

Sprachverarbeitung II / 12 FS 2017

Wissensbasierte Sprachmodellierung

Buch: Kapitel 14.3

Beat Pfister



Programm heute

Vorlesung: Sprachmodellierung (Teil 2)

- Stärken und Schwächen statistischer Sprachmodelle
- Wissensbasierter Ansatz der Sprachmodellierung
 - Einsatz linguistischen Wissens in der Spracherkennung
 - formale Grammatiken und Parsing in der Spracherkennung

Übung: keine

Statistisches Sprachmodell

Allgemein: $P(W_1^K)$

Zerlegung: $P(W_1^K) = P(w_1) \cdot P(w_2|w_1) \cdots P(w_K|w_1 \dots w_{K-1})$

Approximation: $P(w_k|w_1 \dots w_{k-1}) \approx P(w_k|w_{k-N+1} \dots w_{k-1})$

→ N-Gram-Sprachmodell

Stärken von N-Gram-Sprachmodellen

- Training mit tatsächlich gebrauchten Sätzen
 - + N-Gram beschreiben die effektiv gebrauchte Sprache
(enthalten syntaktisches, semantisches und pragmatisches Wissen)
- Wissen wird automatisch aus Daten gelernt
 - + kein linguistisches Wissen über die Sprache nötig
 - + für beliebige Sprachen anwendbar (sofern Text vorhanden)
- einfach in Erkennner mit HMM integrierbar
 - + effiziente Erkennung mit Viterbi-Algorithmus

Schwächen von N-Gram-Sprachmodellen

- Modellschwäche: nur für kleine Werte von N praktikabel
 - Kontext beschränkt
- mangelnde linguistische Relevanz
 - ungesehene und inkorrekte Wortfolgen nicht unterscheidbar
- Überbewertung des Sprachmodells (z.B. Müller ↔ Möller)
 - Sprachmodell kann akustisches Modell überstimmen
- Trainingstext mindestens 50 Mio. Wörter (Buch SPV etwa 110'000)
 - praktisch sehr aufwändig

>>>

Argumente für die wissensbasierte Sprachmodellierung

- Der Grossteil des Wissens über eine natürliche Sprache
 - ist entweder lexikalisch (Aussprachelexikon, Morphemlexikon, ...)
 - oder hat die Form von Regeln (insbes. Morphologie und Syntax)
- Auch noch nie formulierte Sätze können analysiert und damit als zur Sprache gehörig erkannt werden.
- Die Schwächen der statistischen Modelle:
 - zu eingeschränkter Kontext
 - ungewöhnliche und inkorrekte Wortfolgen nicht unterscheidbar
 - Überbewertung des Sprachmodells

Linguistisches Wissen zur Spracherkennung

Linguistisches Wissen, das für die Spracherkennung eingesetzt werden kann, gibt es unter anderem auf der

- akustisch-phonetischen Ebene
- phonemischen Ebene
- morphologischen Ebene
- syntaktischen Ebene

Linguistisches Wissen zur Spracherkennung

Linguistisches Wissen, das für die Spracherkennung eingesetzt werden kann, gibt es unter anderem auf der

- akustisch-phonetischen Ebene
- phonemischen Ebene
- morphologischen Ebene
- syntaktischen Ebene

Akustisch-phonetische Ebene

- Inventar der zu unterscheidenden / erkennenden Laute
- Satz distinktiver Merkmale

stimmhaft \leftrightarrow stimmlos

nasal \leftrightarrow nicht nasal

kurz \leftrightarrow lang

gerundet \leftrightarrow ungerundet

hinten \leftrightarrow vorn

vokalisch \leftrightarrow konsonatisch

⋮
⋮
⋮

Linguistisches Wissen zur Spracherkennung

Linguistisches Wissen, das für die Spracherkennung eingesetzt werden kann, gibt es unter anderem auf der

- akustisch-phonetischen Ebene
- phonemischen Ebene
- morphologischen Ebene
- syntaktischen Ebene

Phonemische Ebene

- Inventar der Phoneme und Allophone
- Regeln zur Beschreibung von Sprechvarianten
(viel häufiger als freie Allophone)

	normal:	schnell:	Regel:
anbinden	[anbindən]	[ambindən]	[nb] → [mb]
ein Bier	[aɪ̯n bɪ̯ɐ̯]	[aɪ̯m bɪ̯ɐ̯]	[nb] → [mb]
am Mittag	[am mitak]	[amitak]	[mm] → [m]
mitkommen	['mitkɔ̯mən]	['mɪkɔ̯mən]	[tk] → [k]
Tagesschau	[ta:gəsʃau̯]	[ta:gəʃau̯]	[sʃ] → [ʃ]

Linguistisches Wissen zur Spracherkennung

Linguistisches Wissen, das für die Spracherkennung eingesetzt werden kann, gibt es unter anderem auf der

- akustisch-phonetischen Ebene
- phonemischen Ebene
- morphologischen Ebene
- syntaktischen Ebene

Morphologische Ebene

Das Inventar der zu erkennenden Wörter (Vokabular) jedes Spracherkenners muss gegeben sein

- direkt als Liste der Wörter (Vollformenlexikon) oder
- indirekt in der Form eines Morphemlexikons und einer Wortgrammatik (erzeugendes System)

Linguistisches Wissen zur Spracherkennung

Linguistisches Wissen, das für die Spracherkennung eingesetzt werden kann, gibt es unter anderem auf der

- akustisch-phonetischen Ebene
- phonemischen Ebene
- morphologischen Ebene
- **syntaktischen Ebene**

Syntaktische Ebene

Die Satzgrammatik definiert zusammen mit dem Vokabular die Menge der zu erkennenden Sätze.

Einsatz linguistischen Wissens in der Spracherkennung

Für die Spracherkennung wird linguistisches Wissen in expliziter Form (z.B. als Regeln und Lexika) eingesetzt

- zur Konzeption des Spracherkenners
- in der Konfigurationsphase
- für die eigentliche Erkennung
- in der Nachverarbeitung

Einsatz linguistischen Wissens in der Spracherkennung

Für die Spracherkennung wird linguistisches Wissen in expliziter Form (z.B. als Regeln und Lexika) eingesetzt

- zur Konzeption des Spracherkenners
- in der Konfigurationsphase
- für die eigentliche Erkennung
- in der Nachverarbeitung

Linguistisches Wissen in der Konzeptionsphase

Linguistisches Wissen verschiedener Ebenen kann bei der Konzeption eines Spracherkenners zum Einsatz kommen:

- beim Festlegen der zu modellierenden Einheiten
(Art der Einheiten: Laute, Diphone . . . ; Inventar der Einheiten)
- zur Definition des Vokabulars, z.B. Kommandosprache
 - Semantik grundsätzlich frei definierbar
(besser: semantische Übereinstimmung mit allgemeinem Sprachgebrauch)
 - akustische Unterscheidbarkeit der Wörter

Einsatz linguistischen Wissens in der Spracherkennung

Für die Spracherkennung wird linguistisches Wissen in expliziter Form (z.B. als Regeln und Lexika) eingesetzt

- zur Konzeption des Spracherkenners
- in der Konfigurationsphase
- für die eigentliche Erkennung
- in der Nachverarbeitung

Linguistisches Wissen in der Konfigurationsphase

- Erzeugen der Wortmodelle aus einem vollständigen Satz von Subwortmodellen und einem Aussprache-Lexikon
- Generieren des Erkennernetzwerkes
 - aus Vokabular und Grammatik alle Sätze erzeugen
 - Erkennernetzwerk so, dass jeder korrekte Satz einem Pfad vom *START*- zum *END*-Knoten entspricht (Viterbi-Algorithmus)
- Generieren von Aussprachevarianten mittels Regeln
 - Ergänzung des Aussprache-Lexikons
(enthält gewöhnlich nur die Standardaussprache)
 - Erweiterung des Erkennernetzwerkes

Einsatz linguistischen Wissens in der Spracherkennung

Für die Spracherkennung wird linguistisches Wissen in expliziter Form (z.B. als Regeln und Lexika) eingesetzt

- zur Konzeption des Spracherkenners
- in der Konfigurationsphase
- für die eigentliche Erkennung
- in der Nachverarbeitung

Linguistisches Wissen in der Erkennungsphase

In der eigentlichen Erkennung explizites linguistisches Wissen einzusetzen bedeutet:

- kein Erkennungsnetzwerk generieren, um Viterbi-Algorithmus anzuwenden,
- sondern direkt die Wort- und / oder Satzgrammatik und das Lexikon mit einem **Parser** einsetzen
- Aufgabe des Parsers: **aus dem Sprachsignal erkannte Elemente (z.B. Laute) unter Einhaltung der Grammatikregeln zu grösseren Elementen (Silben, Wörter, Konstituenten, Sätze) zusammenfassen und bewerten**
Satz mit bester Bewertung gilt als erkannt.

Einsatz linguistischen Wissens in der Spracherkennung

Für die Spracherkennung wird linguistisches Wissen in expliziter Form (z.B. als Regeln und Lexika) eingesetzt

- zur Konzeption des Spracherkenners
- in der Konfigurationsphase
- für die eigentliche Erkennung
- **in der Nachverarbeitung**

Linguistisches Wissen in der Nachverarbeitung

- wissensbasierte Spracherkennung (mit Parsing) ist sehr aufwändig
 - zuerst mit statistischem Spracherkenner N wahrscheinlichste Wortfolgen ermitteln (z.B. N -Best-Viterbi-Algorithmus)
- In diesem stark reduzierten Suchraum:
 - Auswahl der nach linguistischen Kriterien (z.B. syntaktische Korrektheit) am besten bewerteten Wortfolge
 - semantische Interpretation der Ausgabe des Erkenners (nur im Kontext einer bestimmten Anwendung)

Wissensbasierte Sprachmodelle

Statt statistische Beschreibung der Sprache (Wahrscheinlichkeit von Wortfolgen)

→ Einsatz formaler Grammatiken

Formale Grammatiken in der Spracherkennung

Absicht: Einbezug des Kriteriums der syntaktischen Korrektheit
→ vermeiden z.B. Übereinstimmungsfehler [">>>>](#)

Anforderung: Grammatik muss zwischen syntaktisch korrekten und falschen Wortfolgen unterscheiden können!

Probleme:

- Grammatik muss sehr präzise sein
- syntaktisch korrekt heisst aber noch nicht sinnvoll [">>>>](#)

Einsatz eines Parsers in der Spracherkennung

Grundsätzlich zwei Möglichkeiten:

A: Parser setzt aus dem Signal detektierte Lauthypothesen zu Morphemen, Wörtern, Konstituenten und Satz zusammen

>>>

B: Parser verarbeitet eine N-Best-Lösung, weil korrekte Wortfolge meistens in N-Best-Word-Lattice vorhanden

Einsatz eines Parsers in der Spracherkennung

Grundsätzlich zwei Möglichkeiten:

A: Parser setzt aus dem Signal detektierte Lauthypothesen zu Morphemen, Wörtern, Konstituenten und Satz zusammen

>>>

- Integration des Parsers in Erkenner (bidirektionale Kommunikation)

>>>

B: Parser verarbeitet eine N-Best-Lösung, weil korrekte Wortfolge meistens in N-Best-Word-Lattice vorhanden

Einsatz eines Parsers in der Spracherkennung

Grundsätzlich zwei Möglichkeiten:

A: Parser setzt aus dem Signal detektierte Lauthypothesen zu Morphemen, Wörtern, Konstituenten und Satz zusammen

>>>

- Integration des Parsers in Erkenner (bidirektionale Kommunikation)
- Idee gut – Realisation schwierig weil Parser zu wenig effizient

>>>

B: Parser verarbeitet eine N-Best-Lösung, weil korrekte Wortfolge meistens in N-Best-Word-Lattice vorhanden

Einsatz eines Parsers in der Spracherkennung

Grundsätzlich zwei Möglichkeiten:

A: Parser setzt aus dem Signal detektierte Lauthypothesen zu Morphemen, Wörtern, Konstituenten und Satz zusammen

>>>

- Integration des Parsers in Erkenner (bidirektionale Kommunikation)
- Idee gut – Realisation schwierig weil Parser zu wenig effizient

>>>

B: Parser verarbeitet eine N-Best-Lösung, weil korrekte Wortfolge meistens in N-Best-Word-Lattice vorhanden

>>>

Einsatz eines Parsers in der Spracherkennung

Grundsätzlich zwei Möglichkeiten:

A: Parser setzt aus dem Signal detektierte Lauthypothesen zu Morphemen, Wörtern, Konstituenten und Satz zusammen

>>>

- Integration des Parsers in Erkenner (bidirektionale Kommunikation)
- Idee gut – Realisation schwierig weil Parser zu wenig effizient

>>>

B: Parser verarbeitet eine N-Best-Lösung, weil korrekte Wortfolge meistens in N-Best-Word-Lattice vorhanden

>>>

- Parser bildet zweite Stufe im Erkenner

>>>

Einsatz eines Parsers in der Spracherkennung

Grundsätzlich zwei Möglichkeiten:

A: Parser setzt aus dem Signal detektierte Lauthypothesen zu Morphemen, Wörtern, Konstituenten und Satz zusammen

>>>

- Integration des Parsers in Erkenner (bidirektionale Kommunikation)
- Idee gut – Realisation schwierig weil Parser zu wenig effizient

>>>

B: Parser verarbeitet eine N-Best-Lösung, weil korrekte Wortfolge meistens in N-Best-Word-Lattice vorhanden

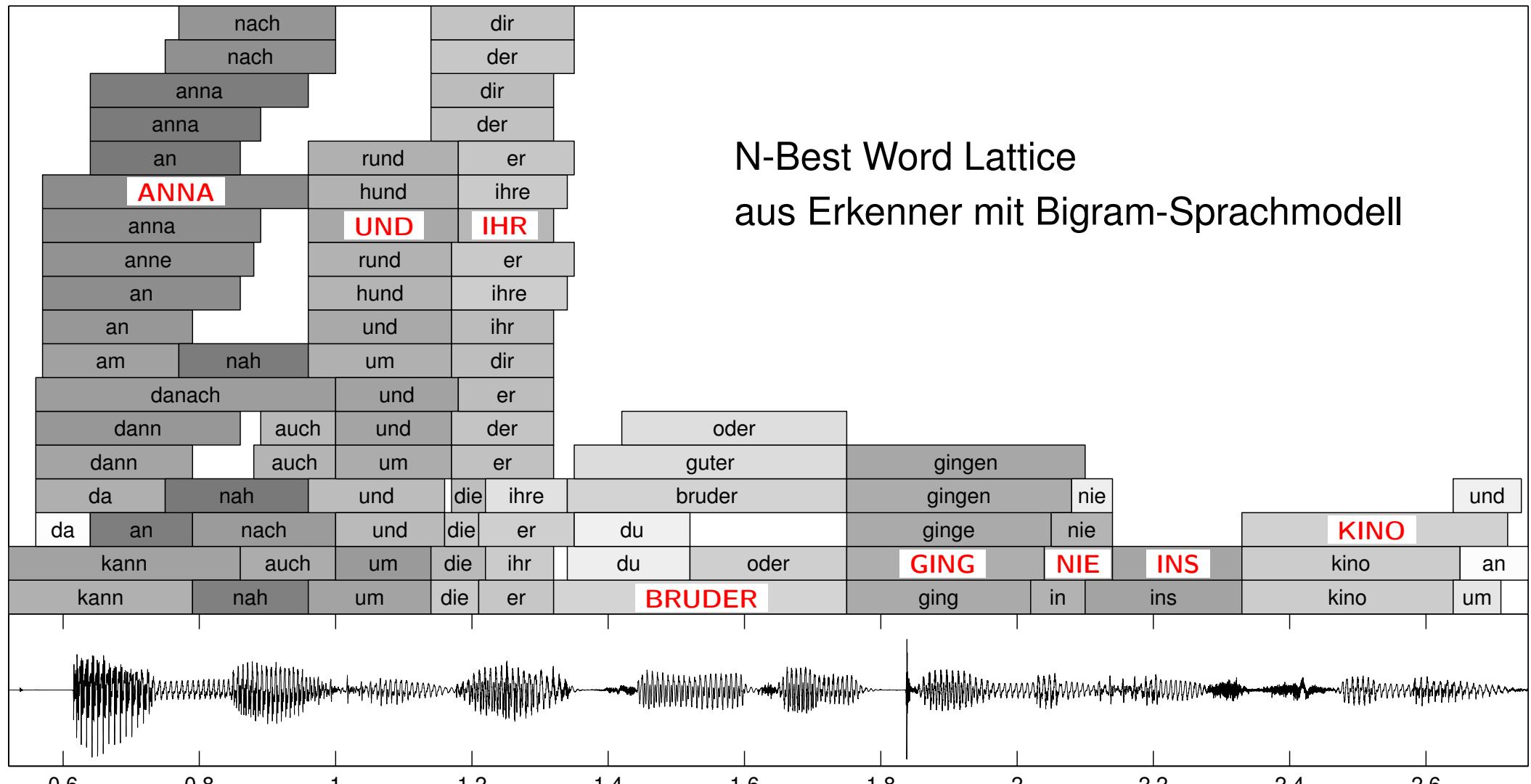
>>>

- Parser bildet zweite Stufe im Erkenner
- Höhere Bewertung syntaktisch korrekter Hypothesen

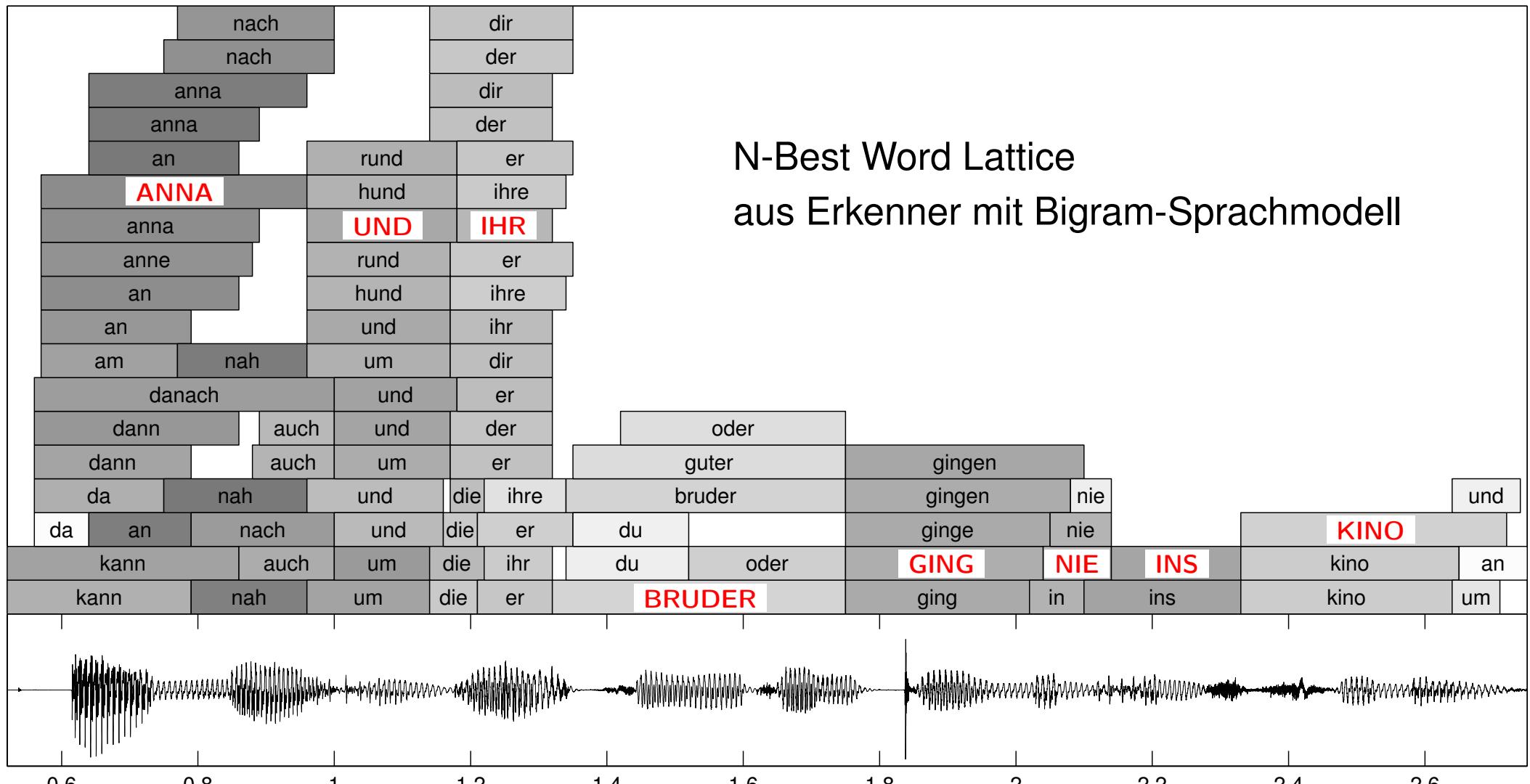
>>>

Ende der Vorlesung Sprachverarbeitung

Zur Übersicht der Vorlesung *Sprachverarbeitung II* [*>>>*](#)



N-Best Word Lattice
aus Erkenner mit Bigram-Sprachmodell



N-Best Word Lattice
aus Erkenner mit Bigram-Sprachmodell

Problem der Mehrdeutigkeit natürlicher Sprache

Ziel: **Syntaktische Korrektheit** als Kriterium verwenden

Fall A: Hypothese korrekt → kein grundsätzliches Problem
(mehr Aufwand für den Parser)

Fall B: Hypothese inkorrekt → **Kriterium ungenügend!**

>>>

<<<

Beispiel

Offensichtlich inkorrekte Hypothese (erkannte Wortfolge)

“scharon denkt nicht daran, dem rückzug der truppen aus einzuordnen”

ist effektiv syntaktisch korrekt!

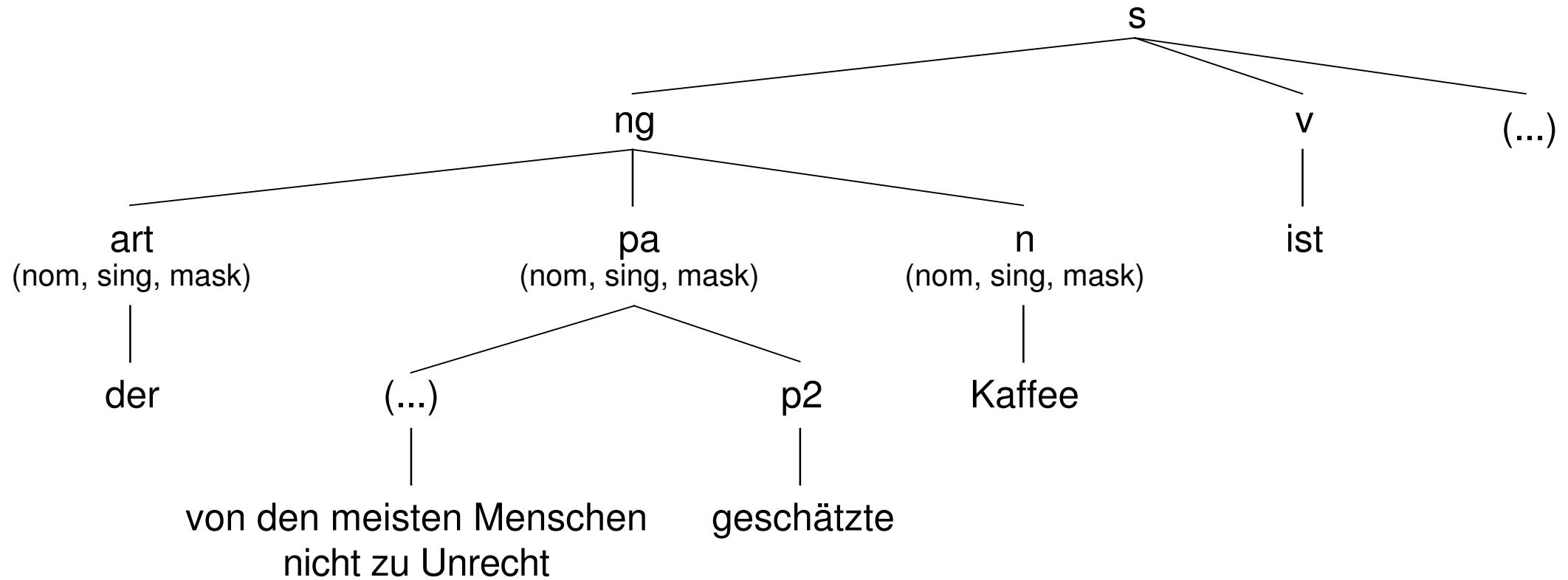
(korrekt wäre: “scharon denkt nicht daran, den rückzug der truppen anzuordnen”)

Erläuterungen: “der truppen aus” analog zu “des Pudels Kern”

“aus” kann auch ein Nomen sein wie in “Es droht bald das Aus.”

<<<

Beispiel einer komplexen Nominalgruppe mit KNG-Übereinstimmung



<<<

Unvollständig analysierbare Äusserungen

- Gründe:
- Äusserung tatsächlich ungrammatisch
 - Äusserung durch Grammatik nicht abgedeckt
 - Äusserung enthält unbekannte Wörter
 - Signal gestört:
 - durch Sprecher (zögern, räuspern, atmen, ...)
 - durch andere Stimmen
 - durch Fremdgeräusche

Folge:

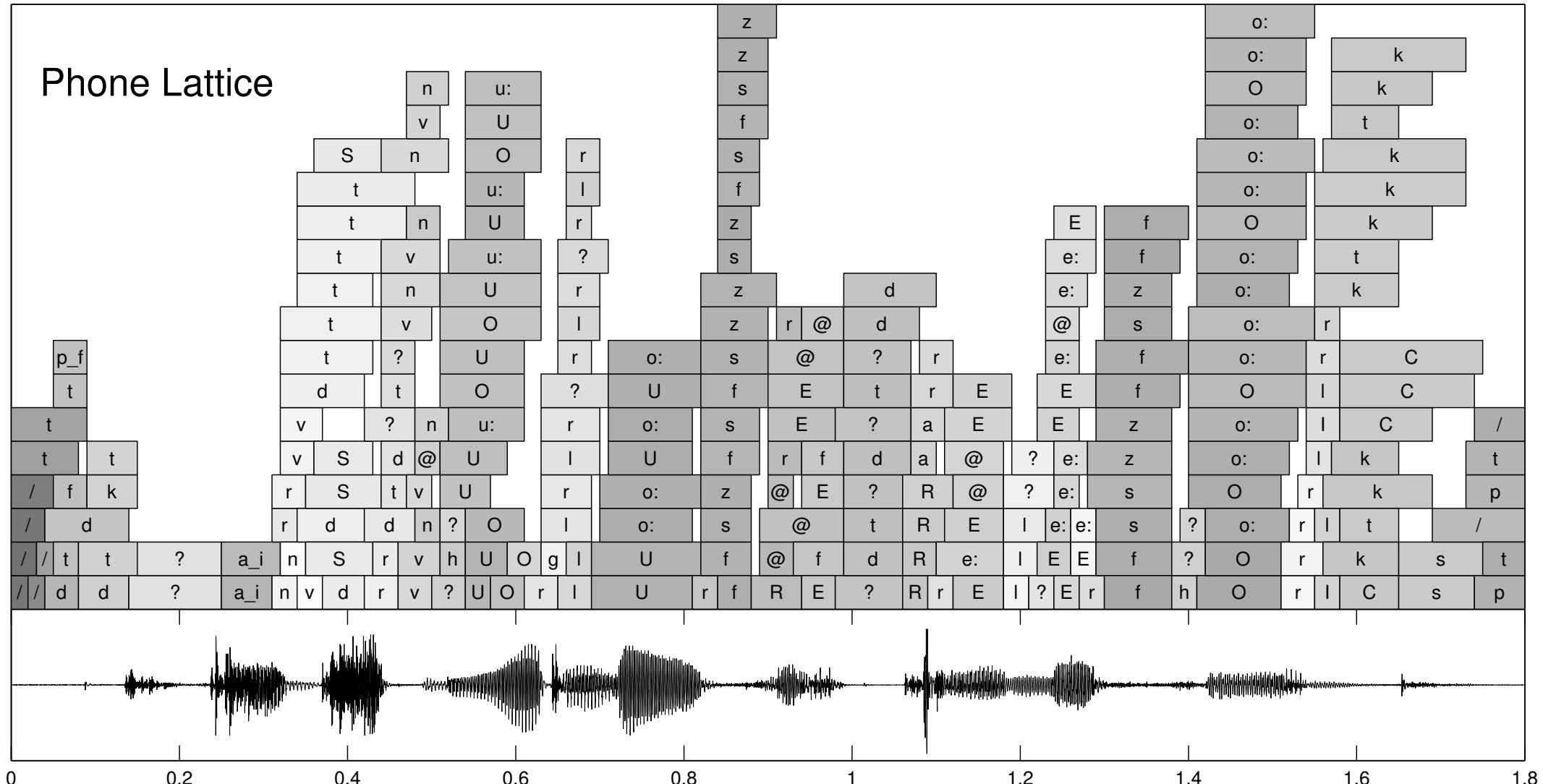
Parser findet keinen vollständigen Syntaxbaum

Lösung:

Ersatz-Syntaxbaum aus Teillösungen in der Chart zusammensetzen

<<<

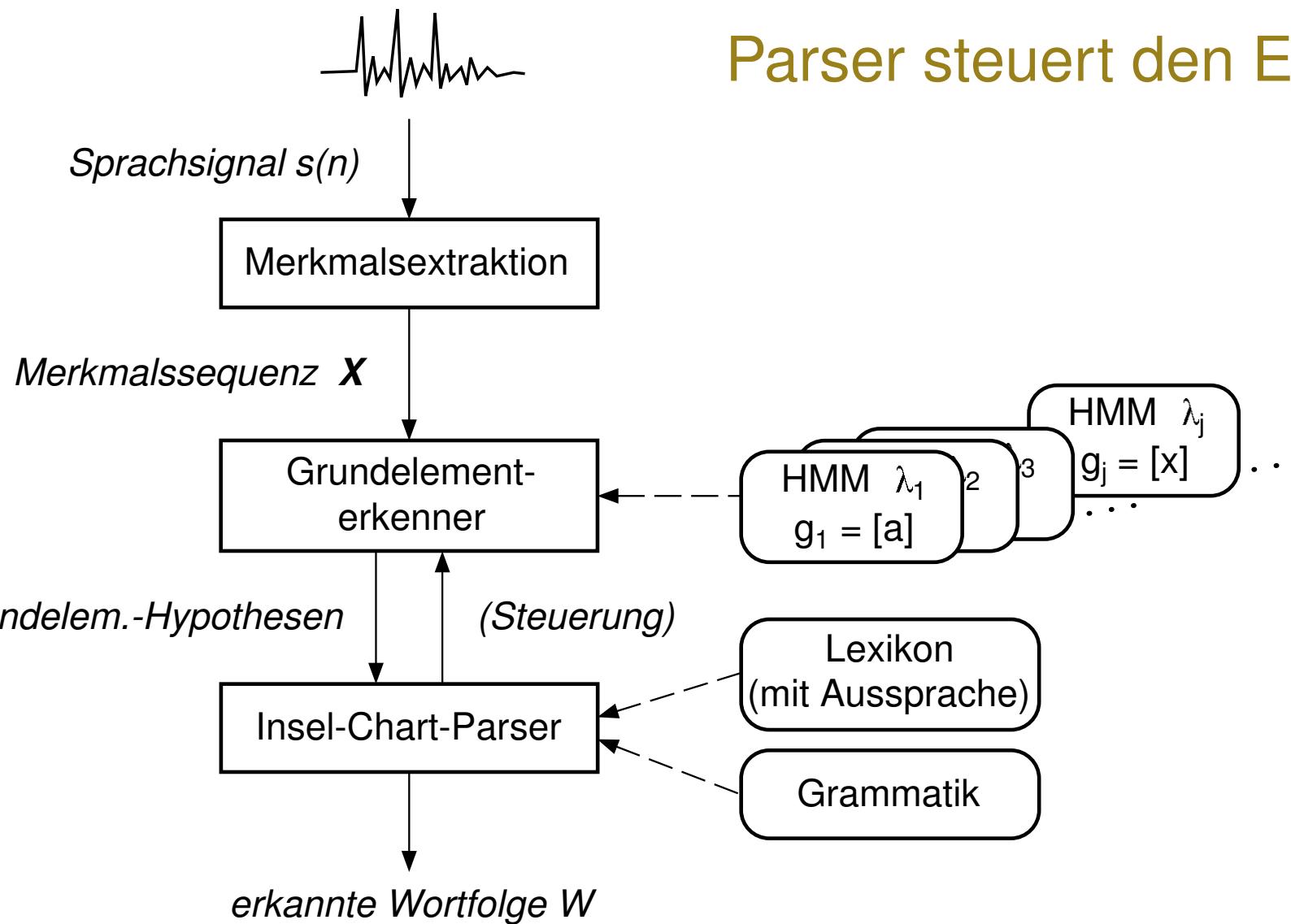
Phone Lattice



Parsing von Lauthypothesen

- Parsing von Text geht davon aus, dass die Graphemfolge und die Wortgrenzen eindeutig gegeben sind
 - Parsing in der Spracherkennung muss berücksichtigen:
 - aus dem Sprachsignal detektierte Laute sind unsicher
 - Laute können durch Koartikulation verändert oder eliminiert werden
 - weder die Anzahl noch der Ort der Laute sind eindeutig
 - die Wortgrenzen sind nicht vorhanden
 - einzelne Laute können im Phone Lattice fehlen
- Eingabe für Parser ist nicht eine Lautfolge, sondern eine unvollständige Sammlung sich teils widersprechender Lauthypothesen!

<<<

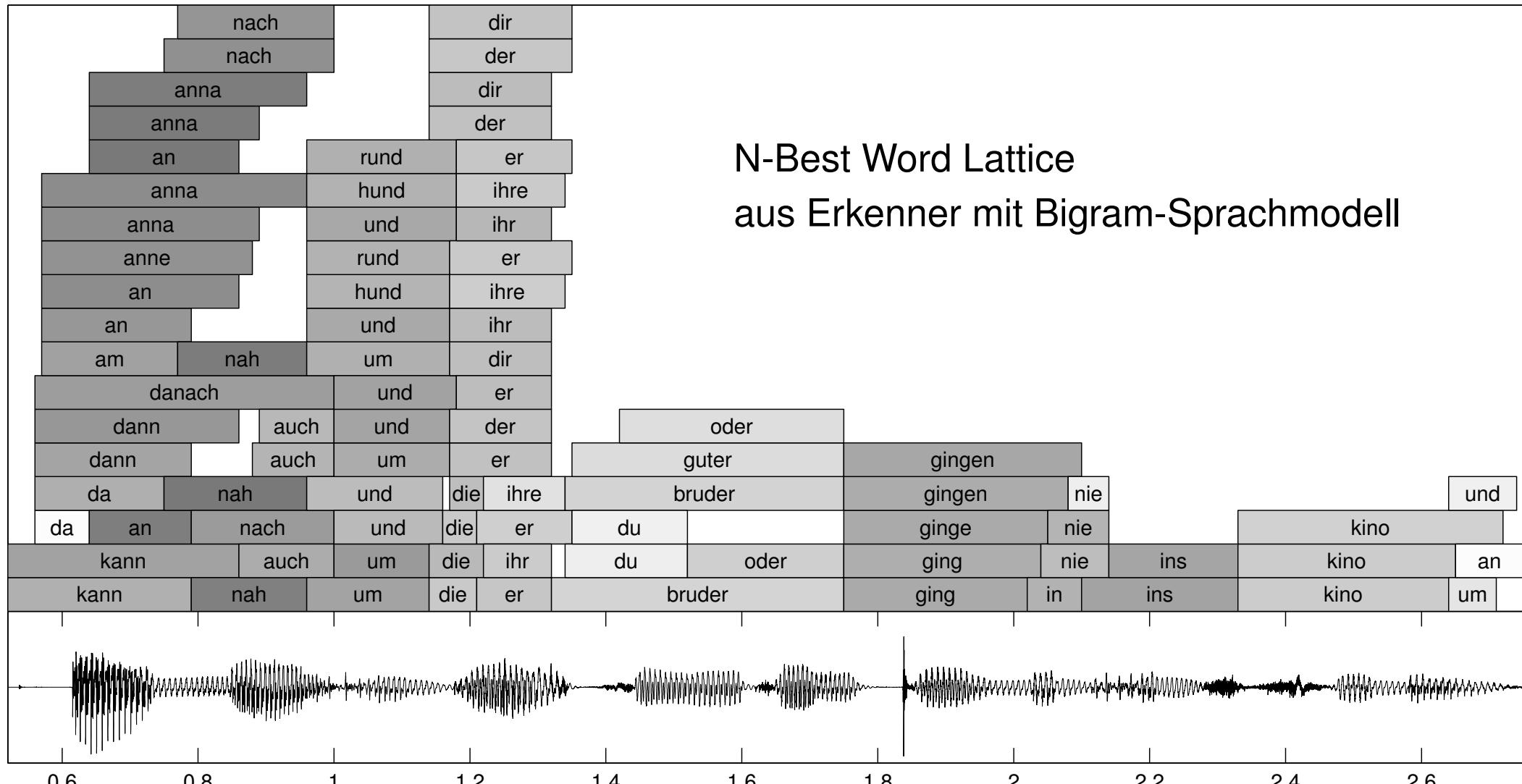


Parser steuert den Erkenner

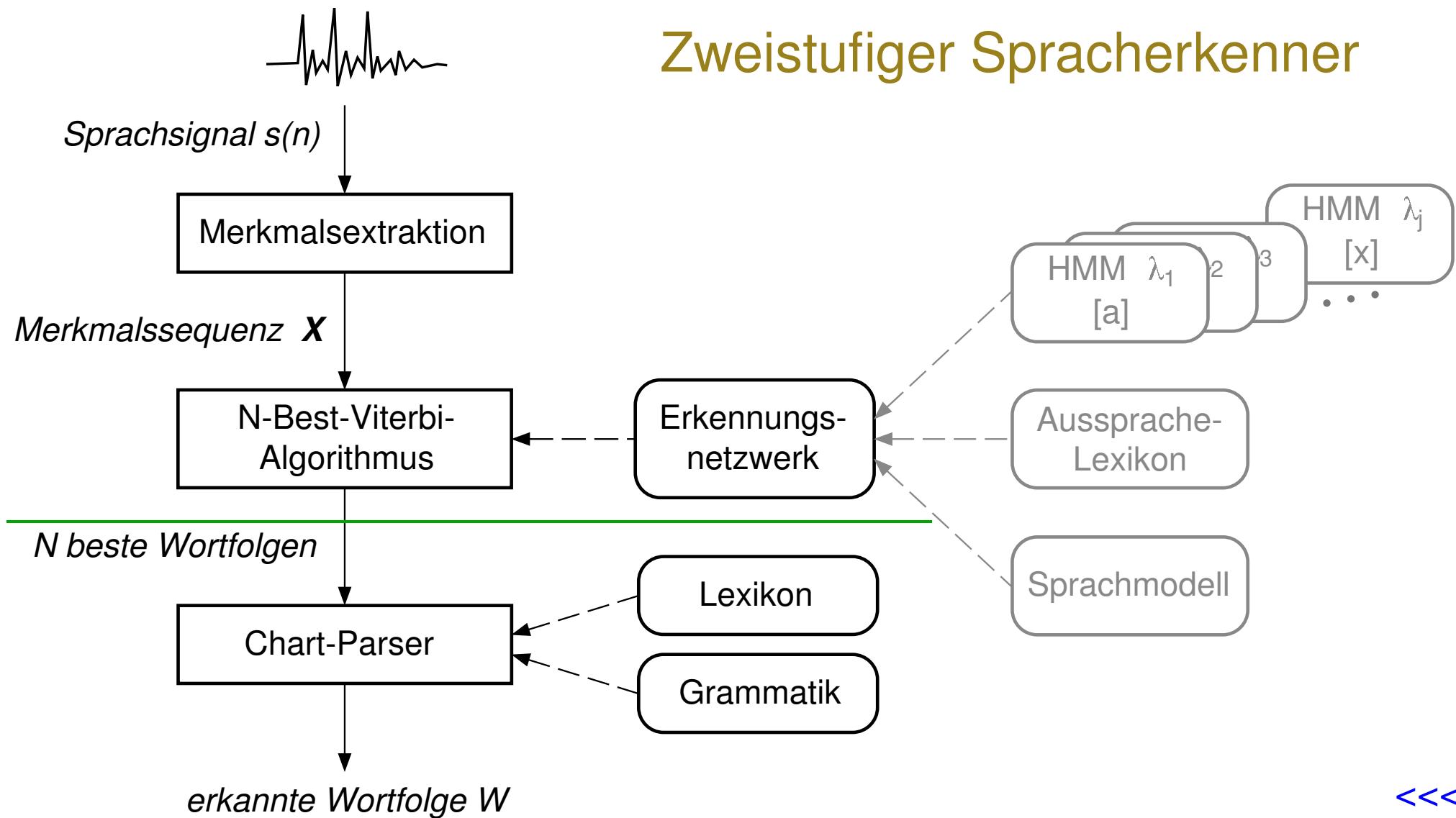
<<<

Ungelöste Probleme beim Parsen von Lauthypothesen

- Fehlende Lauthypothesen aus dem statistischen Erkenner
- Hypothesenexplosion beim Parsing
- Insel-Parsing
 - (Links-Rechts-Verarbeitung ungeeignet, weil die Laute am Anfang einer Äusserung i.a. sehr unsicher sind.)
- Intelligente Suchstrategie statt den ganzen, sehr grossen Suchraum zu durchsuchen



Zweistufiger Spracherkenner



<<<

