, VAA, KAA
- 7: KBG, VBBd, KBBd, VBA, VAA, KAA, VAV, KAV

5.2 Überführung in die Kodierung

Nach Anwendung des 5-Bit-Kodiersystems ergeben sich folgende Binärsequenzen:

```

1 1:
    00000,10000,00100,10100,00101,10101,00100,10100,00101,11001,01001,11100

2 2: 10000,00100,10100,11001,01001,10000,00100,11001,01001

3 3: 00100,10100,11001,01001

4 4:
    00100,10100,00101,10101,00100,10101,01000,11000,01001,11100,01100

5 5: 00000,10000,00100,10100,01001

6 6: 00100,10100,00101,11001,01001

7 7: 00000,10100,00100,10101,11001,01001,11100,01100

```

Listing 1: Kodierte Terminalzeichenketten

5.3 Validierung durch den Automaten

Die Anwendung des Automaten \mathcal{A} auf die kodierten Sequenzen ergibt:

Tabelle 3: Validierungsergebnisse

Transkript	Letzter Zustand	Wohlgeformt
1	q_{AV}	
2	q_{AV}	
3	q_{AV}	
4	q_{AV}	
5	q_{AV}	
6	q_{AV}	
7	q_{AV}	

Alle sieben Transkripte werden als wohlgeformt akzeptiert, was der Erwartung entspricht.

6 Diskussion

6.1 Methodologische Bedeutung

Das vorgestellte Verfahren löst ein zentrales methodologisches Problem der qualitativen Sequenzanalyse: Die Validität einer Interpretation wird nicht mehr durch

externe Kriterien oder statistische Plausibilität begründet, sondern durch formale Entscheidbarkeit. Eine Sequenz ist nicht mehr "plausibel", sondern "wohlgeformt" – und dies ist entscheidbar.

Dies entspricht der in der objektiven Hermeneutik formulierten Forderung nach strikter Regelgeleitetheit sozialer Interaktion (Oevermann et al., 1979, S. 372). Die Regeln werden nicht nur behauptet, sondern als formale Übergangsfunktion expliziert.

6.2 Verhältnis zur XAI-Diskussion

Die Explainable AI (XAI) hat die Forderung nach Transparenz und Rekonstruierbarkeit technischer Systeme formuliert (Samek & Müller, 2019; Barredo Arrieta et al., 2020). Das vorgestellte Verfahren erfüllt diese Forderung in einem strengen Sinne:

- **Verständlichkeit:** Die Zustände und Übergänge sind semantisch interpretierbar.
- **Genauigkeit:** Die Entscheidung folgt exakt den definierten Regeln.
- **Wissensgrenzen:** Die Grenzen des Verfahrens sind mit der Zustandsmenge Q explizit gegeben.

Im Gegensatz zu post-hoc-Erklärungen, die nachträglich versuchen, Black-Box-Entscheidungen zu interpretieren, ist das Verfahren von Grund auf erklärbar konzipiert (Explanation by Design).

6.3 Grenzen des Verfahrens

Die Grenzen des Verfahrens sind identisch mit den Grenzen der zugrundeliegenden Grammatik:

- Das Verfahren erfasst nur die vorgesehenen Phasen und Übergänge.
- Komplexere Interaktionsmuster (Unterbrechungen, Parallelität) erfordern eine Erweiterung des Zustandsraums.
- Die Kodierung ist auf das binäre System beschränkt; feinere Differenzierungen erfordern mehr Bits.

7 Fazit und Ausblick

Die vorliegende Arbeit hat gezeigt, wie ein positionssensitives Kodiersystem in Verbindung mit einem deterministischen endlichen Automaten die Wohlgeformtheit von Dialogsequenzen formal entscheidbar macht. Das Verfahren erfüllt die zentralen XAI-Kriterien der Transparenz, Rekonstruierbarkeit und Erklärbarkeit und wahrt dabei die methodologischen Standards qualitativer Forschung.

Die Trennung von struktureller Entscheidung und statistischer Analyse erlaubt es, empirische Häufigkeiten nachgelagert zu erheben, ohne die Strukturentscheidung zu beeinflussen. Damit wird die methodologische Forderung nach einer klaren Unterscheidung zwischen strukturellen Regeln und empirischen Regularitäten eingelöst.

Weiterführende Forschung könnte:

1. Das Verfahren auf komplexere Interaktionstypen erweitern (Mehrpersoneninteraktionen, Unterbrechungen).
2. Die Kodierung um weitere Dimensionen ergänzen (emotionale Tönung, prosodische Merkmale).
3. Das Zusammenspiel mit statistischen Verfahren systematisch untersuchen (PCFG auf den kodierten Sequenzen).

Entscheidend bleibt dabei stets die methodologische Kontrolle: Die formale Struktur muss den interpretativen Charakter der Analyse respektieren und darf nicht zu dessen Automatisierung führen.

Literatur

- Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barredo, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82-115.
- Flick, U. (2019). *Qualitative Sozialforschung: Eine Einführung* (9. Aufl.). Rowohlt.
- Oevermann, U., Allert, T., Konau, E., & Krambeck, J. (1979). Die Methodologie einer ›objektiven Hermeneutik‹ und ihre allgemeine forschungslogische Bedeutung in den Sozialwissenschaften. In H.-G. Soeffner (Hrsg.), *Interpretative Verfahren in den Sozial- und Textwissenschaften* (S. 352-434). Metzler.
- Przyborski, A., & Wohlrab-Sahr, M. (2021). *Qualitative Sozialforschung: Ein Arbeitsbuch* (5. Aufl.). De Gruyter Oldenbourg.
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4), 696-735.
- Samek, W., & Müller, K.-R. (2019). Towards Explainable Artificial Intelligence. In W. Samek, G. Montavon, A. Vedaldi, L. K. Hansen, & K.-R. Müller (Hrsg.), *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning* (S. 1-10). Springer.

A Die sieben Transkripte in kodierter Form

A.1 Transkript 1

Original: KBG, VBG, KBBd, VBBd, KBA, VBA, KBBd, VBBd, KBA, VAA, KAA, VAV, KAV

Kodiert: 00000, 10000, 00100, 10100, 00101, 10101, 00100, 10100, 00101, 11001, 01001, 11100, 01100

A.2 Transkript 2

Original: VBG, KBBd, VBBd, VAA, KAA, VBG, KBBd, VAA, KAA

Kodiert: 10000, 00100, 10100, 11001, 01001, 10000, 00100, 11001, 01001

A.3 Transkript 3

Original: KBBd, VBBd, VAA, KAA

Kodiert: 00100, 10100, 11001, 01001

A.4 Transkript 4

Original: KBBd, VBBd, KBA, VBA, KBBd, VBA, KAE, VAE, KAA, VAV, KAV

Kodiert: 00100, 10100, 00101, 10101, 00100, 10101, 01000, 11000, 01001, 11100, 01100

A.5 Transkript 5

Original: KBG, VBG, KBBd, VBBd, KAA

Kodiert: 00000, 10000, 00100, 10100, 01001

A.6 Transkript 6

Original: KBBd, VBBd, KBA, VAA, KAA

Kodiert: 00100, 10100, 00101, 11001, 01001

A.7 Transkript 7

Original: KBG, VBBd, KBBd, VBA, VAA, KAA, VAV, KAV

Kodiert: 00000, 10100, 00100, 10101, 11001, 01001, 11100, 01100