Kommunikation mit gesprochener Sprache ...

... ist mehr als Hören und Sprechen:

Kommunikation mit gesprochener Sprache ...

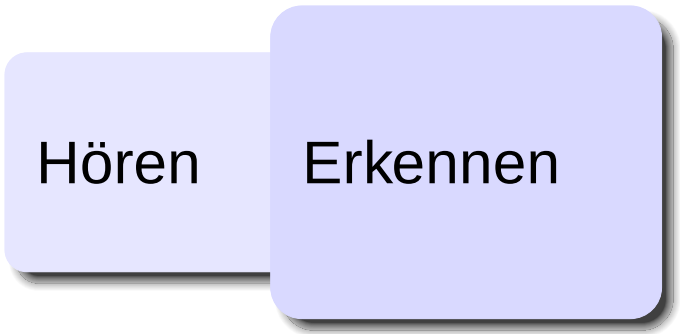
... ist mehr als Hören und Sprechen:

Hören

Kikala brint tovoluti?

Kommunikation mit gesprochener Sprache ...

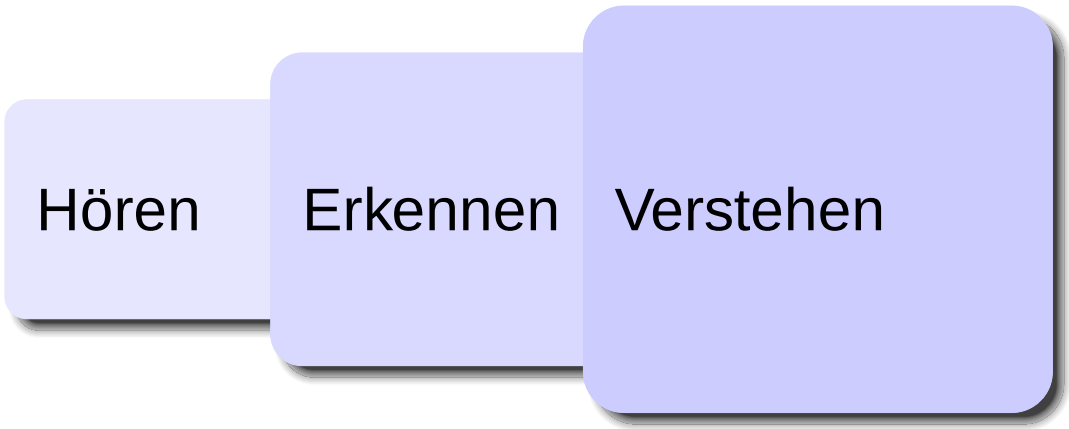
... ist mehr als Hören und Sprechen:



Winter kochtest ganz Blatt?

Kommunikation mit gesprochener Sprache ...

... ist mehr als Hören und Sprechen:



Wann steigt die Party?

Kommunikation mit gesprochener Sprache ...

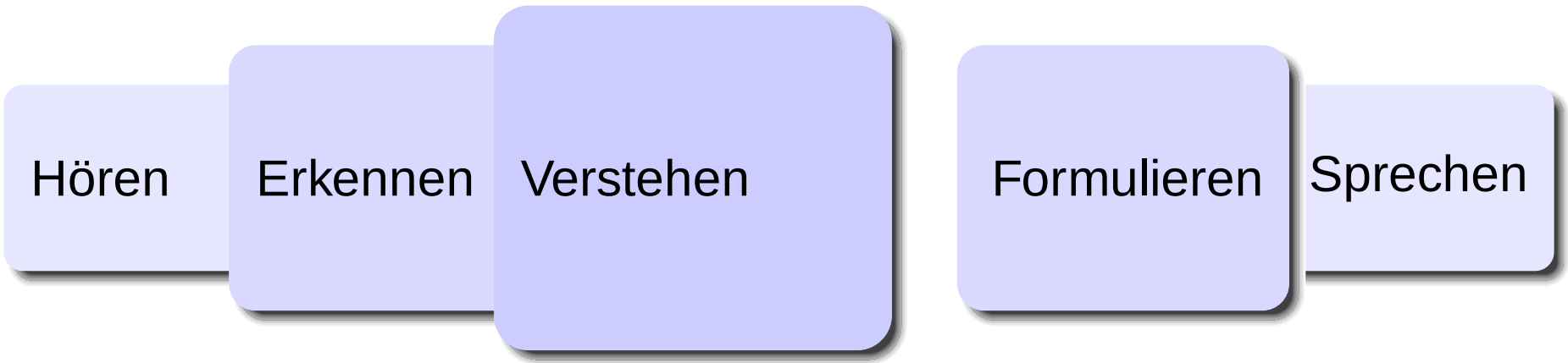
... ist mehr als Hören und Sprechen:

Wann steigt die Party?

Sintu högafi notsi!

Kommunikation mit gesprochener Sprache ...

... ist mehr als Hören und Sprechen:



Wann steigt die Party?

Sonderbar werfen die Wellen hinab!

Kommunikation mit gesprochener Sprache ...

... ist mehr als Hören und Sprechen:

Wann steigt die Party?

Am Freitag, im Hirsch!

Kommunikation mit gesprochener Sprache ...

... ist mehr als Hören und Sprechen:

Wann steigt die Party?

Am Freitag, im Hirsch!

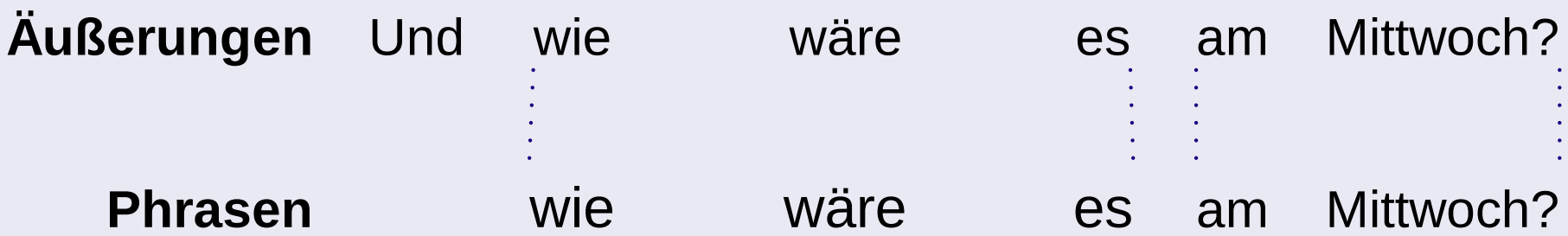
Gesprochene Sprache

Segmentstruktur

Äußerungen	Und	wie	wäre	es	am	Mittwoch?

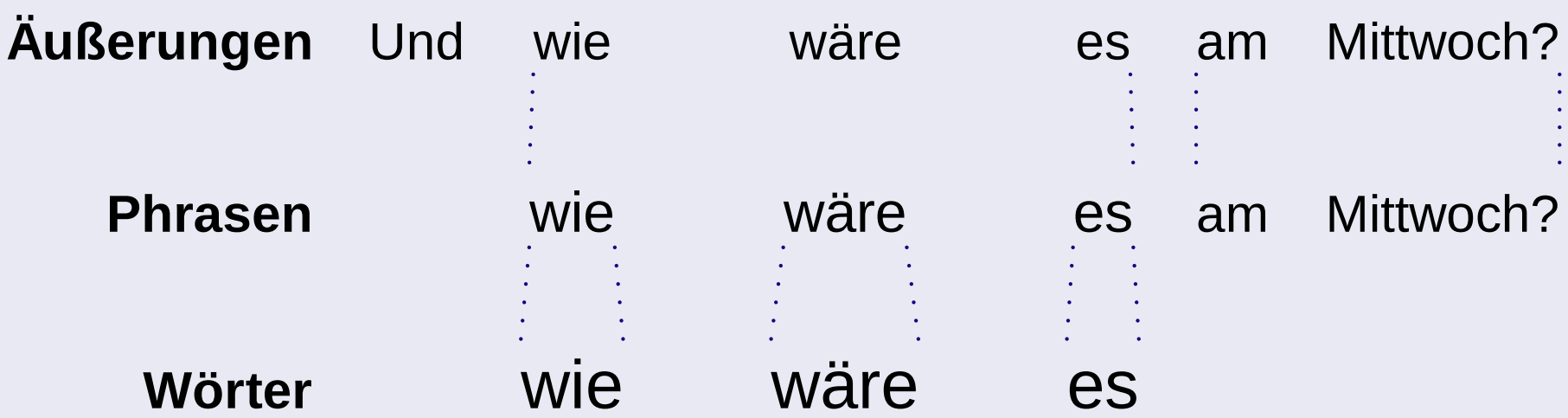
Gesprochene Sprache

Segmentstruktur



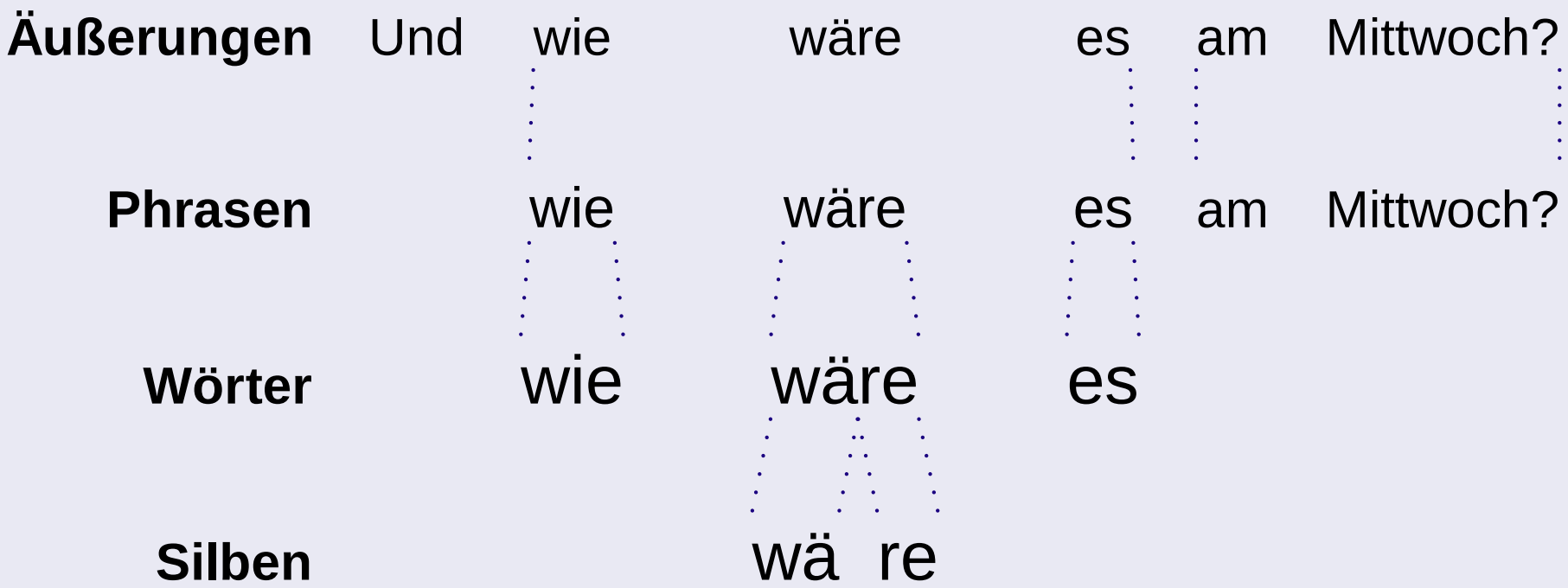
Gesprochene Sprache

Segmentstruktur



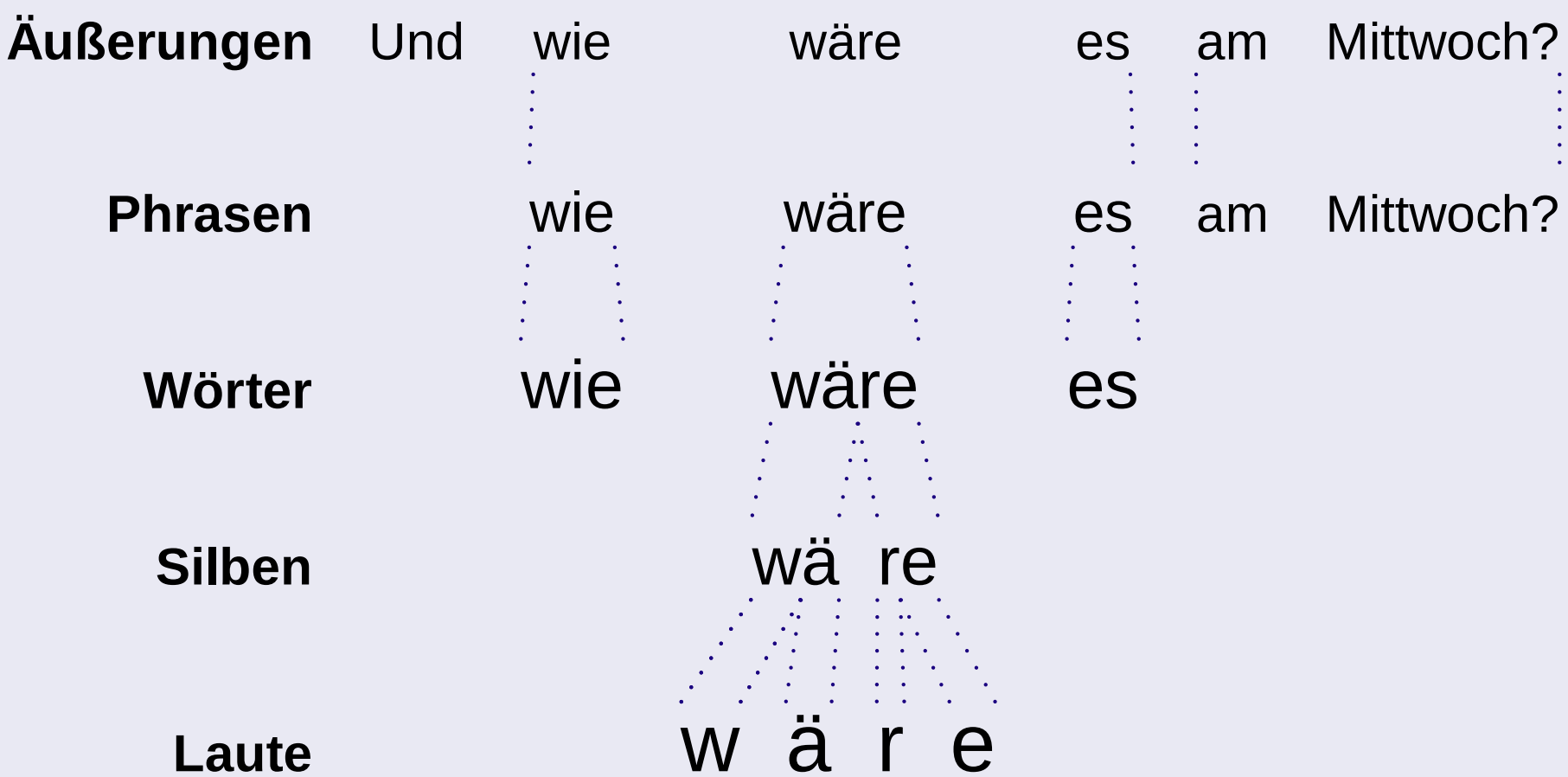
Gesprochene Sprache

Segmentstruktur



Gesprochene Sprache

Segmentstruktur



Gesprochene Sprache

Segmentübergreifende Information (Prosodie)

Gesprochene Sprache

Segmentübergreifende Information (Prosodie)

- Grundfrequenz

Gesprochene Sprache

Segmentübergreifende Information (Prosodie)

- Grundfrequenz
- Rhythmus

Gesprochene Sprache

Segmentübergreifende Information (Prosodie)

- Grundfrequenz
- Rhythmus
- Lautstärke

Gesprochene Sprache

Segmentübergreifende Information (Prosodie)

- Grundfrequenz
- Rhythmus
- Lautstärke

wird verwendet ...

Gesprochene Sprache

Segmentübergreifende Information (Prosodie)

- Grundfrequenz
- Rhythmus
- Lautstärke

wird verwendet ...

- ... für Hervorhebungen (Neues, Wichtiges, Unerwartetes, ...)

Gesprochene Sprache

Segmentübergreifende Information (Prosodie)

- Grundfrequenz
- Rhythmus
- Lautstärke

wird verwendet ...

- ... für Hervorhebungen (Neues, Wichtiges, Unerwartetes, ...)
- ... zur Gliederung (Phrasen, Sätze)

Gesprochene Sprache

Segmentübergreifende Information (Prosodie)

- Grundfrequenz
- Rhythmus
- Lautstärke

wird verwendet ...

- ... für Hervorhebungen (Neues, Wichtiges, Unerwartetes, ...)
- ... zur Gliederung (Phrasen, Sätze)
- ... zur Moduskennzeichnung (Aussage, Frage, ...)

Gesprochene Sprache

Segmentübergreifende Information (Prosodie)

- Grundfrequenz
- Rhythmus
- Lautstärke

wird verwendet ...

- ... für Hervorhebungen (Neues, Wichtiges, Unerwartetes, ...)
- ... zur Gliederung (Phrasen, Sätze)
- ... zur Moduskennzeichnung (Aussage, Frage, ...)
- ... zur Dialogsteuerung (Vergabe der Initiative)

Gesprochene Sprache

Segmentübergreifende Information (Prosodie)

- Grundfrequenz
- Rhythmus
- Lautstärke

wird verwendet ...

- ... für Hervorhebungen (Neues, Wichtiges, Unerwartetes, ...)
- ... zur Gliederung (Phrasen, Sätze)
- ... zur Moduskennzeichnung (Aussage, Frage, ...)
- ... zur Dialogsteuerung (Vergabe der Initiative)
- ... zum Ausdruck von Emotionen (Freude, Angst, Überraschung, Verlegenheit, ...)

Sprachsynthese

Zwei Herangehensweisen:

Sprachsynthese

Zwei Herangehensweisen:

Vollsynthese

- Erzeugen des Sprachsignals durch Ton- und Rauschgeneratoren
- akzeptable Verständlichkeit
- geringe Natürlichkeit

Sprachsynthese

Zwei Herangehensweisen:

Vollsynthese

- Erzeugen des Sprachsignals durch Ton- und Rauschgeneratoren
- akzeptable Verständlichkeit
- geringe Natürlichkeit

reproduktive Synthese

- Aufnehmen und Wiedergeben menschlicher Sprachsignale
- hohe Verständlichkeit
- gute bis hohe Natürlichkeit

Sprachsynthese

reproduktive Synthese

Was sind geeignete Basiseinheiten?

- ganze Phrasen: nur für Spezialanwendungen
- Laute: schlechte Qualität
- Kompromiss: flexible Ermittlung der Basiselemente aus Korpusdaten

Sprachsynthese

reproduktive Synthese

Was sind geeignete Basiseinheiten?

- ganze Phrasen: nur für Spezialanwendungen
- Laute: schlechte Qualität
- Kompromiss: flexible Ermittlung der Basiselemente aus Korpusdaten

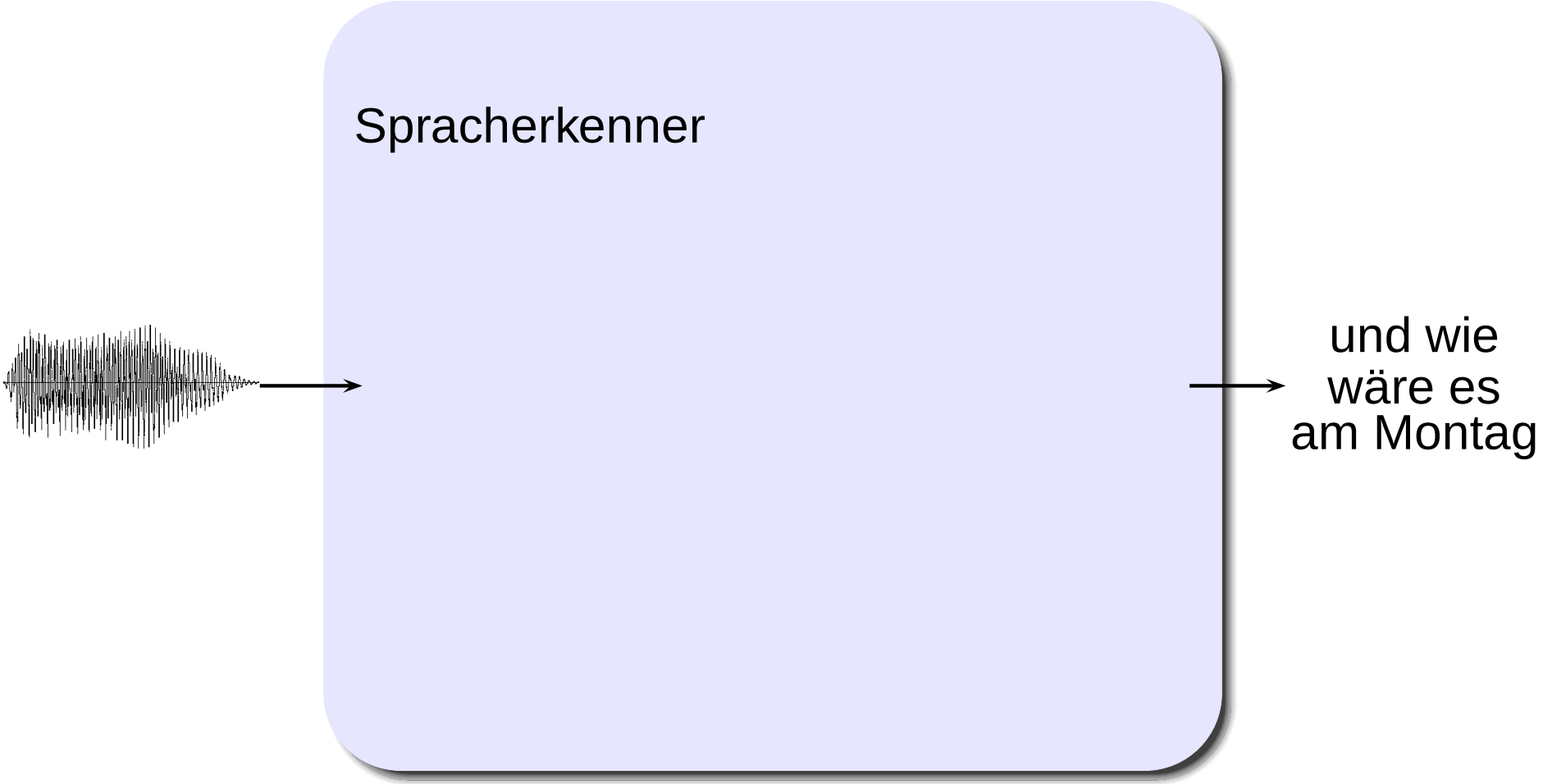
Wie werden die Basiselemente verkettet?

- harter Schnitt erzeugt Knackgeräusche
- ”weiche” Übergänge erforderlich
- prosodische Variation durch spezielle Transformationsverfahren

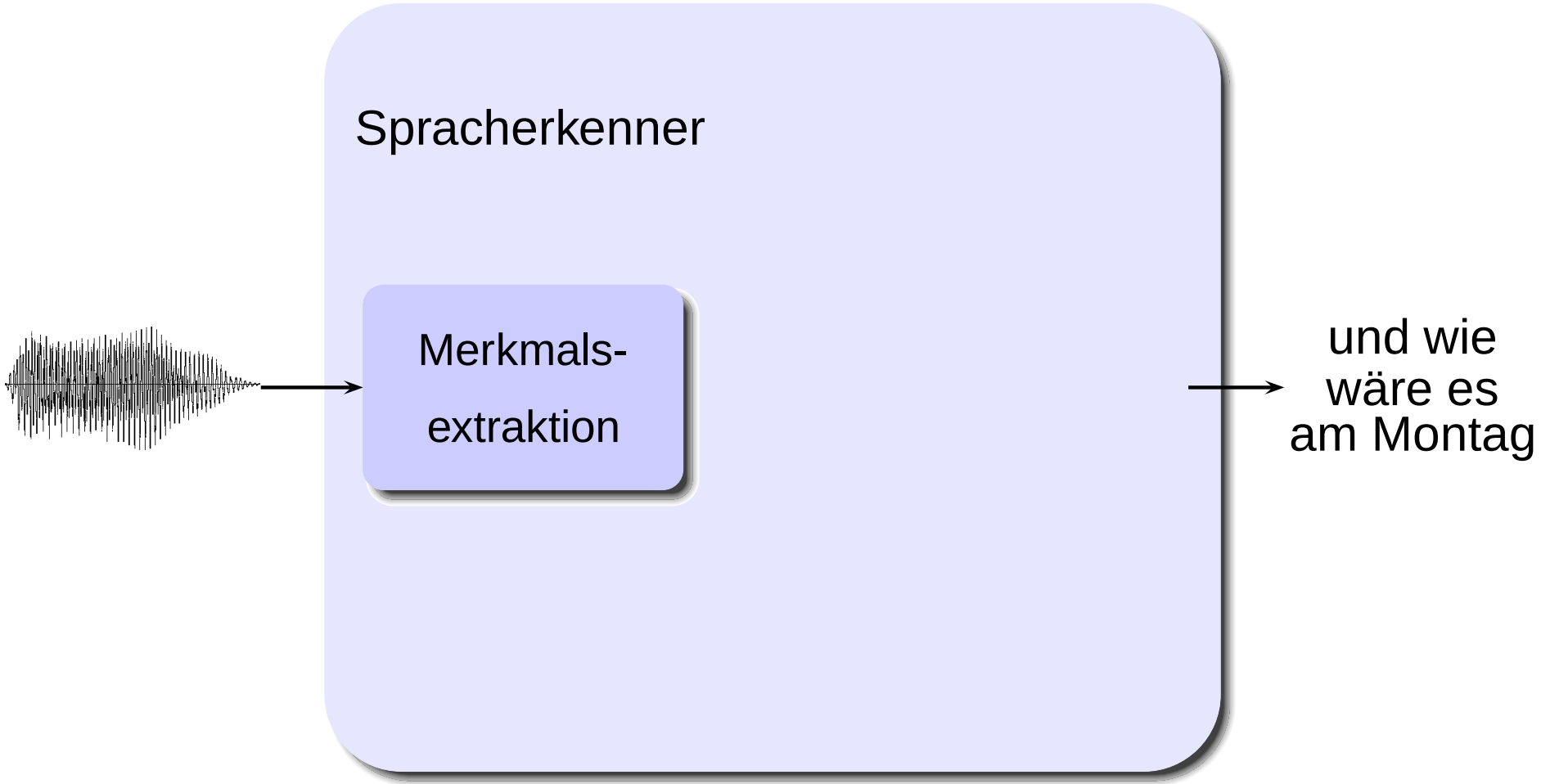
Spracherkennung

- nur Berücksichtigung von Lautcharakteristika
- ”Training” von Modellen auf großen Sprachdatensammlungen
- Vernachlässigung der Prosodie
- nur Erkennung, kein Sprachverstehen!

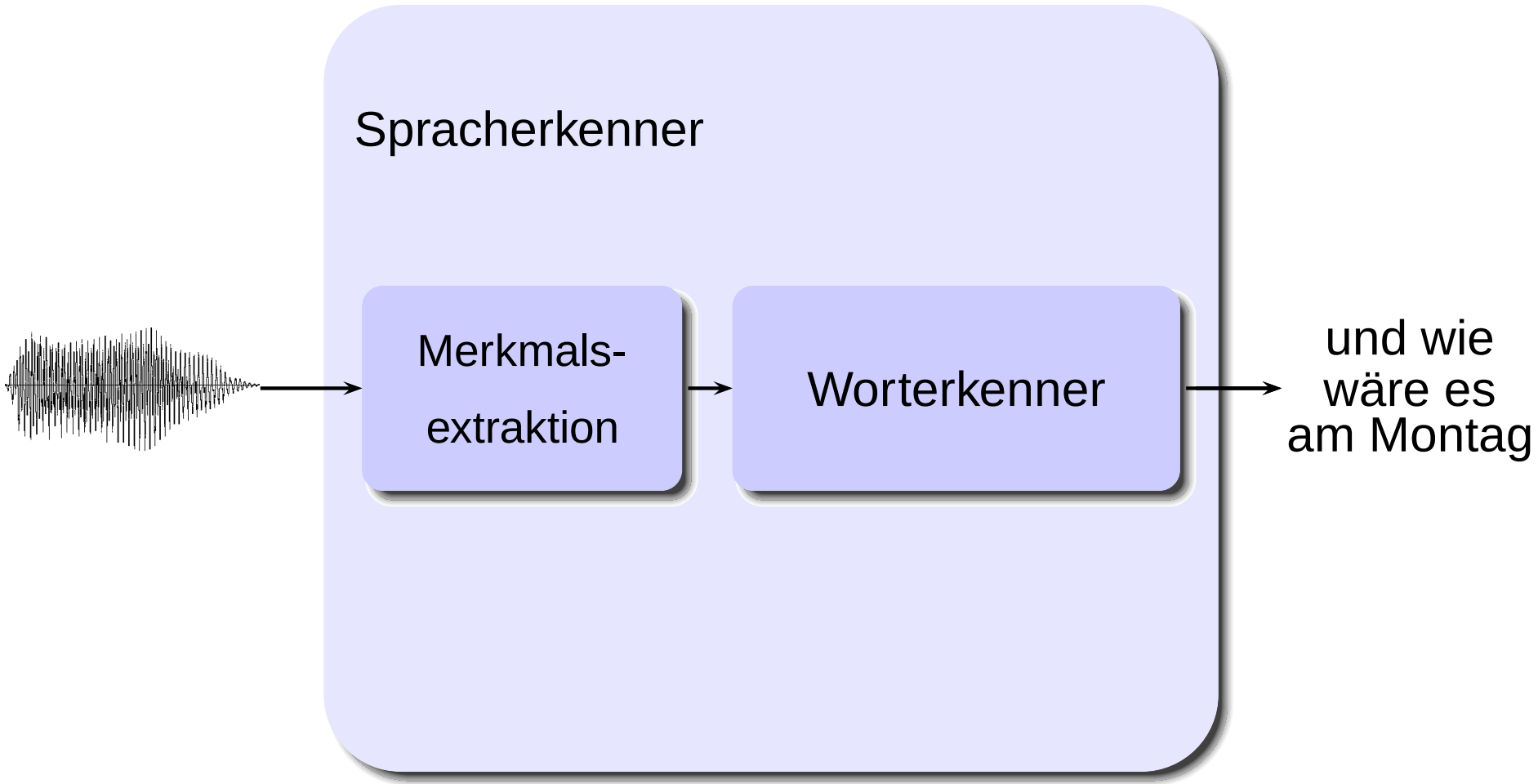
Spracherkennung



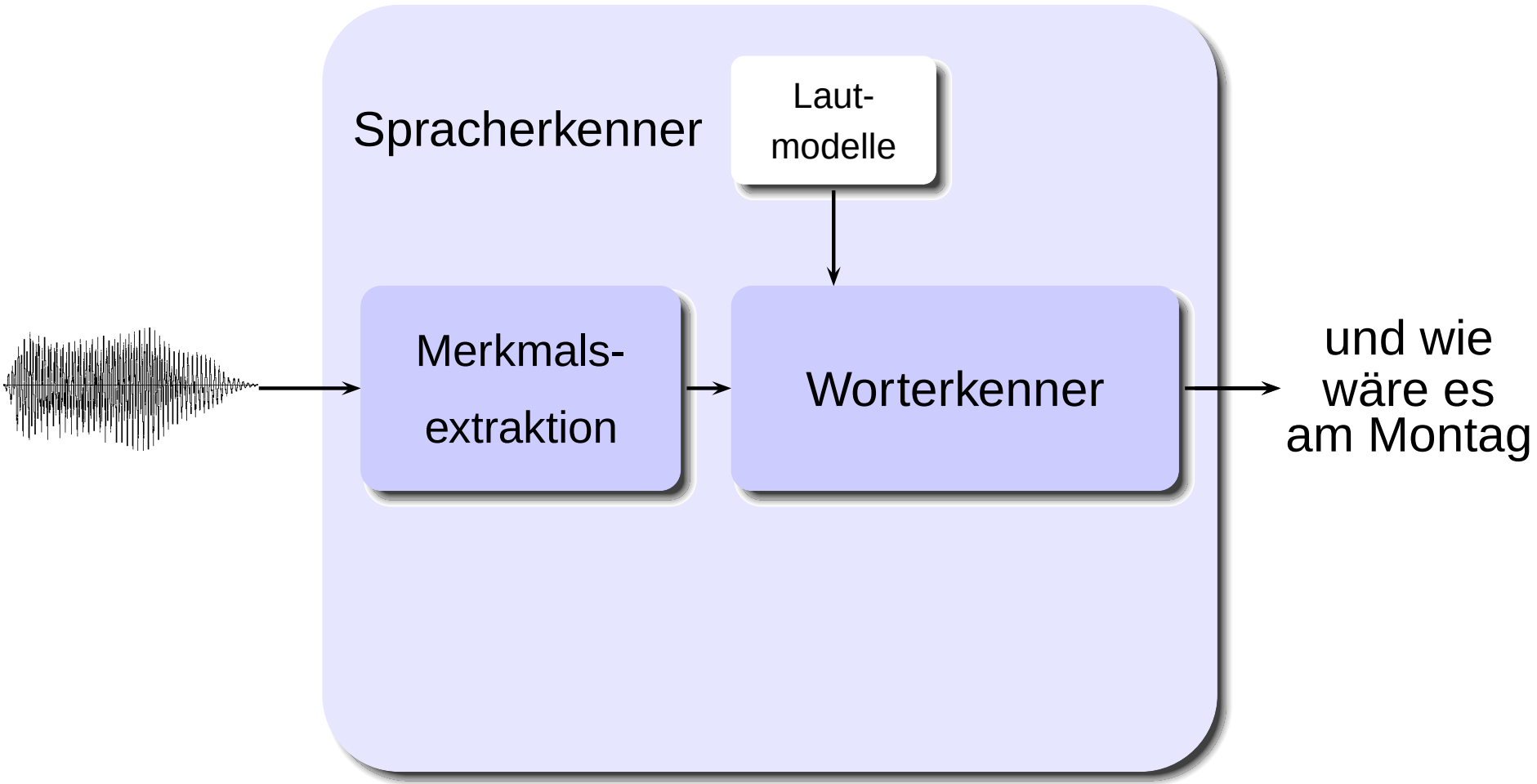
Spracherkennung



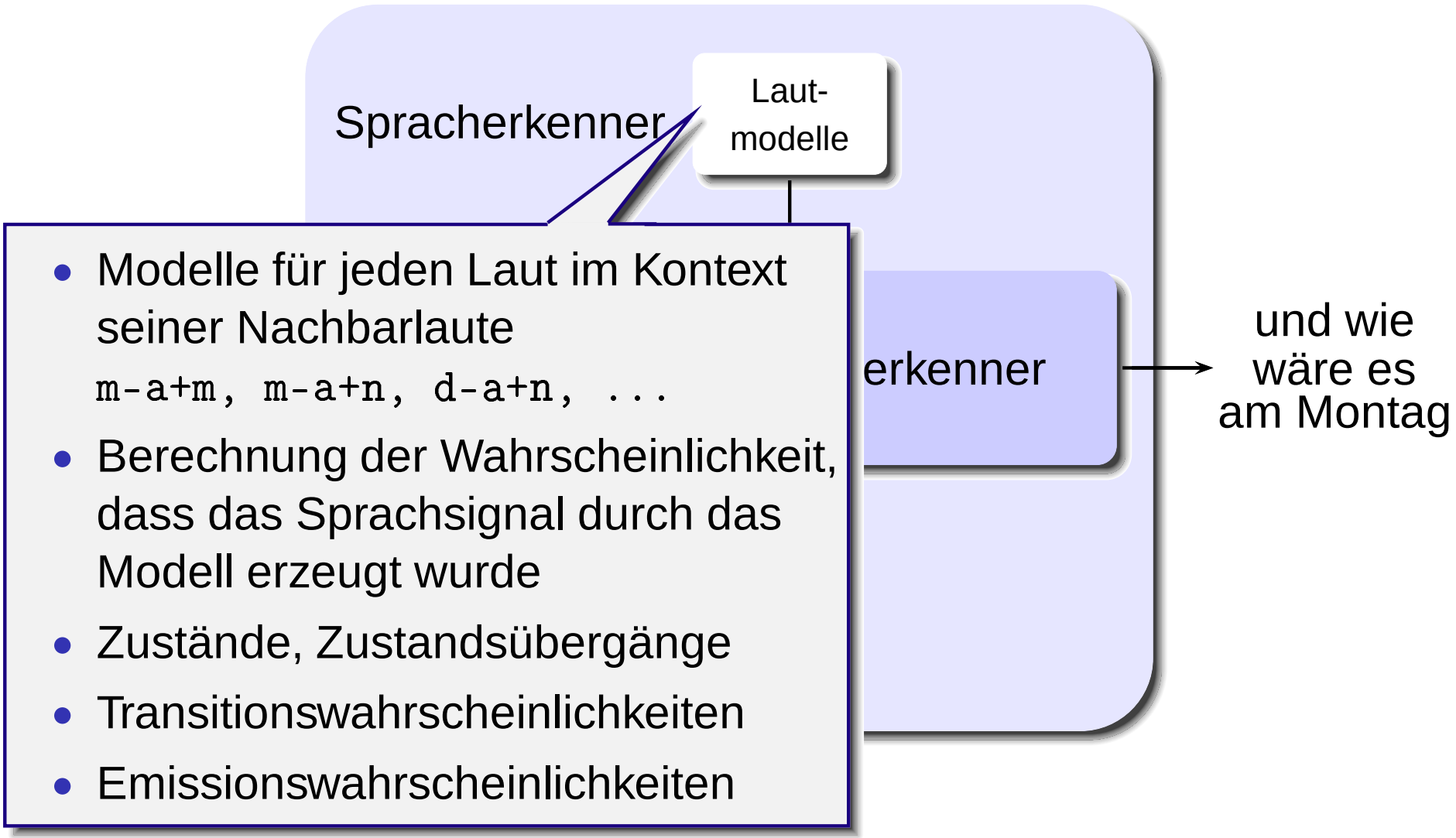
Spracherkennung



Spracherkennung

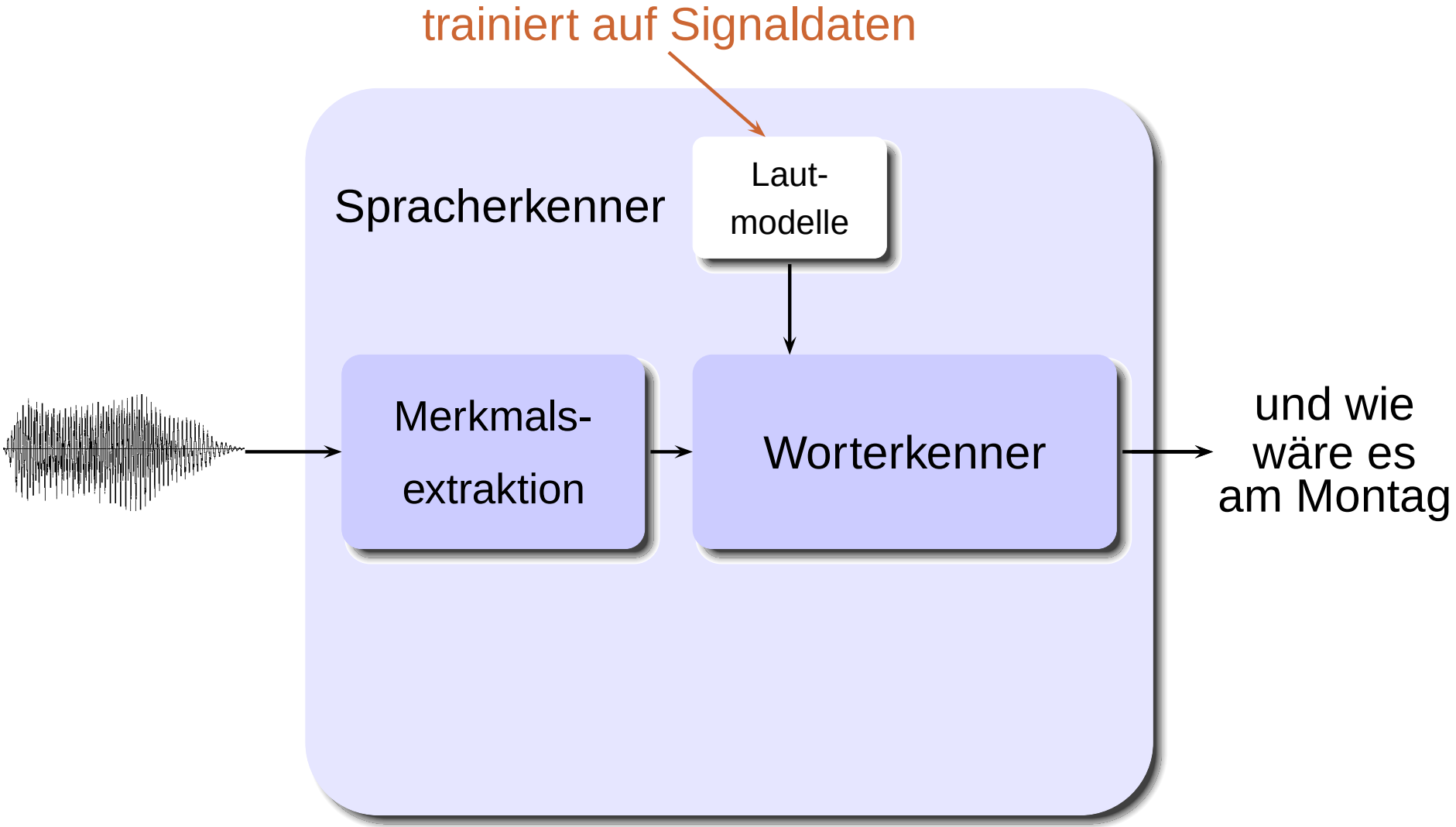


Spracherkennung

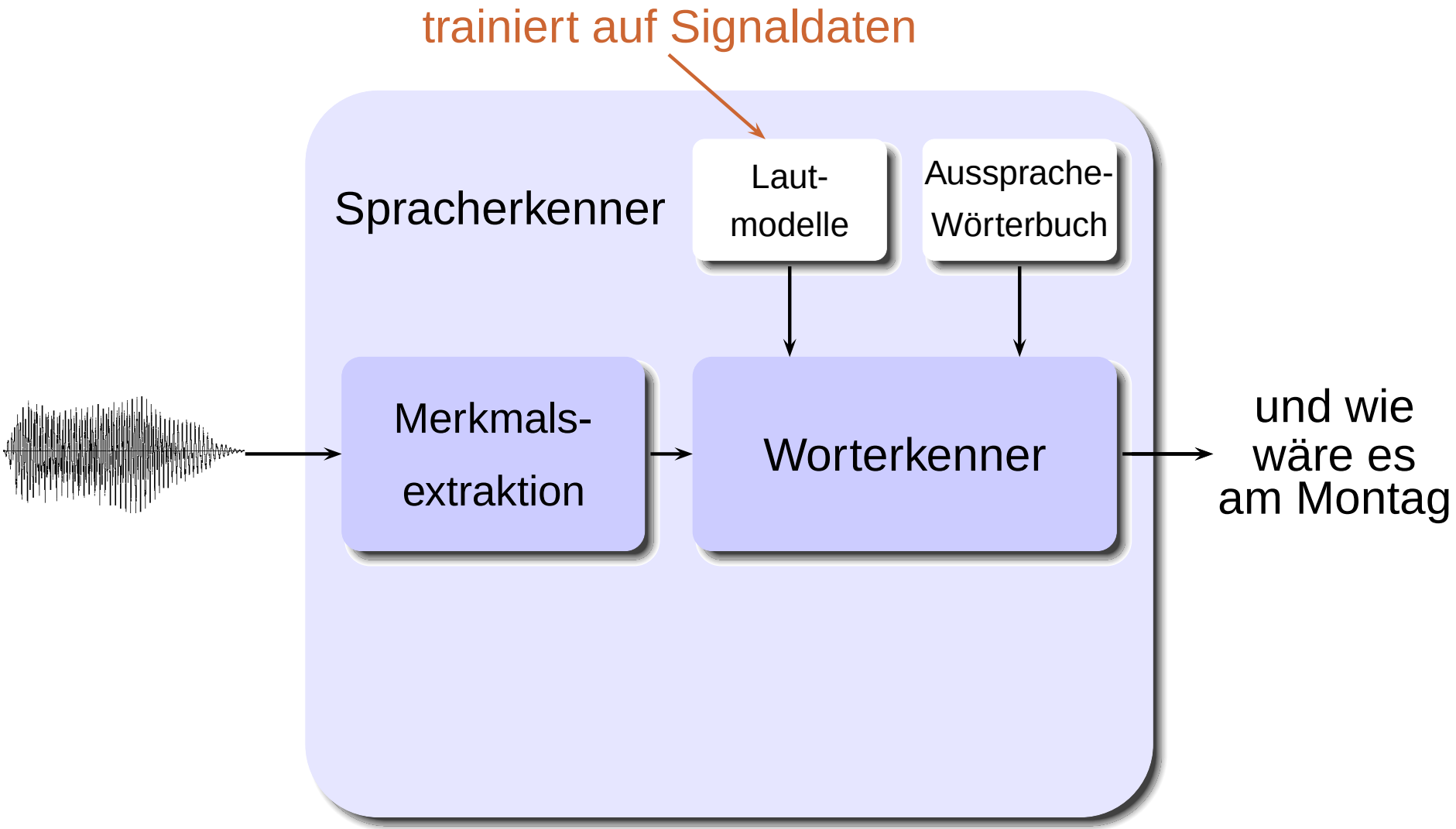


- Modelle für jeden Laut im Kontext seiner Nachbarlaute
  `m-a+m, m-a+n, d-a+n, ...`
- Berechnung der Wahrscheinlichkeit, dass das Sprachsignal durch das Modell erzeugt wurde
- Zustände, Zustandsübergänge
- Transitionswahrscheinlichkeiten
- Emissionswahrscheinlichkeiten

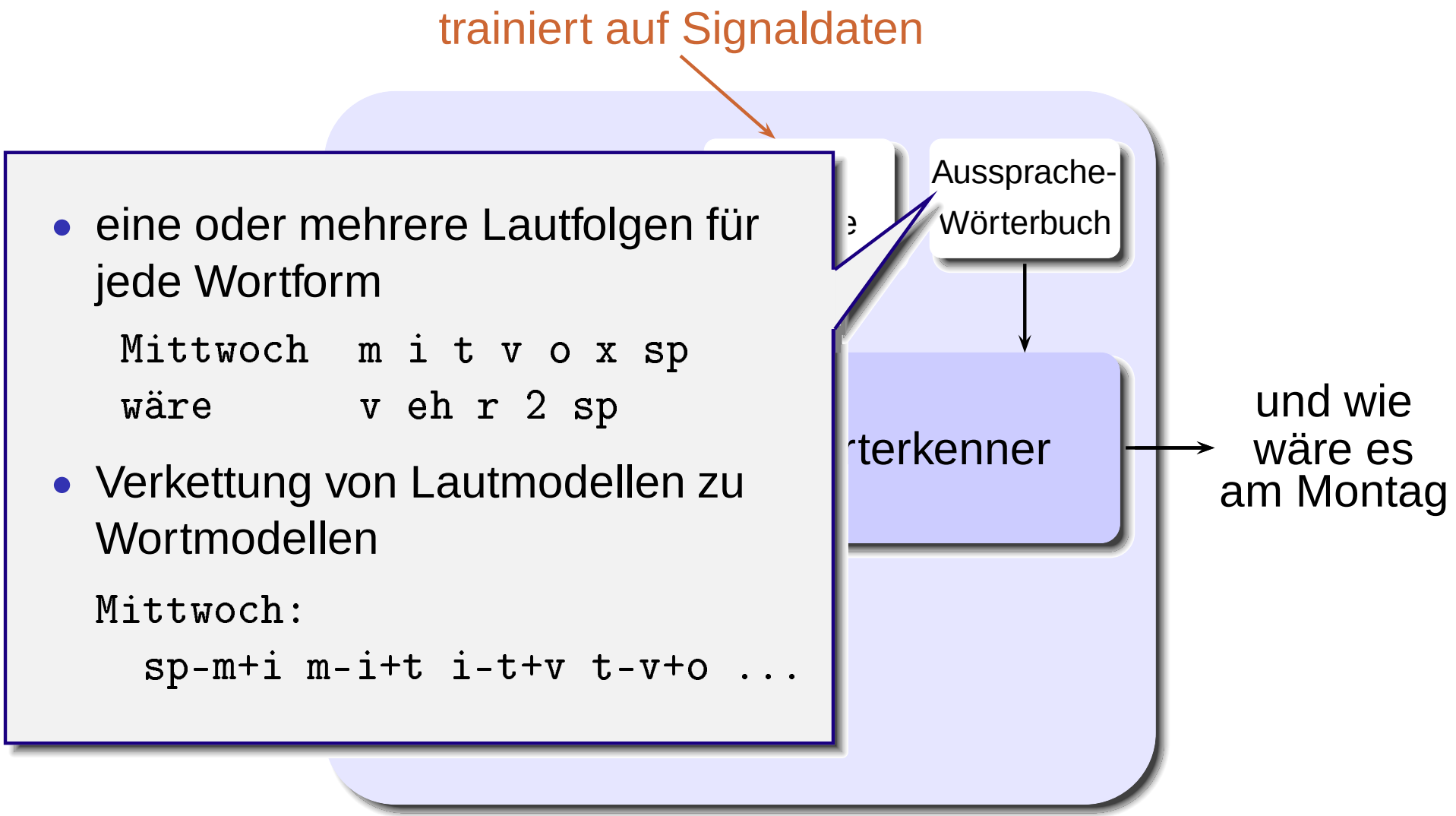
Spracherkennung



Spracherkennung



Spracherkennung



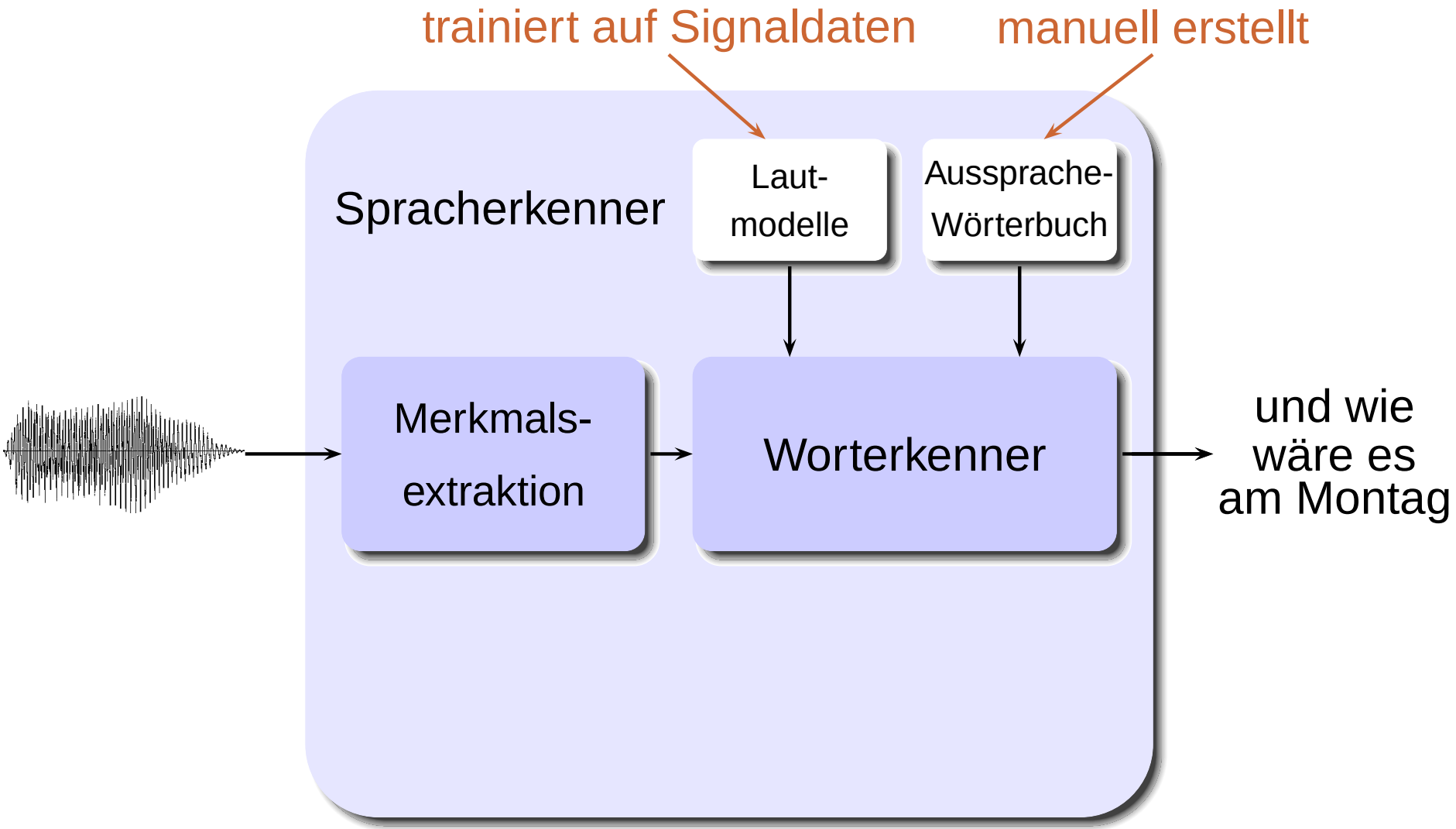
- eine oder mehrere Lautfolgen für jede Wortform

```
Mittwoch   m i t v o x sp
wäre       v eh r 2 sp
```

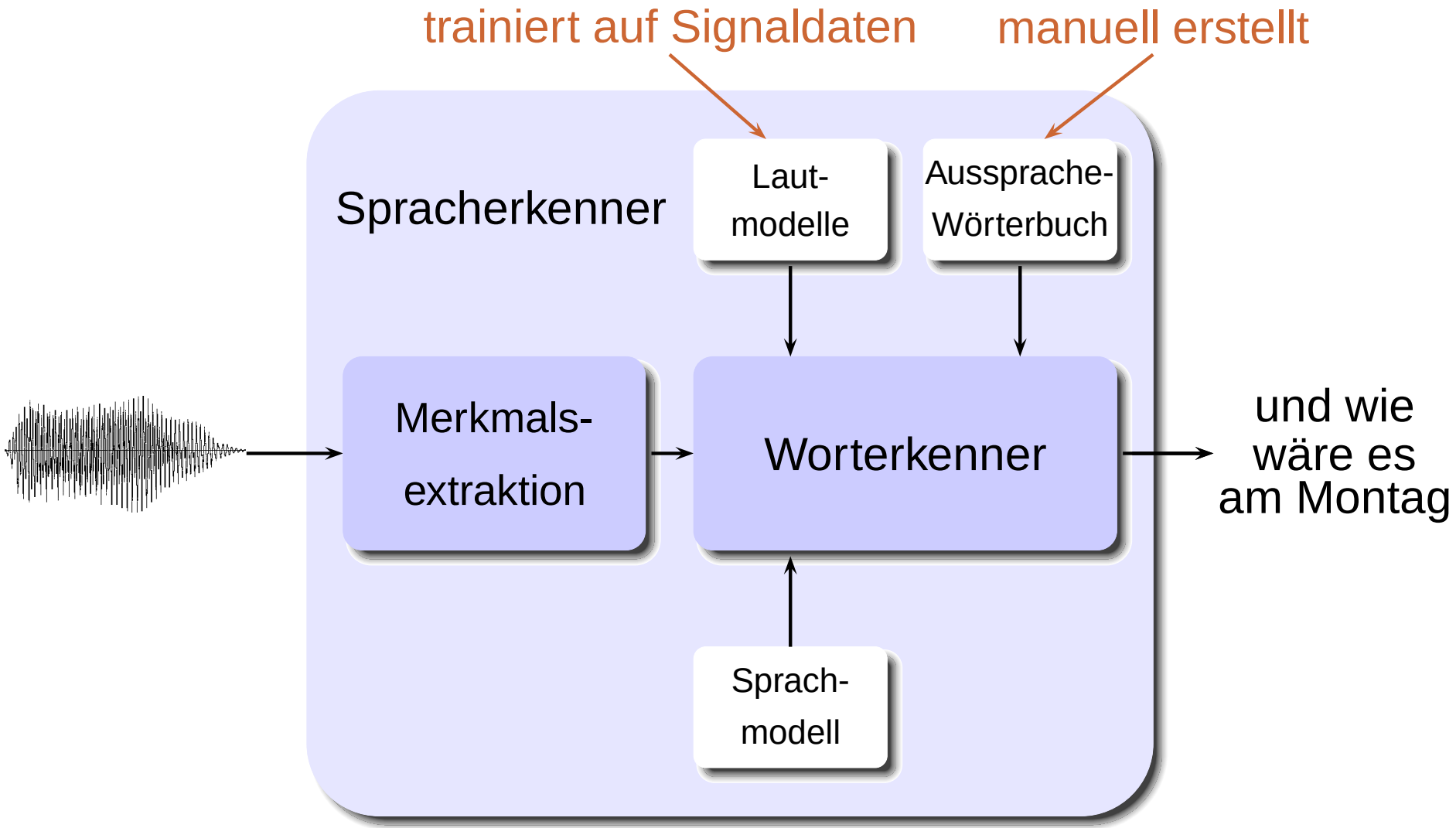
- Verkettung von Lautmodellen zu Wortmodellen

```
Mittwoch:
  sp-m+i m-i+t i-t+v t-v+o ...
```

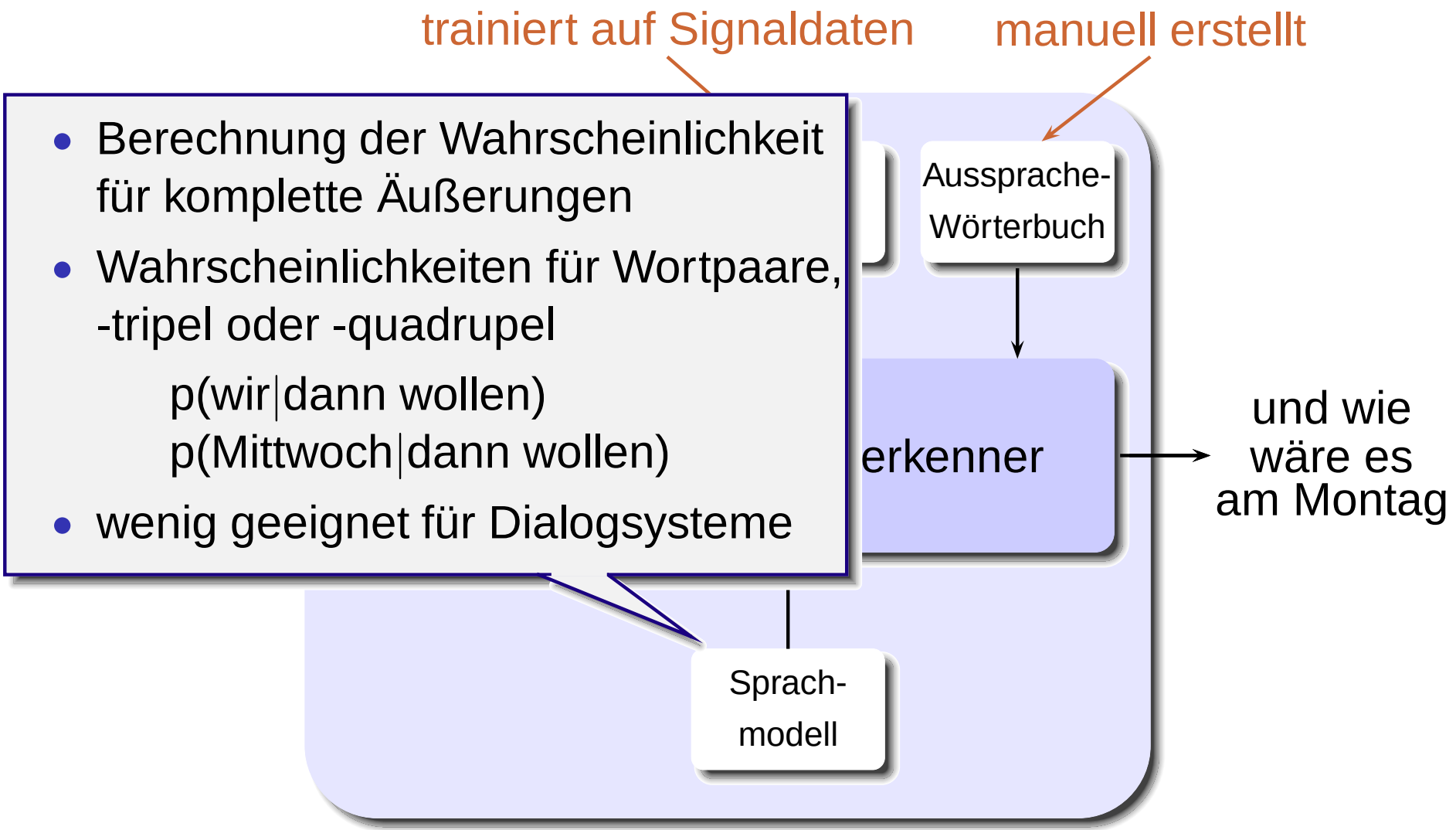
Spracherkennung



Spracherkennung



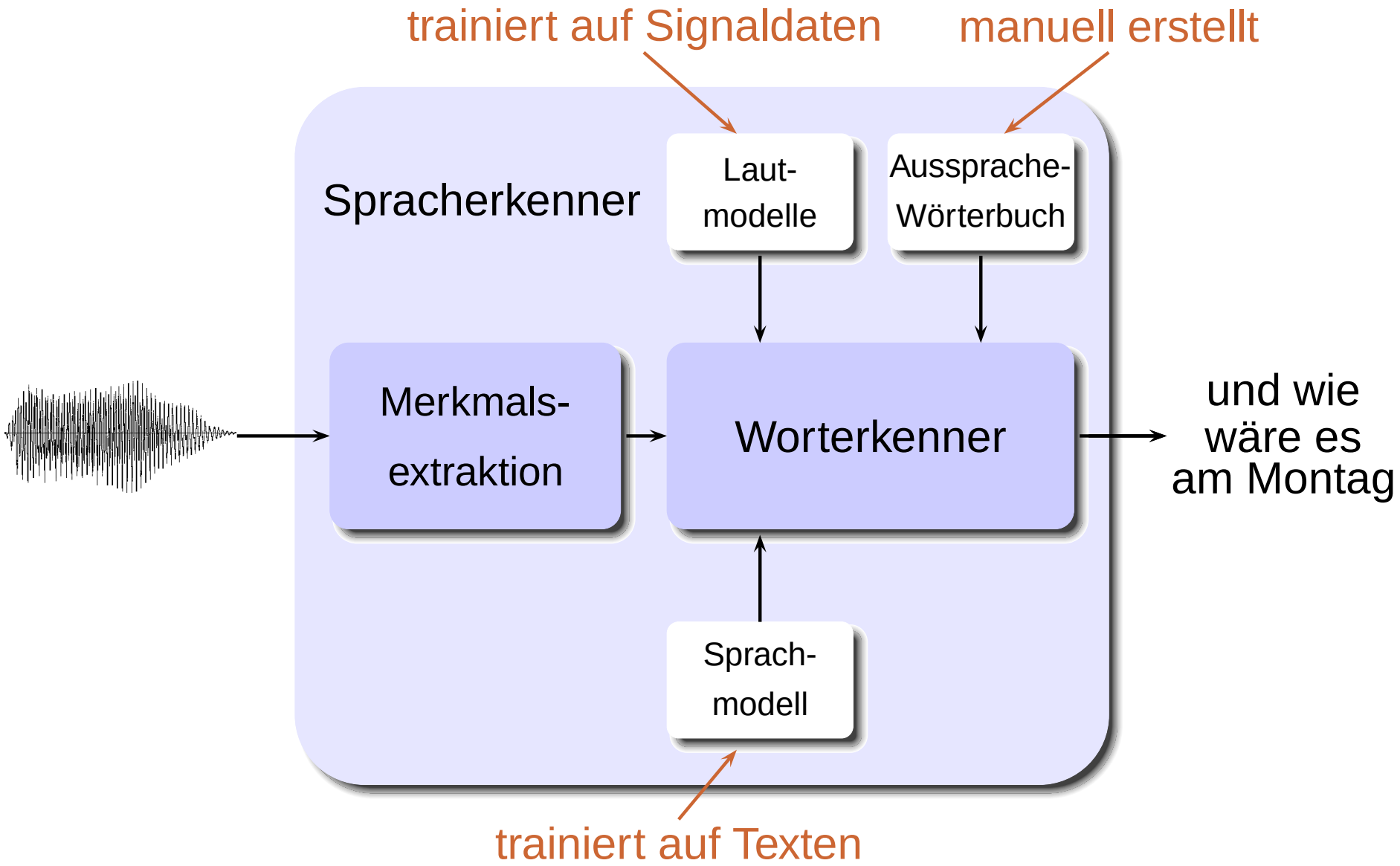
Spracherkennung



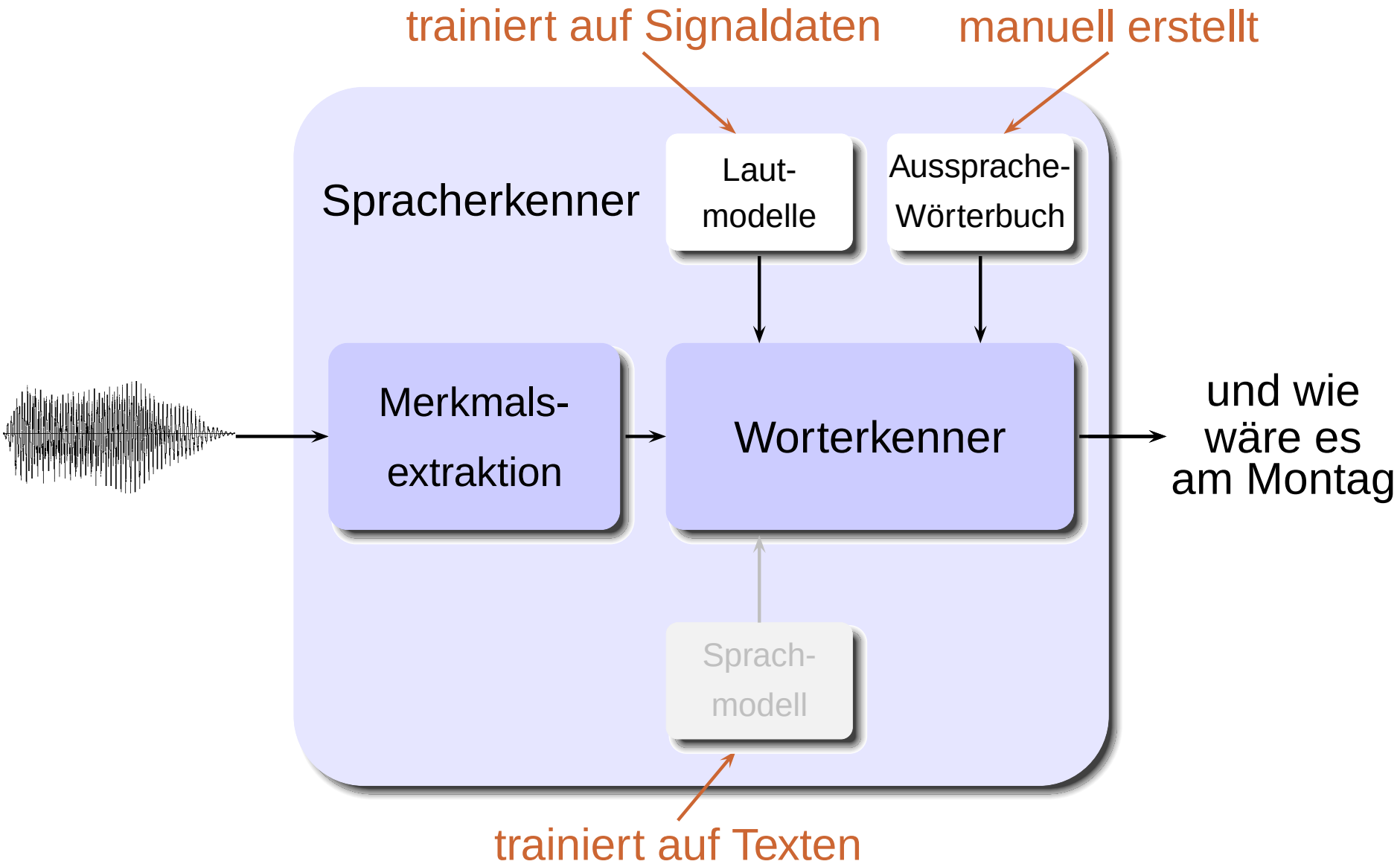
- Berechnung der Wahrscheinlichkeit für komplette Äußerungen
- Wahrscheinlichkeiten für Wortpaare, -tripel oder -quadrupel

  p(wir|dann wollen)
  p(Mittwoch|dann wollen)
- wenig geeignet für Dialogsysteme

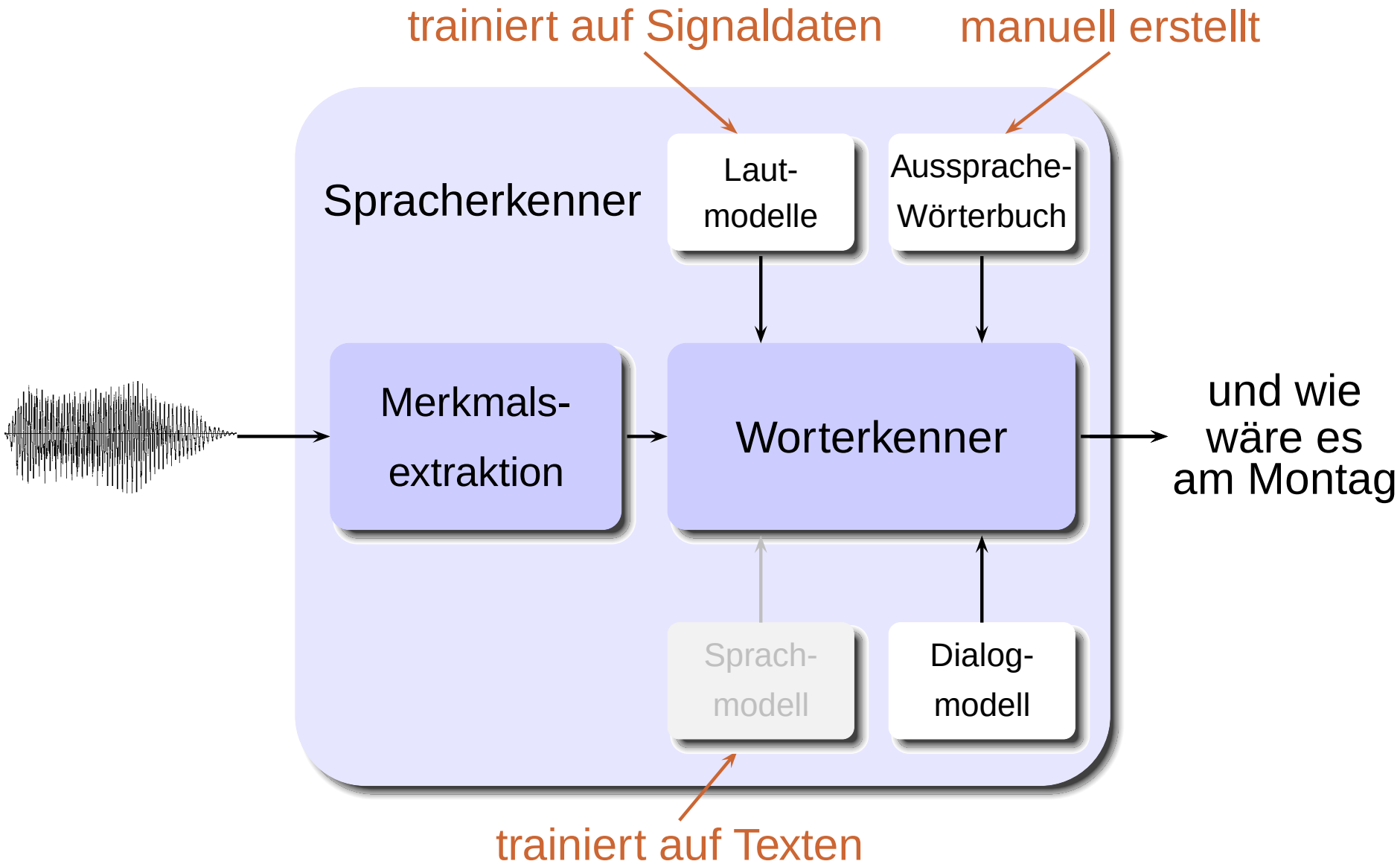
Spracherkennung



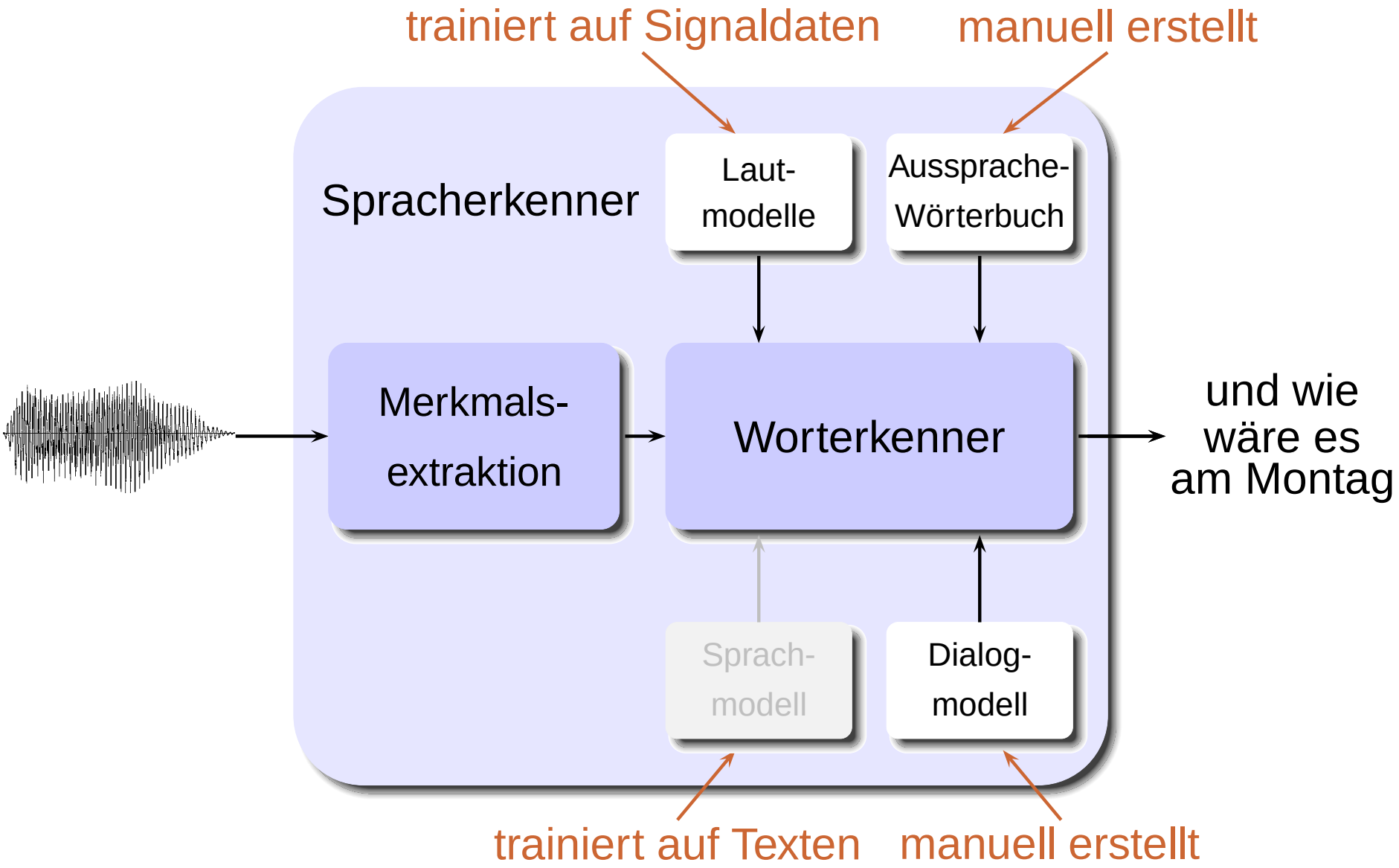
Spracherkennung



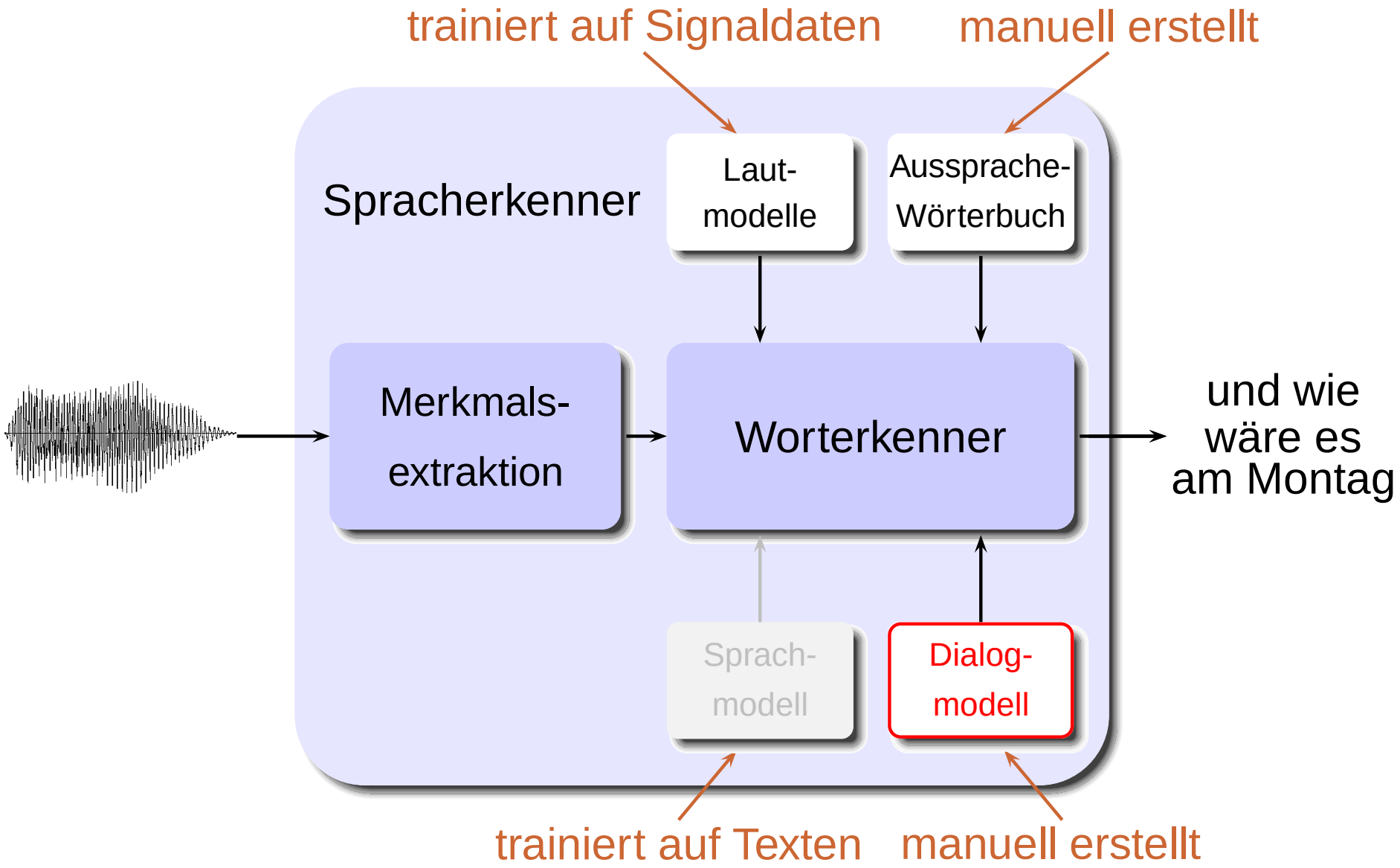
Spracherkennung



Spracherkennung



Spracherkennung



Dialogmodellierung

- dynamische Einschränkung des Erkennerwortschatzes in Abhängigkeit vom Dialogzustand

Dialogmodellierung

- dynamische Einschränkung des Erkennerwortschatzes in Abhängigkeit vom Dialogzustand
- Wozu braucht man das?

Dialogmodellierung

- dynamische Einschränkung des Erkennerwortschatzes in Abhängigkeit vom Dialogzustand
- Wozu braucht man das?
  - Erkennungssicherheit erhöhen
    $\longrightarrow$ Was wurde gesagt?

Dialogmodellierung

- dynamische Einschränkung des Erkennerwortschatzes in Abhängigkeit vom Dialogzustand
- Wozu braucht man das?
  - Erkennungssicherheit erhöhen
    $\longrightarrow$ Was wurde gesagt?
    - ähnliche Aussprache: *Mai* oder *drei*?
    - verschiedene Sprecher
    - schlechte Übertragungsqualität

Dialogmodellierung

- dynamische Einschränkung des Erkennerwortschatzes in Abhängigkeit vom Dialogzustand
- Wozu braucht man das?
  - Erkennungssicherheit erhöhen
    $\longrightarrow$ Was wurde gesagt?
    - ähnliche Aussprache: *Mai* oder *drei*?
    - verschiedene Sprecher
    - schlechte Übertragungsqualität
  - semantische Interpretation erleichtern
    $\longrightarrow$ Was wird von der Maschine erwartet?

Dialogmodellierung

- dynamische Einschränkung des Erkennerwortschatzes in Abhängigkeit vom Dialogzustand
- Wozu braucht man das?
  - Erkennungssicherheit erhöhen
    $\rightarrow$ Was wurde gesagt?
    - ähnliche Aussprache: *Mai* oder *drei*?
    - verschiedene Sprecher
    - schlechte Übertragungsqualität
  - semantische Interpretation erleichtern
    $\rightarrow$ Was wird von der Maschine erwartet?
    - Zahl $\rightarrow$ Geldbetrag, Uhrzeit, Datum, Kontonummer, ...
    - Mehrdeutigkeit: *Wann werden S/sie kommen?*
    - Referenzauflösung: Was wird durch *sie* bezeichnet?

Dialogmodellierung

- dynamische Einschränkung des Erkennerwortschatzes in Abhängigkeit vom Dialogzustand
- Wozu braucht man das?
  - Erkennungssicherheit erhöhen
    $\rightarrow$ Was wurde gesagt?
    - ähnliche Aussprache: *Mai* oder *drei*?
    - verschiedene Sprecher
    - schlechte Übertragungsqualität
  - semantische Interpretation erleichtern
    $\rightarrow$ Was wird von der Maschine erwartet?
    - Zahl $\rightarrow$ Geldbetrag, Uhrzeit, Datum, Kontonummer, ...
    - Mehrdeutigkeit: *Wann werden S/sie kommen?*
    - Referenzauflösung: Was wird durch *sie* bezeichnet?
  - eine Maschine hat keinen gesunden Menschenverstand!

Dialogmodellierung

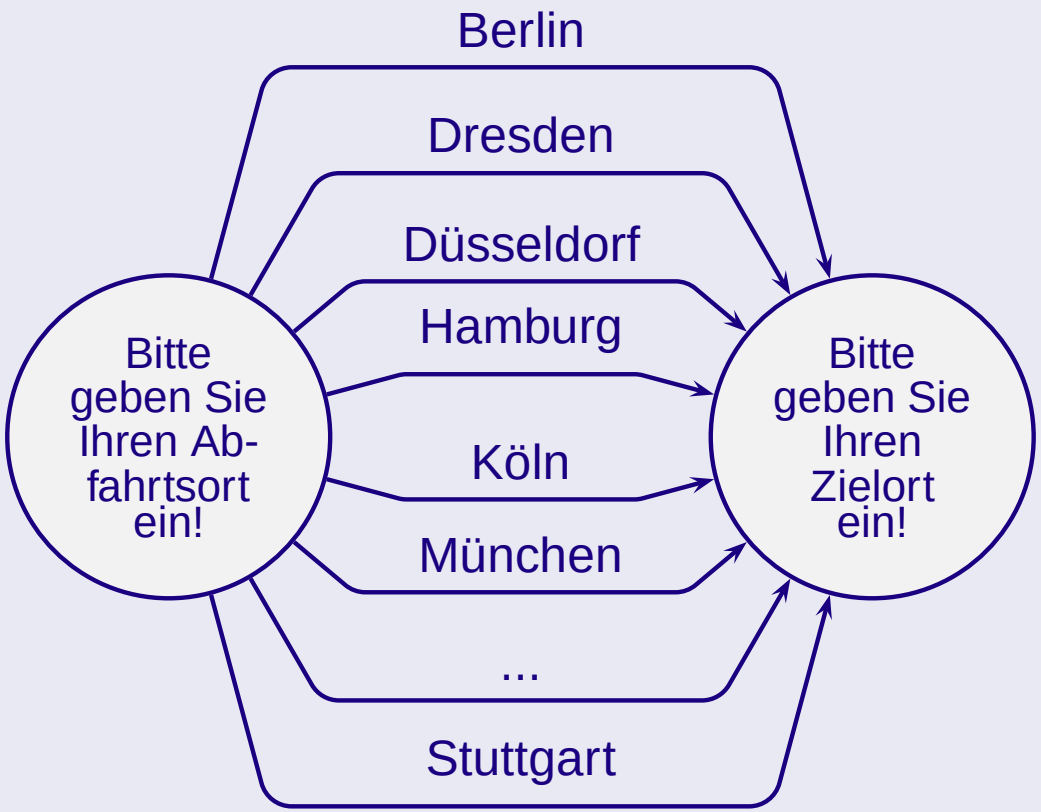
- Dialogzustände: Aufforderung zur Eingabe (Prompt)
- Übergänge zwischen Dialogzuständen: Erkennung von Nutzeräußerungen

Dialogmodellierung

- Dialogzustände: Aufforderung zur Eingabe (Prompt)
- Übergänge zwischen Dialogzuständen: Erkennung von Nutzeräußerungen

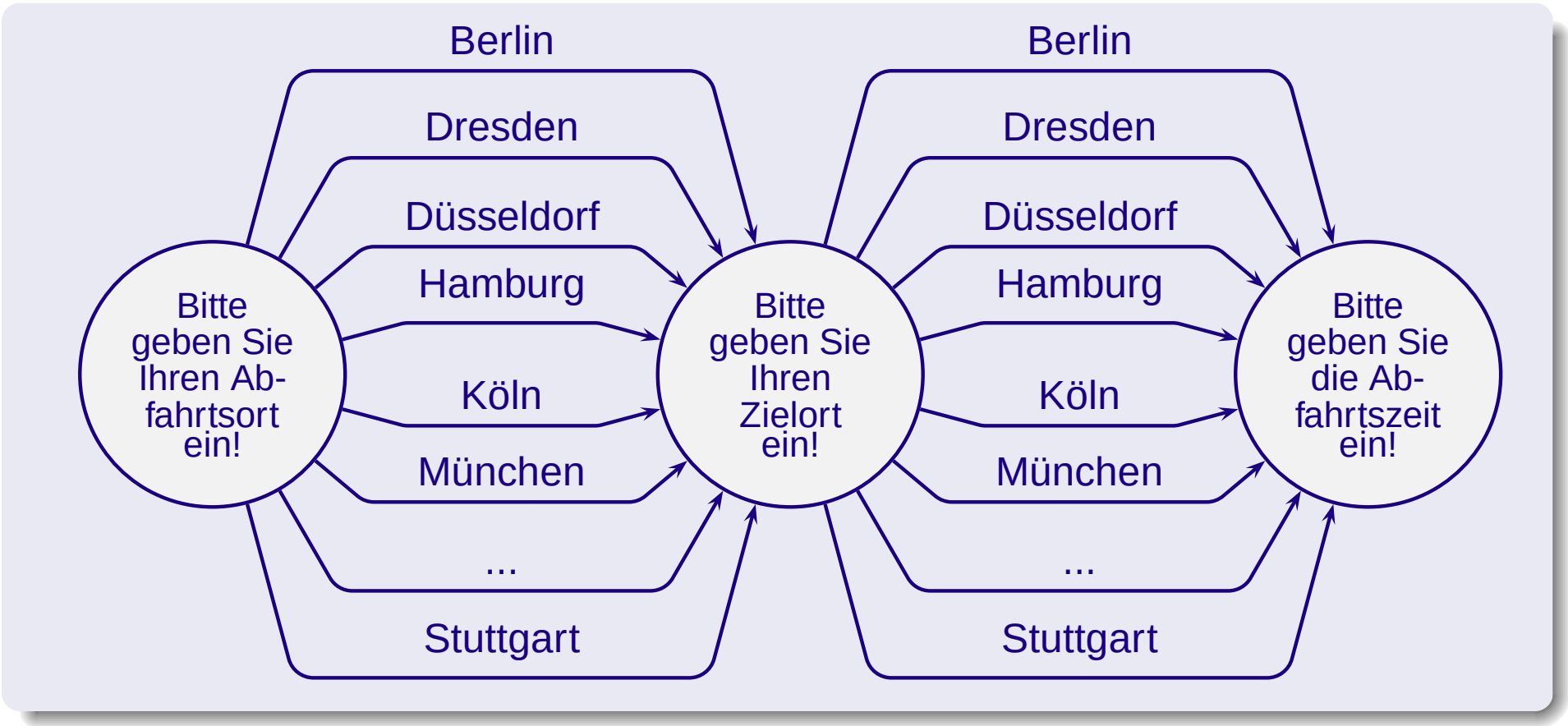
Dialogmodellierung

- Dialogzustände: Aufforderung zur Eingabe (Prompt)
- Übergänge zwischen Dialogzuständen: Erkennung von Nutzeräußerungen



Dialogmodellierung

- Dialogzustände: Aufforderung zur Eingabe (Prompt)
- Übergänge zwischen Dialogzuständen: Erkennung von Nutzeräußerungen



Dialogmodellierung

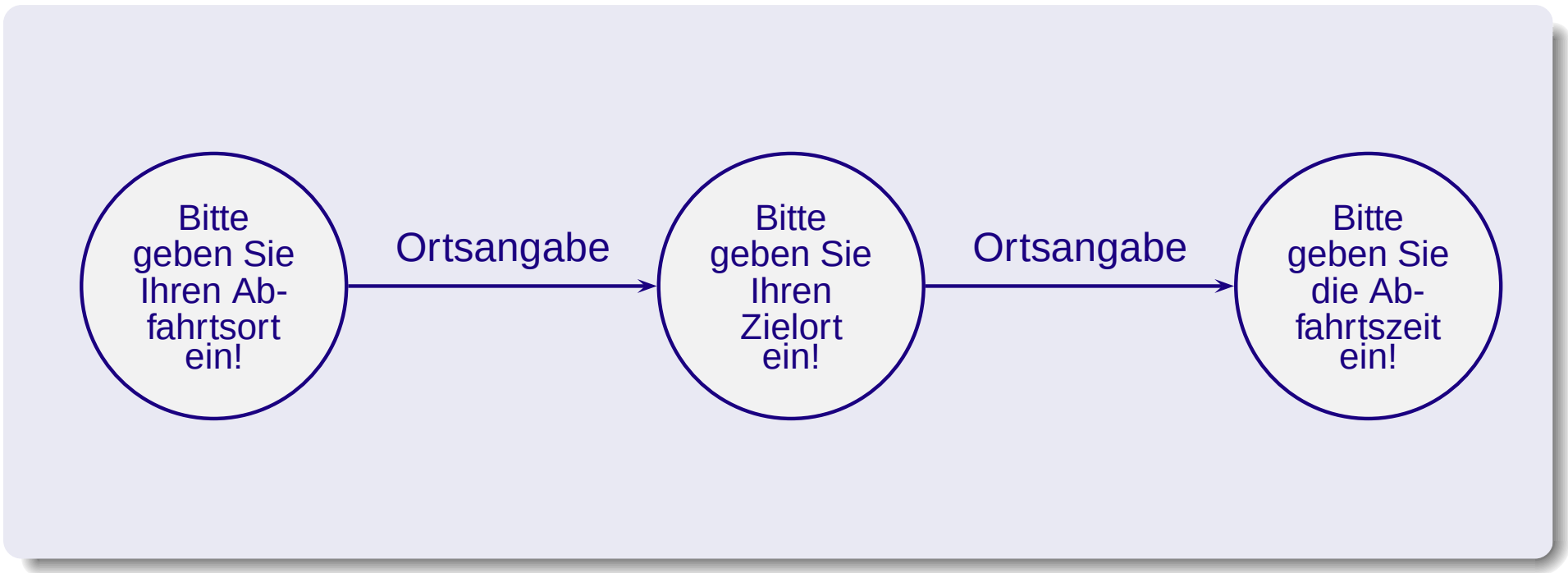
- Mehrfachverwendung von Teilnetzen

Dialogmodellierung

- Mehrfachverwendung von Teilnetzen

Dialogmodellierung

- Mehrfachverwendung von Teilnetzen



Dialogmodellierung

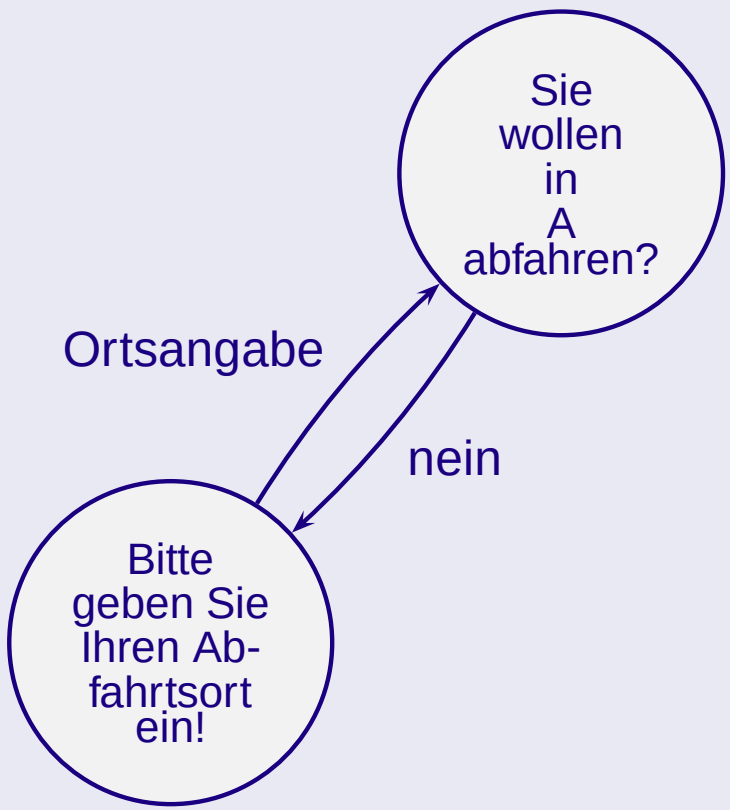
- sprecherunabhängige Spracherkennung ist unsicher
- insbesondere bei Telefoneingabe
- Erhöhen der Zuverlässigkeit durch Rückfragen

Dialogmodellierung

- sprecherunabhängige Spracherkennung ist unsicher
- insbesondere bei Telefoneingabe
- Erhöhen der Zuverlässigkeit durch Rückfragen

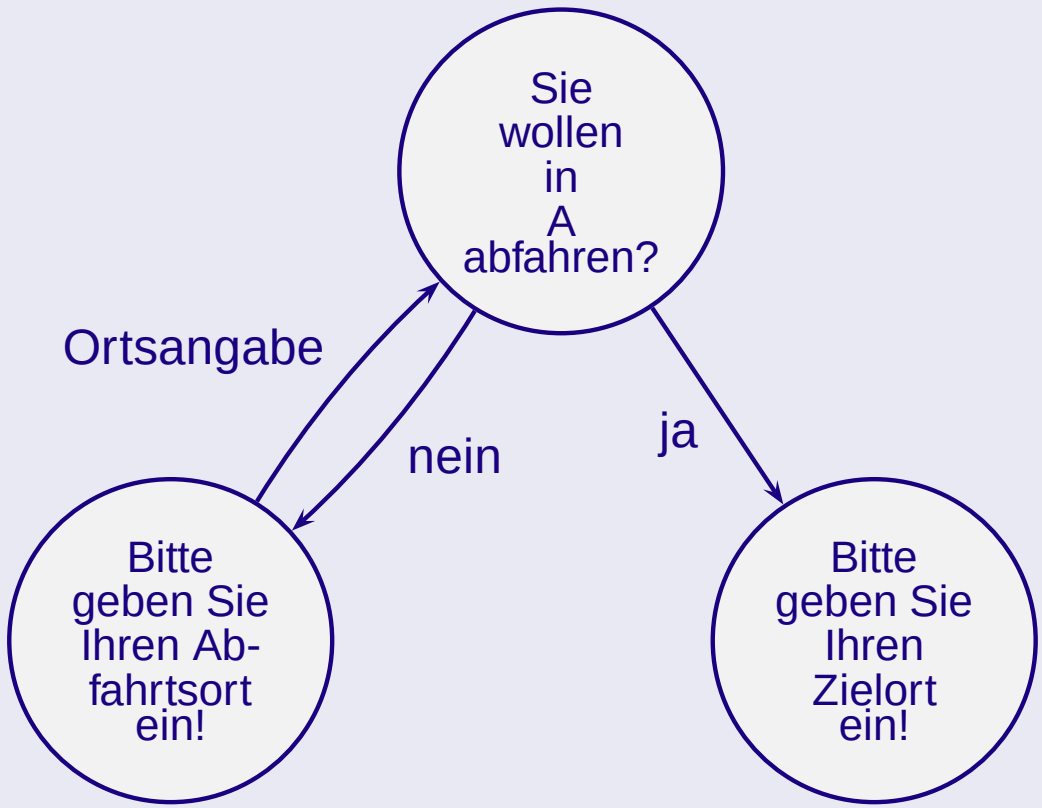
Dialogmodellierung

- sprecherunabhängige Spracherkennung ist unsicher
- insbesondere bei Telefoneingabe
- Erhöhen der Zuverlässigkeit durch Rückfragen



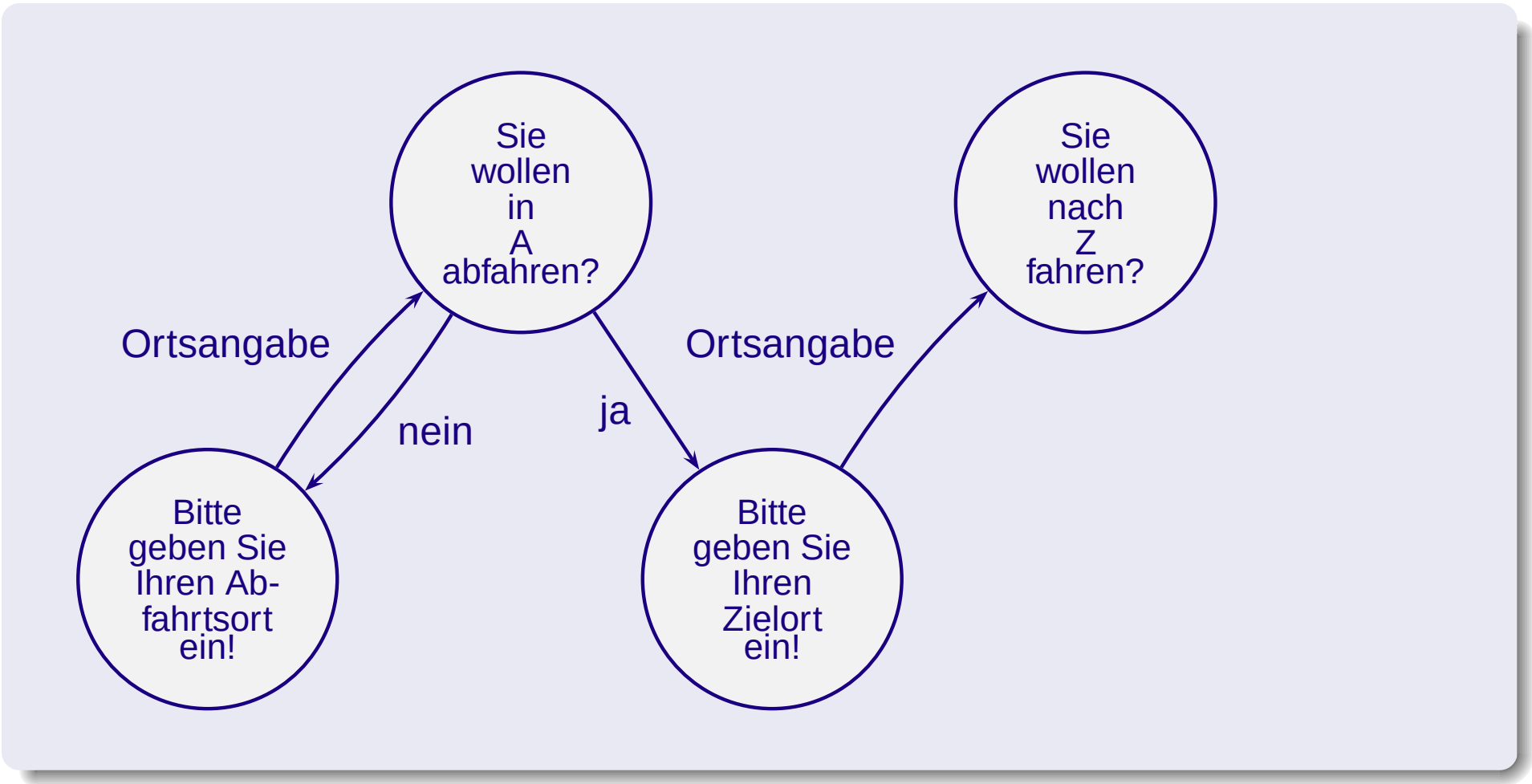
Dialogmodellierung

- sprecherunabhängige Spracherkennung ist unsicher
- insbesondere bei Telefoneingabe
- Erhöhen der Zuverlässigkeit durch Rückfragen



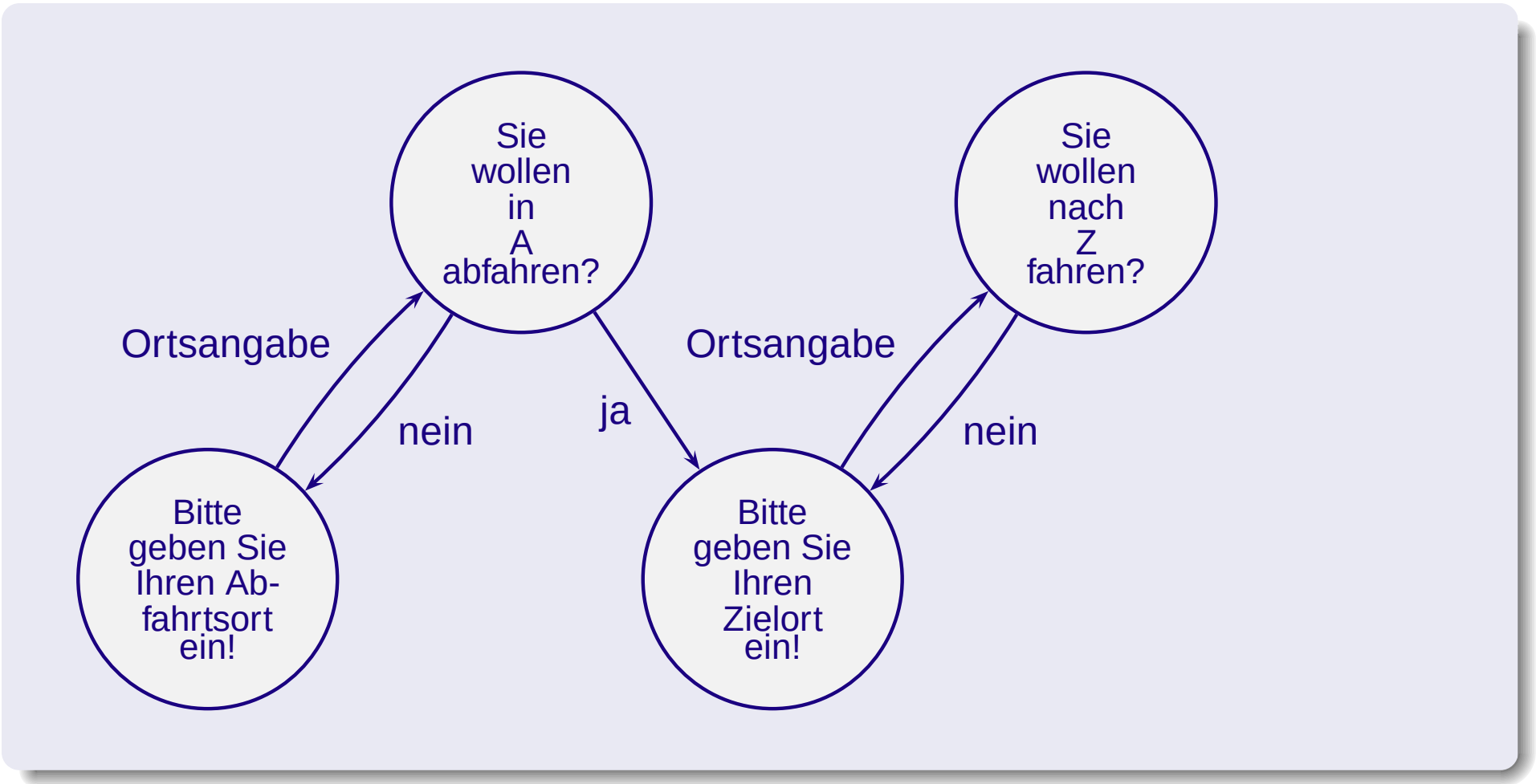
Dialogmodellierung

- sprecherunabhängige Spracherkennung ist unsicher
- insbesondere bei Telefoneingabe
- Erhöhen der Zuverlässigkeit durch Rückfragen



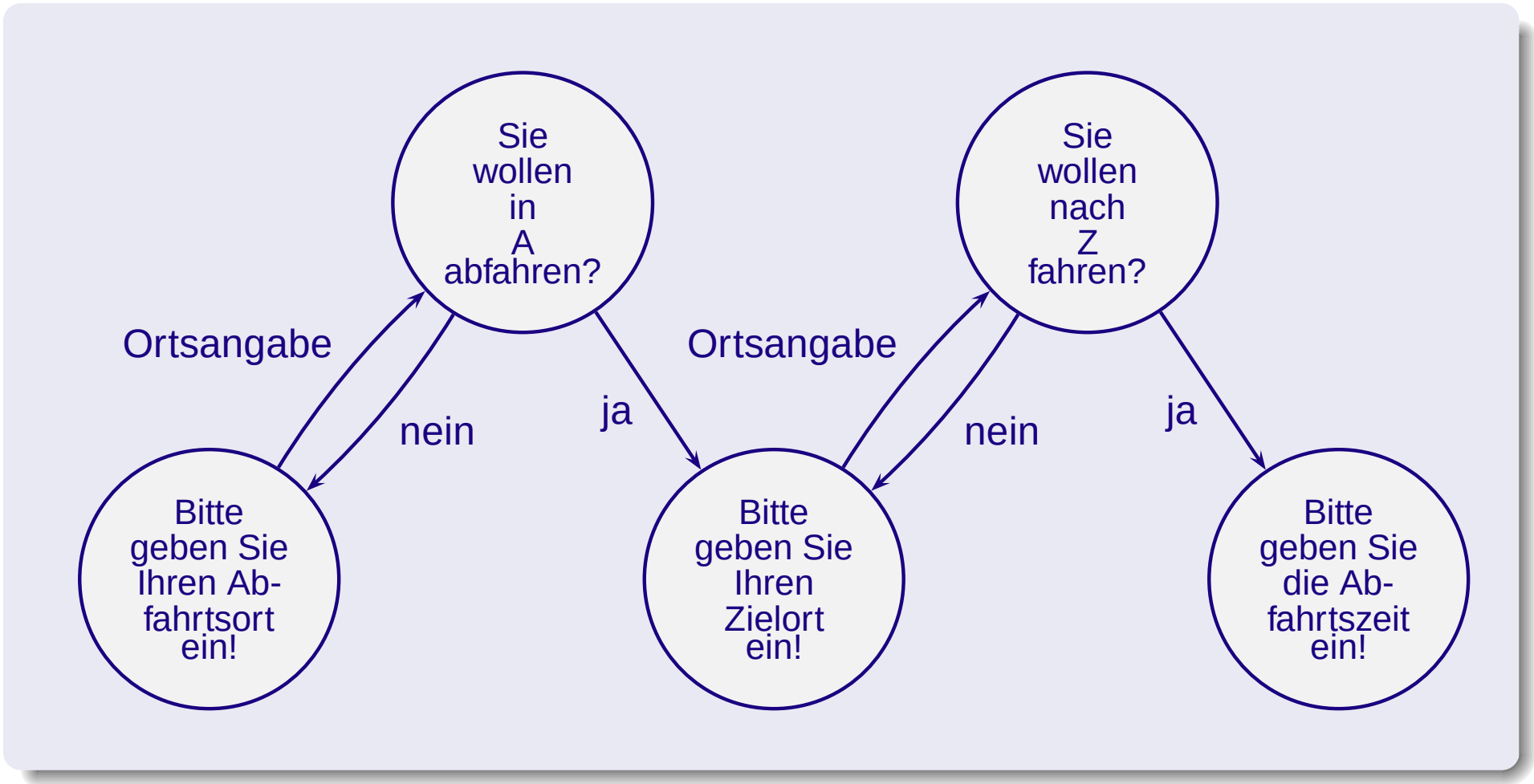
Dialogmodellierung

- sprecherunabhängige Spracherkennung ist unsicher
- insbesondere bei Telefoneingabe
- Erhöhen der Zuverlässigkeit durch Rückfragen



Dialogmodellierung

- sprecherunabhängige Spracherkennung ist unsicher
- insbesondere bei Telefoneingabe
- Erhöhen der Zuverlässigkeit durch Rückfragen



Dialogmodellierung

- theoretische Grundlage: deterministischer endlicher Automat

Dialogmodellierung

- theoretische Grundlage: deterministischer endlicher Automat
  - einfachstes Automatenmodell der Informatik
  - effiziente Implementierung
  - gute Vorhersagefähigkeit $\rightarrow$ starke Einschränkung des aktiven Wortschatzes

Dialogmodellierung

- theoretische Grundlage: deterministischer endlicher Automat
  - einfachstes Automatenmodell der Informatik
  - effiziente Implementierung
  - gute Vorhersagefähigkeit $\rightarrow$ starke Einschränkung des aktiven Wortschatzes
- für natürliche Dialogführung zu rigide $\rightarrow$ Erweiterungen nötig

Dialogmodellierung

- theoretische Grundlage: deterministischer endlicher Automat
  - einfachstes Automatenmodell der Informatik
  - effiziente Implementierung
  - gute Vorhersagefähigkeit $\rightarrow$ starke Einschränkung des aktiven Wortschatzes

- für natürliche Dialogführung zu rigide $\rightarrow$ Erweiterungen nötig
  - wechselnde Prompts
  - ”Hineinreden” in den Prompt (barge in)
  - ...

Dialogmodellierung

- Dialogmodellierung erfordert Vorhersehen möglicher Nutzerreaktionen
  - Hineinversetzen in den Nutzer
  - Wizard-of-Oz-Experimente

Dialogmodellierung

- Dialogmodellierung erfordert Vorhersehen möglicher Nutzerreaktionen
  - Hineinversetzen in den Nutzer
  - Wizard-of-Oz-Experimente
- Dialogmodellierung schränkt die sprachliche Möglichkeiten eines Nutzers stark ein
  - Lenkung des Nutzers durch Vorgabe zulässiger Äußerungen (akustisches Menü)