

November 17, 2024

1 Algorithmisch Rekursive Sequenzanalyse 2.0

1.1 Grammatikinduktion aus weiteren Transkripten nach extensiver Sequenzanalyse eines Transkriptes

In diesem Dokument wird die Methode der algorithmisch rekursiven Sequenzanalyse vorgestellt. Das Ziel dieser Methode ist es, grammatikalische Strukturen aus natürlichen Sprachsequenzen zu extrahieren und darauf aufbauend eine Grammatik zu induzieren. Wir fokussieren uns auf die Analyse von Transkripten, die durch umfangreiche Sequenzanalysen vorbereitet wurden, um die zugrundeliegenden Regeln und Muster in den Daten zu erkennen.

1.1.1 Hintergrund

Die Sequenzanalyse in natürlichen Sprachdaten ist ein weit verbreiteter Ansatz in der Sprachverarbeitung, um wiederkehrende Muster, Strukturen und Abhängigkeiten zwischen verschiedenen Teilen einer Sequenz zu identifizieren. In dieser erweiterten Version (2.0) integrieren wir die grammatikalische Induktion als eine Möglichkeit, aus den analysierten Sequenzen eine formale Grammatik abzuleiten.

1.1.2 Zielsetzung

Die Hauptziele dieses Ansatzes umfassen:

1. **Erweiterung der Sequenzanalyse:** Die Analyse wird auf eine breitere Menge an Transkripten angewendet, um zu prüfen, ob die entdeckten Muster über ein einzelnes Transkript hinaus generalisiert werden können.
2. **Grammatikinduktion:** Mit den gewonnenen Erkenntnissen aus den Sequenzen sollen grammatikalische Regeln extrahiert und formalisierte Grammatiken entwickelt werden.
3. **Evaluation der Qualität:** Die Qualität der induzierten Grammatik wird anhand ihrer Fähigkeit bewertet, zukünftige Sequenzen zu generieren, die den Mustern in den Transkripten entsprechen.

1.1.3 Methodik

Die Methode besteht aus mehreren Schritten, die iterativ durchgeführt werden:

1. **Datenvorbereitung:** Sammlung und Preprocessing der Transkripte.
2. **Sequenzanalyse:** Durchführung einer extensiven Sequenzanalyse mit Lesartenproduktion und Lesartenfalsifikation an einem Transkript (a.a.O in diesem Repository). Durchführung

einer umfangreichen Analyse der Transkripte, bei der statistische und syntaktische Muster identifiziert werden.

3. **Grammatikinduktion:** Auf Basis der identifizierten Muster wird eine formale Grammatik (Python hier im Dokument) induziert, die die generierten Sequenzen beschreiben kann.
4. **Optimierung:** Die Grammatik wird optimiert (Python hier im Dokument) und validiert, um ihre Genauigkeit und Generalisierbarkeit zu erhöhen.

Nach der Auswertung von Fachliteratur zu Verkaufsgesprächen ergibt sich als vorläufige hypothetische Grammatik: Ein Verkaufsgespräch (VKG) besteht aus einer Begrüßung (BG), einem Verkaufsteil (VT) und einer Verabschiedung (AV).

- Die Begrüßung besteht aus einem Gruß des Kunden (KBG) und einem Gruß des Verkäufers (VBG).
- Die Verabschiedung besteht aus einer Verabschiedung durch den Kunden (KAV) und einer Verabschiedung durch den Verkäufer (VAV).
- Der Verkaufsteil besteht aus einem Bedarfsteil (B) und einem Abschlussteil (A).
 - Der Bedarfsteil umfasst eine Bedarfsklärung (BBd) und eine Bedarfsargumentation (BA).
 - * Die Bedarfsklärung besteht aus Bedarfen des Kunden (KBBd) und den dazugehörigen Klärungen des Verkäufers (VBBd).
 - * Die Bedarfsargumentation setzt sich aus Argumenten des Kunden (KBA) und des Verkäufers (VBA) zusammen.
 - Der Abschlussteil (A) besteht aus Einwänden (AE) und einem Verkaufsabschluss (AA).
 - * Die Einwände setzen sich aus Argumenten des Kunden (KAE) und des Verkäufers (VAE) zusammen.
 - * Der Verkaufsabschluss besteht aus Argumenten des Kunden (KAA) und des Verkäufers (VAA).

Die Verabschiedung (AV) besteht aus einem Abschied des Kunden (KAV) und des Verkäufers (VAV).

Das Startsymbol ist demnach VKG, und die Terminalzeichen für die Kategorien sind: KBG, VBG, KBBd, VBBd, KBA, VBA, KAE, VAE, KAA, VAA, KAV und VAV. ERstelle aus diesen Angaben eine probabilistische kontextfreie Grammatik mit zunächst vermuteten Übergangswahrscheinlichkeiten.

Zunächst wurde dann ein Transkript einer intensiven Sequenzanalyse mit Lesartenproduktion und Lesartenfalsifikation, Kategorienbildung, Interkodierkorrelation und Grammatikinduktion mit Schem und Grammatiktransduktion mit Lisp und Parsen der Terminalkette mit Object Pascal durchgeführt (a.a.O. in diesem Repository).

Das Vorgehen lässt sich in mehrere klar definierte Schritte unterteilen. Hier ist eine Übersicht, um sicherzustellen, dass alle Schritte verständlich sind:

1.1.4 1. Hypothesenbildung zur Grammatik

- Durchführung einer Voruntersuchung, um eine **hypothetische PCFG** zu erstellen. Diese dient als Orientierung, ist aber nicht verbindlich für die Analyse der Transkripte.

1.1.5 2. Analyse der Transkripte

- Jedes **Transkript** wird **einzeln** analysiert. Die Intentionen der Sprecher und die Struktur der Interaktionen werden identifiziert.
- Jeder Interaktion im Transkript wird ein **Terminalzeichen** zugeordnet. Dies ergibt für jedes Transkript eine **Terminalzeichenkette**.

1.1.6 3. Induktion einer Grammatik für jedes Transkript

- Aus der Terminalzeichenkette eines jeden Transkripts wird eine spezifische **Grammatik induziert**.
- Es gibt am Ende eine eigene **Grammatik pro Transkript** (insgesamt acht).

1.1.7 4. Vereinigung der Grammatiken

- Die acht individuellen Grammatiken werden zu einer **vereinigten Grammatik** zusammengeführt, die die Struktur aller Transkripte abdeckt.

1.1.8 5. Parsing der Terminalzeichenketten

- Die acht Terminalzeichenketten der Transkripte werden auf **Wohlgeformtheit** in Bezug auf die vereinigte Grammatik geprüft. Dies bedeutet, dass jede Zeichenkette durch die Grammatik korrekt erkannt und geparkt werden sollte.

1.1.9 6. Optimierung der PCFG

- Eine erste Version einer PCFG wird aus den vorherigen Analysen erstellt.
- Mit der erstellten PCFG werden **künstliche Terminalzeichenketten** erzeugt.
- Die **anteilige Häufigkeit der Terminalzeichen** aus den künstlichen Zeichenketten wird mit der Häufigkeit der Terminalzeichen in den realen Daten (den acht Transkripten) verglichen.
- Die **Signifikanz** dieser Korrelation wird gemessen.
- Anpassungen an der PCFG werden vorgenommen, bis die Korrelation zufriedenstellend ist.

1.1.10 7. Wiederholungsprozess

- Schritte 6 und 7 werden so lange wiederholt, bis eine gute Anpassung zwischen den **empirischen Daten** und der erzeugten PCFG besteht.

Hier ist eine probabilistische kontextfreie Grammatik (PCFG), die auf der Struktur basiert. Bei einer PCFG werden Wahrscheinlichkeiten für die Übergänge zwischen verschiedenen Produktionsregeln definiert. Da du nur eine Struktur ohne spezifische Wahrscheinlichkeiten angegeben hast, werde ich annahmeweise gleiche Wahrscheinlichkeiten für die möglichen Optionen innerhalb der gleichen Ebene verwenden, um eine Grundlage zu schaffen. Diese Wahrscheinlichkeiten lassen sich später anpassen, wenn empirische Daten verfügbar sind.

2 Probabilistische hypothetische kontextfreie Grammatik (PCFG) für ein Verkaufsgespräch

VKG -> BG VT AV [1.0]

3 Begrüßung

BG -> KBG VBG [1.0] KBG -> ‘Kunden-Gruß’ VBG -> ‘Verkäufer-Gruß’

4 Verabschiedung

AV -> KAV VAV [1.0] KAV -> ‘Kunden-Verabschiedung’ VAV -> ‘Verkäufer-Verabschiedung’

5 Verkaufsteil

VT -> B A [1.0]

6 Bedarfsteil

B -> BBd BA [1.0]

7 Bedarfsklärung

BBd -> KBBd VBBd [1.0] KBBd -> ‘Kunden-Bedarf’ VBBd -> ‘Verkäufer-Klä rung’

8 Bedarfsargumentation

BA -> KBA VBA [0.5] | VBA KBA [0.5] KBA -> ‘Kunden-Argument’ VBA -> ‘Verkäufer-Argument’

9 Abschlussteil

A -> AE AA [1.0]

10 Einwände

AE -> KAE VAE [0.5] | VAE KAE [0.5] KAE -> ‘Kunden-Einwand’ VAE -> ‘Verkäufer-Einwand’

11 Verkaufsabschluss

AA -> KAA VAA [0.5] | VAA KAA [0.5] KAA -> ‘Kunden-Abschluss’ VAA -> ‘Verkäufer-Abschluss’

Es folgen die acht Transkripte. (Tonbandprotokolle a.a.O. in diesem Repository.

11.0.1 Text 1

Datum: 28. Juni 1994, **Ort:** Metzgerei, Aachen, 11:00 Uhr

(Die Geräusche eines geschäftigen Marktplatzes im Hintergrund, Stimmen und Gemurmel)

Verkäuferin: Guten Tag, was darf es sein?

Kunde: Einmal von der groben Leberwurst, bitte.

Verkäuferin: Wie viel darf's denn sein?

Kunde: Zwei hundert Gramm.

Verkäuferin: Zwei hundert Gramm. Sonst noch etwas?

Kunde: Ja, dann noch ein Stück von dem Schwarzwälder Schinken.

Verkäuferin: Wie groß soll das Stück sein?

Kunde: So um die dreihundert Gramm.

Verkäuferin: Alles klar. Kommt sofort. *(Geräusche von Papier und Verpackung)*

Kunde: Danke schön.

Verkäuferin: Das macht dann acht Mark zwanzig.

Kunde: Bitte. *(Klimpern von Münzen, Geräusche der Kasse)*

Verkäuferin: Danke und einen schönen Tag noch!

Kunde: Danke, ebenfalls!

Ende Text 1

11.0.2 Text 2

Datum: 28. Juni 1994, **Ort:** Marktplatz, Aachen

(Ständige Hintergrundgeräusche von Stimmen und Marktatmosphäre)

Verkäufer: Kirschen kann jeder probieren hier, Kirschen kann jeder probieren hier!

Kunde 1: Ein halbes Kilo Kirschen, bitte.

Verkäufer: Ein halbes Kilo? Oder ein Kilo?

(Unverständliches Gespräch, Münzen klimpern)

Verkäufer: Danke schön!

Verkäufer: Kirschen kann jeder probieren hier! Drei Mark, bitte.

Kunde 1: Danke schön!

Verkäufer: Kirschen kann jeder probieren hier, Kirschen kann jeder probieren hier!

(Weitere Stimmen im Hintergrund, unverständliches Gespräch, Münzen klimpern)

Kunde 2: Ein halbes Kilo, bitte.

(Unverständliches Gespräch)

Ende Text 2

11.0.3 Text 3

Datum: 28. Juni 1994, **Ort:** Fischstand, Marktplatz, Aachen

(Marktatmosphäre, Gespräch im Hintergrund, teilweise unverständlich)

Kunde: Ein Pfund Seelachs, bitte.

Verkäufer: Seelachs, alles klar.

(Geräusche von Verpackung und Verkaufsvorbereitungen)

Verkäufer: Vier Mark neunzehn, bitte.

(Geräusche von Verpackung, Münzen klappern)

Verkäufer: Schönen Dank!

Kunde: Ja, danke schön!

Ende Text 3

11.0.4 Text 4

Datum: 28. Juni 1994, **Ort:** Gemüsestand, Aachen, Marktplatz, 11:00 Uhr

(Marktatmosphäre, teilweise unverständlich)

Kunde: Hören Sie, ich nehme ein paar Champignons mit.

Verkäufer: Braune oder helle?

Kunde: Nehmen wir die hellen.

Verkäufer: Alles klar, die hellen.

(Unverständliche Unterhaltung im Hintergrund)

Verkäufer: Die sind beide frisch, keine Sorge.

Kunde: Wie ist es mit Pfifferlingen?

Verkäufer: Ah, die sind super!

(Unverständliches Gespräch)

Kunde: Kann ich die in Reissalat tun?

Verkäufer: Eher kurz anbraten in der Pfanne.

Kunde: Okay, mache ich.

Verkäufer: Die können Sie roh verwenden, aber ein bisschen anbraten ist besser.

Kunde: Verstanden.

(Weitere Unterhaltung, unverständliche Kommentare)

Verkäufer: Noch etwas anderes?

Kunde: Ja, dann nehme ich noch Erdbeeren.

(Pause, Hintergrundgeräusche von Verpackung und Stimmen)

Verkäufer: Schönen Tag noch!

Kunde: Gleichfalls!

Ende Text 4

11.0.5 Text 5

Datum: 26. Juni 1994, **Ort:** Gemüsestand, Aachen, Marktplatz, 11:00 Uhr

(Marktatmosphäre, teilweise unverständlich)

Verkäufer: So, bitte schön.

Kunde 1: Auf Wiedersehen!

Kunde 2: Ich hätte gern ein Kilo von den Granny Smith Äpfeln hier.

(Unverständliches Gespräch im Hintergrund)

Verkäufer: Sonst noch etwas?

Kunde 2: Ja, noch ein Kilo Zwiebeln.

Verkäufer: Alles klar.

(Unverständliches Gespräch, Hintergrundgeräusche)

Kunde 2: Das war's.

Verkäufer: Sechs Mark fünfundzwanzig, bitte.

(Unverständliches Gespräch, Geräusche von Münzen und Verpackung)

Verkäufer: Wiedersehen!

Kunde 2: Wiedersehen!

Ende Text 5

11.0.6 Text 6

Datum: 28. Juni 1994, **Ort:** Käseverkaufsstand, Aachen, Marktplatz

(Marktatmosphäre, Begrüßungen)

Kunde 1: Guten Morgen!

Verkäufer: Guten Morgen!

Kunde 1: Ich hätte gerne fünfhundert Gramm holländischen Gouda.

Verkäufer: Am Stück?

Kunde 1: Ja, am Stück, bitte.

Ende Text 6

11.0.7 Text 7

Datum: 28. Juni 1994, **Ort:** Bonbonstand, Aachen, Marktplatz, 11:30 Uhr

(Geräusche von Stimmen und Marktatmosphäre, teilweise unverständlich)

Kunde: Von den gemischten hätte ich gerne hundert Gramm.

(Unverständliche Fragen und Antworten)

Verkäufer: Für zu Hause oder zum Mitnehmen?

Kunde: Zum Mitnehmen, bitte.

Verkäufer: Fünfzig Pfennig, bitte.

(Klimpern von Münzen, Geräusche von Verpackung)

Kunde: Danke!

Ende Text 7

11.0.8 Text 8

Datum: 9. Juli 1994, **Ort:** Bäckerei, Aachen, 12:00 Uhr

(Schritte hörbar, Hintergrundgeräusche, teilweise unverständlich)

Kunde: Guten Tag!

(Unverständliche Begrüßung im Hintergrund)

Verkäuferin: Einmal unser bester Kaffee, frisch gemahlen, bitte.

(Geräusche der Kaffeemühle, Verpackungsgeräusche)

Verkäuferin: Sonst noch etwas?

Kunde: Ja, noch zwei Stück Obstsalat und ein Schälchen Sahne.

Verkäuferin: In Ordnung!

(Geräusche der Kaffeemühle, Papiergeräusche)

Verkäuferin: Ein kleines Schälchen Sahne, ja?

Kunde: Ja, danke.

(Türgeräusch, Lachen, Papiergeräusche)

Verkäuferin: Keiner kümmert sich darum, die Türen zu ölen.

Kunde: Ja, das ist immer so.

(Lachen, Geräusche von Münzen und Verpackung)

Verkäuferin: Das macht vierzehn Mark und neunzehn Pfennig, bitte.

Kunde: Ich zahle in Kleingeld.

(Lachen und Geräusche von Münzen)

Verkäuferin: Vielen Dank, schönen Sonntag noch!

Kunde: Danke, Ihnen auch!

Ende Text 8

Auf Grundlage der vorläufigen probabilistischen kontextfreien Grammatik (PCFG) werden die Transkripte in die entsprechenden Terminalzeichen umwandeln. Hier ist die detaillierte Zuordnung für jedes Transkript, wobei für jede relevante Aktion oder Aussage des Gesprächs ein passendes Terminalzeichen aus der vorgegebenen Grammatik verwende.

11.0.9 Transkript 1 - Terminalzeichen

Datum: 28. Juni 1994, **Ort:** Metzgerei, Aachen, 11:00 Uhr

Transkriptausschnitt	Terminalzeichen
Kunde: Guten Tag	KBG (Kunden-Gruß)
Verkäuferin: Guten Tag	VBG (Verkäufer-Gruß)
Kunde: Einmal von der groben Leberwurst, bitte.	KBBd (Kunden-Bedarf)
Verkäuferin: Wie viel darf's denn sein?	VBBd (Verkäufer-Klärung)
Kunde: Zwei hundert Gramm.	KBA (Kunden-Argument)
Verkäuferin: Sonst noch etwas?	VBA (Verkäufer-Argument)
Kunde: Ja, dann noch ein Stück von dem Schwarzwäldle	KBBd (Kunden-Bedarf)
Verkäuferin: Wie groß soll das Stück sein?	VBBd (Verkäufer-Klärung)
Kunde: So um die dreihundert Gramm.	KBA (Kunden-Argument)
Verkäuferin: Das macht dann acht Mark zwanzig.	VAA (Verkäufer-Abschluss)
Kunde: Bitte.	KAA (Kunden-Abschluss)
Verkäuferin: Danke und einen schönen Tag noch!	VAV (Verkäufer-Verabschiedung)
Kunde: Danke, ebenfalls!	KAV (Kunden-Verabschiedung)

11.0.10 Transkript 2 - Terminalzeichen

Datum: 28. Juni 1994, **Ort:** Marktplatz, Aachen

Transkriptausschnitt	Terminalzeichen
Verkäufer: Kirschen kann jeder probieren hier!	VBG (Verkäufer-Gruß)
Kunde 1: Ein halbes Kilo Kirschen, bitte.	KBBd (Kunden-Bedarf)
Verkäufer: Ein halbes Kilo? Oder ein Kilo?	VBBd (Verkäufer-Klärung)
Verkäufer: Drei Mark, bitte.	VAA (Verkäufer-Abschluss)
Kunde 1: Danke schön!	KAA (Kunden-Abschluss)
Verkäufer: Kirschen kann jeder probieren hier!	VBG (Verkäufer-Gruß)
Kunde 2: Ein halbes Kilo, bitte.	KBBd (Kunden-Bedarf)
Verkäufer: Drei Mark, bitte.	VAA (Verkäufer-Abschluss)
Kunde 2: Danke schön!	KAA (Kunden-Abschluss)

11.0.11 Transkript 3 - Terminalzeichen

Datum: 28. Juni 1994, **Ort:** Fischstand, Marktplatz, Aachen

Transkriptausschnitt	Terminalzeichen
Kunde: Ein Pfund Seelachs, bitte.	KBBd (Kunden-Bedarf)
Verkäufer: Seelachs, alles klar.	VBBd (Verkäufer-Klärung)
Verkäufer: Vier Mark neunzehn, bitte.	VAA (Verkäufer-Abschluss)
Kunde: Danke schön!	KAA (Kunden-Abschluss)

11.0.12 Transkript 4 - Terminalzeichen

Datum: 28. Juni 1994, **Ort:** Gemüsestand, Aachen, Marktplatz, 11:00 Uhr

Transkriptausschnitt	Terminalzeichen
Kunde: Hören Sie, ich nehme ein paar Champignons mit.	KBBd (Kunden-Bedarf)
Verkäufer: Braune oder helle?	VBBd (Verkäufer-Klärung)
Kunde: Nehmen wir die hellen.	KBA (Kunden-Argument)
Verkäufer: Die sind beide frisch, keine Sorge.	VBA (Verkäufer-Argument)
Kunde: Wie ist es mit Pfifferlingen?	KBBd (Kunden-Bedarf)
Verkäufer: Ah, die sind super!	VBA (Verkäufer-Argument)
Kunde: Kann ich die in Reissalat tun?	KAE (Kunden-Einwand)
Verkäufer: Eher kurz anbraten in der Pfanne.	VAE (Verkäufer-Einwand)

Transkriptausschnitt	Terminalzeichen
Kunde: Okay, mache ich.	KA (Kunden-Abschluss)
Verkäufer: Schönen Tag noch!	VAV (Verkäufer-Verabschiedung)
Kunde: Gleichfalls!	KAV (Kunden-Verabschiedung)

11.0.13 Transkript 5 - Terminalzeichen

Datum: 26. Juni 1994, **Ort:** Gemüsestand, Aachen, Marktplatz, 11:00 Uhr

Transkriptausschnitt	Terminalzeichen
Kunde 1: Auf Wiedersehen!	KAV (Kunden-Verabschiedung)
Kunde 2: Ich hätte gern ein Kilo von den Granny Smi	KBBd (Kunden-Bedarf)
Verkäufer: Sonst noch etwas?	VBBd (Verkäufer-Klärung)
Kunde 2: Ja, noch ein Kilo Zwiebeln.	KBBd (Kunden-Bedarf)
Verkäufer: Sechs Mark fünfundzwanzig, bitte.	VAA (Verkäufer-Abschluss)
Kunde 2: Auf Wiedersehen!	KAV (Kunden-Verabschiedung)

11.0.14 Transkript 6 - Terminalzeichen

Datum: 28. Juni 1994, **Ort:** Käseverkaufsstand, Aachen, Marktplatz

Transkriptausschnitt	Terminalzeichen
Kunde 1: Guten Morgen!	KBG (Kunden-Gruß)
Verkäufer: Guten Morgen!	VBG (Verkäufer-Gruß)
Kunde 1: Ich hätte gerne fünfhundert Gramm holländi	KBBd (Kunden-Bedarf)
Verkäufer: Am Stück?	VBBd (Verkäufer-Klärung)
Kunde 1: Ja, am Stück, bitte.	KAA (Kunden-Abschluss)

11.0.15 Transkript 7 - Terminalzeichen

Datum: 28. Juni 1994, **Ort:** Bonbonstand, Aachen, Marktplatz, 11:30 Uhr

Transkriptausschnitt	Terminalzeichen
Kunde: Von den gemischten hätte ich gerne hundert G	KBBd (Kunden-Bedarf)
Verkäufer: Für zu Hause oder zum Mitnehmen?	VBBd (Verkäufer-Klärung)
Kunde: Zum Mitnehmen, bitte.	KBA (Kunden-Argument)
Verkäufer: Fünfzig Pfennig, bitte.	VAA (Verkäufer-Abschluss)
Kunde: Danke!	KAA (Kunden-Abschluss)

11.0.16 Transkript 8 - Terminalzeichen

Datum: 9. Juli 1994, **Ort:** Bäckerei, Aachen, 12:00 Uhr

Transkriptausschnitt	Terminalzeichen
Kunde: Guten Tag!	KBG (Kunden-Gruß)
Verkäuferin: Einmal unser bester Kaffee, frisch gem	VBBd (Verkäufer-Klärung)
Kunde: Ja, noch zwei Stück Obstsalat und ein Schäl	KBBd (Kunden-Bedarf)
Verkäuferin: In Ordnung!	VBA (Verkäufer-Argument)
Verkäuferin: Das macht vierzehn Mark und neunzehn P	VAA (Verkäufer-Abschluss)
Kunde: Ich zahle in Kleingeld.	KAA (Kunden-Abschluss)
Verkäuferin: Vielen Dank, schönen Sonntag noch!	VAV (Verkäufer-Verabschiedung)
Kunde: Danke, Ihnen auch!	KAV (Kunden-Verabschiedung)

Dann folg auf der Basis dieser acht Terminalzeichenketten die Induktion und Optimierung einer auf dem gesamten erhobenen empirischen Material beruhende Induktion einer probabilistischen Grammatik und ihre Optimierung.

[]:

Auf Grundlage der vorläufigen probabilistischen kontextfreien Grammatik (PCFG) werden die Transkripte in die entsprechenden Terminalzeichen umwandeln. Hier ist die detaillierte Zuordnung für jedes Transkript, wobei für jede relevante Aktion oder Aussage des Gesprächs ein passendes Terminalzeichen aus der vorgegebenen Grammatik verwende.

11.0.17 Transkript 1 - Terminalzeichen

Datum: 28. Juni 1994, **Ort:** Metzgerei, Aachen, 11:00 Uhr

Transkriptausschnitt	Terminalzeichen
Kunde: Guten Tag	KBG (Kunden-Gruß)
Verkäuferin: Guten Tag	VBG (Verkäufer-Gruß)
Kunde: Einmal von der groben Leberwurst, bitte.	KBBd (Kunden-Bedarf)
Verkäuferin: Wie viel darf's denn sein?	VBBd (Verkäufer-Klärung)
Kunde: Zwei hundert Gramm.	KBA (Kunden-Argument)
Verkäuferin: Sonst noch etwas?	VBA (Verkäufer-Argument)

Transkriptausschnitt	Terminalzeichen
Kunde: Ja, dann noch ein Stück von dem Schwarzwalde	KBBd (Kunden-Bedarf)
Verkäuferin: Wie groß soll das Stück sein?	VBBd (Verkäufer-Klärung)
Kunde: So um die dreihundert Gramm.	KBA (Kunden-Argument)
Verkäuferin: Das macht dann acht Mark zwanzig.	VAA (Verkäufer-Abschluss)
Kunde: Bitte.	KAA (Kunden-Abschluss)
Verkäuferin: Danke und einen schönen Tag noch!	VAV (Verkäufer- Verabschiedung)
Kunde: Danke, ebenfalls!	KAV (Kunden- Verabschiedung)

11.0.18 Transkript 2 - Terminalzeichen

Datum: 28. Juni 1994, **Ort:** Marktplatz, Aachen

Transkriptausschnitt	Terminalzeichen
Verkäufer: Kirschen kann jeder probieren hier!	VBG (Verkäufer-Gruß)
Kunde 1: Ein halbes Kilo Kirschen, bitte.	KBBd (Kunden-Bedarf)
Verkäufer: Ein halbes Kilo? Oder ein Kilo?	VBBd (Verkäufer-Klärung)
Verkäufer: Drei Mark, bitte.	VAA (Verkäufer-Abschluss)
Kunde 1: Danke schön!	KAA (Kunden-Abschluss)
Verkäufer: Kirschen kann jeder probieren hier!	VBG (Verkäufer-Gruß)
Kunde 2: Ein halbes Kilo, bitte.	KBBd (Kunden-Bedarf)
Verkäufer: Drei Mark, bitte.	VAA (Verkäufer-Abschluss)
Kunde 2: Danke schön!	KAA (Kunden-Abschluss)

11.0.19 Transkript 3 - Terminalzeichen

Datum: 28. Juni 1994, **Ort:** Fischstand, Marktplatz, Aachen

Transkriptausschnitt	Terminalzeichen
Kunde: Ein Pfund Seelachs, bitte.	KBBd (Kunden-Bedarf)
Verkäufer: Seelachs, alles klar.	VBBd (Verkäufer-Klärung)
Verkäufer: Vier Mark neunzehn, bitte.	VAA (Verkäufer-Abschluss)
Kunde: Danke schön!	KAA (Kunden-Abschluss)

11.0.20 Transkript 4 - Terminalzeichen

Datum: 28. Juni 1994, **Ort:** Gemüsestand, Aachen, Marktplatz, 11:00 Uhr

Transkriptausschnitt	Terminalzeichen
Kunde: Hören Sie, ich nehme ein paar Champignons mit.	KBBd (Kunden-Bedarf)
Verkäufer: Braune oder helle?	VBBd (Verkäufer-Klärung)
Kunde: Nehmen wir die hellen.	KBA (Kunden-Argument)
Verkäufer: Die sind beide frisch, keine Sorge.	VBA (Verkäufer-Argument)
Kunde: Wie ist es mit Pfifferlingen?	KBBd (Kunden-Bedarf)
Verkäufer: Ah, die sind super!	VBA (Verkäufer-Argument)
Kunde: Kann ich die in Reissalat tun?	KAE (Kunden-Einwand)
Verkäufer: Eher kurz anbraten in der Pfanne.	VAE (Verkäufer-Einwand)
Kunde: Okay, mache ich.	KAA (Kunden-Abschluss)
Verkäufer: Schönen Tag noch!	VAV (Verkäufer-Verabschiedung)
Kunde: Gleichfalls!	KAV (Kunden-Verabschiedung)

11.0.21 Transkript 5 - Terminalzeichen

Datum: 26. Juni 1994, **Ort:** Gemüsestand, Aachen, Marktplatz, 11:00 Uhr

Transkriptausschnitt	Terminalzeichen
Kunde 1: Auf Wiedersehen!	KAV (Kunden-Verabschiedung)
Kunde 2: Ich hätte gern ein Kilo von den Granny Smi	KBBd (Kunden-Bedarf)
Verkäufer: Sonst noch etwas?	VBBd (Verkäufer-Klärung)
Kunde 2: Ja, noch ein Kilo Zwiebeln.	KBBd (Kunden-Bedarf)
Verkäufer: Sechs Mark fünfundzwanzig, bitte.	VAA (Verkäufer-Abschluss)
Kunde 2: Auf Wiedersehen!	KAV (Kunden-Verabschiedung)

11.0.22 Transkript 6 - Terminalzeichen

Datum: 28. Juni 1994, **Ort:** Käseverkaufsstand, Aachen, Marktplatz

Transkriptausschnitt	Terminalzeichen
Kunde 1: Guten Morgen!	KBG (Kunden-Gruß)
Verkäufer: Guten Morgen!	VBG (Verkäufer-Gruß)
Kunde 1: Ich hätte gerne fünfhundert Gramm holländi	KBBd (Kunden-Bedarf)
Verkäufer: Am Stück?	VBBd (Verkäufer-Klärung)

Transkriptausschnitt	Terminalzeichen
Kunde 1: Ja, am Stück, bitte.	KAA (Kunden-Abschluss)

11.0.23 Transkript 7 - Terminalzeichen

Datum: 28. Juni 1994, **Ort:** Bonbonstand, Aachen, Marktplatz, 11:30 Uhr

Transkriptausschnitt	Terminalzeichen
Kunde: Von den gemischten hätte ich gerne hundert G	KBBd (Kunden-Bedarf)
Verkäufer: Für zu Hause oder zum Mitnehmen?	VBBd (Verkäufer-Klärung)
Kunde: Zum Mitnehmen, bitte.	KBA (Kunden-Argument)
Verkäufer: Fünfzig Pfennig, bitte.	VAA (Verkäufer-Abschluss)
Kunde: Danke!	KAA (Kunden-Abschluss)

11.0.24 Transkript 8 - Terminalzeichen

Datum: 9. Juli 1994, **Ort:** Bäckerei, Aachen, 12:00 Uhr

Transkriptausschnitt	Terminalzeichen
Kunde: Guten Tag!	KBG (Kunden-Gruß)
Verkäuferin: Einmal unser bester Kaffee, frisch gem	VBBd (Verkäufer-Klärung)
Kunde: Ja, noch zwei Stück Obstsalat und ein Schäl	KBBd (Kunden-Bedarf)
Verkäuferin: In Ordnung!	VBA (Verkäufer-Argument)
Verkäuferin: Das macht vierzehn Mark und neunzehn P	VAA (Verkäufer-Abschluss)
Kunde: Ich zahle in Kleingeld.	KAA (Kunden-Abschluss)
Verkäuferin: Vielen Dank, schönen Sonntag noch!	VAV (Verkäufer-Verabschiedung)
Kunde: Danke, Ihnen auch!	KAV (Kunden-Verabschiedung)

```
[4]: import numpy as np
from scipy.stats import pearsonr

# Neue empirische Terminalzeichenketten
empirical_chains = [
    ['KBG', 'VBG', 'KBBd', 'VBBd', 'KBA', 'VBA', 'KBBd', 'VBBd', 'KBA', 'VAA', 'KAA', 'VAV', 'KAV'],
    ['VBG', 'KBBd', 'VBBd', 'VAA', 'KAA', 'VBG', 'KBBd', 'VAA', 'KAA'],
    ['KBBd', 'VBBd', 'VAA', 'KAA'],
```

```

    ['KBBd', 'VBBd', 'KBA', 'VBA', 'KBBd', 'VBA', 'KAE', 'VAE', 'KAA', 'VAV',
    ↪ 'KAV'],
    ['KAV', 'KBBd', 'VBBd', 'KBBd', 'VAA', 'KAV'],
    ['KBG', 'VBG', 'KBBd', 'VBBd', 'KAA'],
    ['KBBd', 'VBBd', 'KBA', 'VAA', 'KAA'],
    ['KBG', 'VBBd', 'KBBd', 'VBA', 'VAA', 'KAA', 'VAV', 'KAV']
]

# Übergangszählung initialisieren
transitions = {}
for chain in empirical_chains:
    for i in range(len(chain) - 1):
        start, end = chain[i], chain[i + 1]
        if start not in transitions:
            transitions[start] = {}
        if end not in transitions[start]:
            transitions[start][end] = 0
        transitions[start][end] += 1

# Normalisierung: Übergangswahrscheinlichkeiten berechnen
probabilities = {}
for start in transitions:
    total = sum(transitions[start].values())
    probabilities[start] = {end: count / total for end, count in
    ↪ transitions[start].items()}

# Terminalzeichen und Startzeichen definieren
terminal_symbols = list(set([item for sublist in empirical_chains for item in
    ↪ sublist]))
start_symbol = empirical_chains[0][0]

# Funktion zur Generierung von Ketten basierend auf der Grammatik
def generate_chain(max_length=10):
    chain = [start_symbol]
    while len(chain) < max_length:
        current = chain[-1]
        if current not in probabilities:
            break
        next_symbol = np.random.choice(list(probabilities[current].keys()),
    ↪ p=list(probabilities[current].values()))
        chain.append(next_symbol)
        if next_symbol not in probabilities:
            break
    return chain

# Funktion zur Berechnung relativer Häufigkeiten
def compute_frequencies(chains, terminals):

```



```

frequency_array = np.zeros(len(terminals))
terminal_index = {term: i for i, term in enumerate(terminals)}

for chain in chains:
    for symbol in chain:
        if symbol in terminal_index:
            frequency_array[terminal_index[symbol]] += 1

total = frequency_array.sum()
if total > 0:
    frequency_array /= total # Normierung der Häufigkeiten

return frequency_array

# Iterative Optimierung
max_iterations = 1000
tolerance = 0.01 # Toleranz für Standardmessfehler
best_correlation = 0
best_significance = 1

# Relativ Häufigkeiten der empirischen Ketten berechnen
empirical_frequencies = compute_frequencies(empirical_chains, terminal_symbols)

for iteration in range(max_iterations):
    # Generiere 8 künstliche Ketten
    generated_chains = [generate_chain() for _ in range(8)]

    # Relativ Häufigkeiten der generierten Ketten berechnen
    generated_frequencies = compute_frequencies(generated_chains,
↳terminal_symbols)

    # Berechne die Korrelation
    correlation, p_value = pearsonr(empirical_frequencies,
↳generated_frequencies)

    print(f"Iteration {iteration + 1}, Korrelation: {correlation:.3f},
↳Signifikanz: {p_value:.3f}")

    # Überprüfen, ob Korrelation und Signifikanz akzeptabel sind
    if correlation >= 0.9 and p_value < 0.05:
        best_correlation = correlation
        best_significance = p_value
        break

# Anpassung der Wahrscheinlichkeiten basierend auf Standardmessfehler
for start in probabilities:
    for end in probabilities[start]:

```

```

# Fehlerabschätzung basierend auf Differenz der Häufigkeiten
empirical_prob = empirical_frequencies[terminal_symbols.index(end)]
generated_prob = generated_frequencies[terminal_symbols.index(end)]
error = empirical_prob - generated_prob

# Anpassung der Wahrscheinlichkeit
probabilities[start][end] += error * tolerance
probabilities[start][end] = max(0, min(1,
↳probabilities[start][end])) # Begrenzen auf [0,1]

# Normalisierung
for start in probabilities:
    total = sum(probabilities[start].values())
    if total > 0:
        probabilities[start] = {end: prob / total for end, prob in
↳probabilities[start].items()}

# Ergebnis ausgeben
print("\nOptimierte probabilistische Grammatik:")
for start, transitions in probabilities.items():
    print(f"{start} → {transitions}")

print(f"\nBeste Korrelation: {best_correlation:.3f}, Signifikanz:
↳{best_significance:.3f}")

```

Iteration 1, Korrelation: 0.925, Signifikanz: 0.000

Optimierte probabilistische Grammatik:

```

KBG → {'VBG': 0.6666666666666666, 'VBBd': 0.3333333333333333}
VBG → {'KBBd': 1.0}
KBBd → {'VBBd': 0.6666666666666666, 'VAA': 0.1666666666666666, 'VBA':
0.1666666666666666}
VBBd → {'KBA': 0.4444444444444444, 'VAA': 0.2222222222222222, 'KBBd':
0.2222222222222222, 'KAA': 0.1111111111111111}
KBA → {'VBA': 0.5, 'VAA': 0.5}
VBA → {'KBBd': 0.5, 'KAE': 0.25, 'VAA': 0.25}
VAA → {'KAA': 0.8571428571428571, 'KAV': 0.14285714285714285}
KAA → {'VAV': 0.75, 'VBG': 0.25}
VAV → {'KAV': 1.0}
KAE → {'VAE': 1.0}
VAE → {'KAA': 1.0}
KAV → {'KBBd': 1.0}

```

Beste Korrelation: 0.925, Signifikanz: 0.000

[]: