

Proseminar

Verarbeitung gesprochener Sprache

Timo Baumann

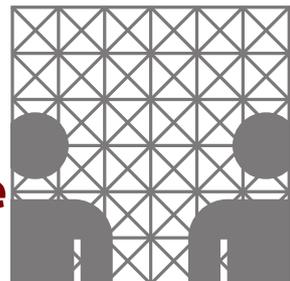
baumann@informatik.uni-hamburg.de



nats-www.informatik.uni-hamburg.de/VGS20/WebHome

Universität Hamburg, Department of

Informatics

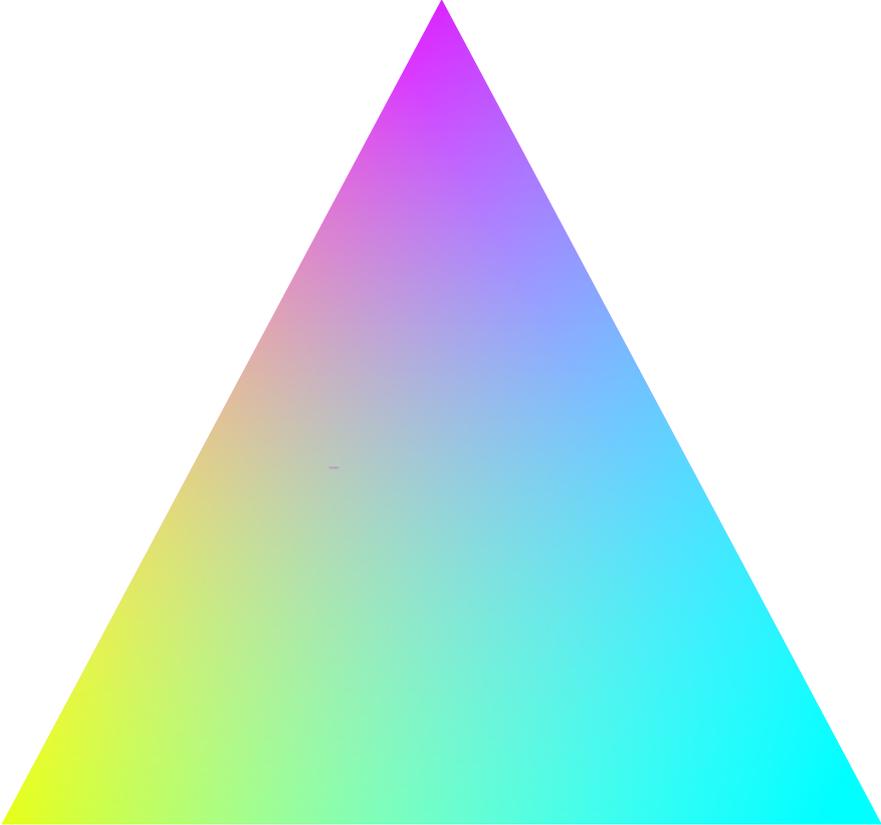


Heute

- Wissenschaftliches Schreiben
 - Was für Sorten von Proseminararbeiten gibt es?
 - Was ist die Kernfrage meiner Arbeit? ...und wo kommt sie her?
 - (Wie strukturiere ich, wie argumentiere ich?)
 - Wie zitiere ich?

Recap: Typen von Proseminararbeiten

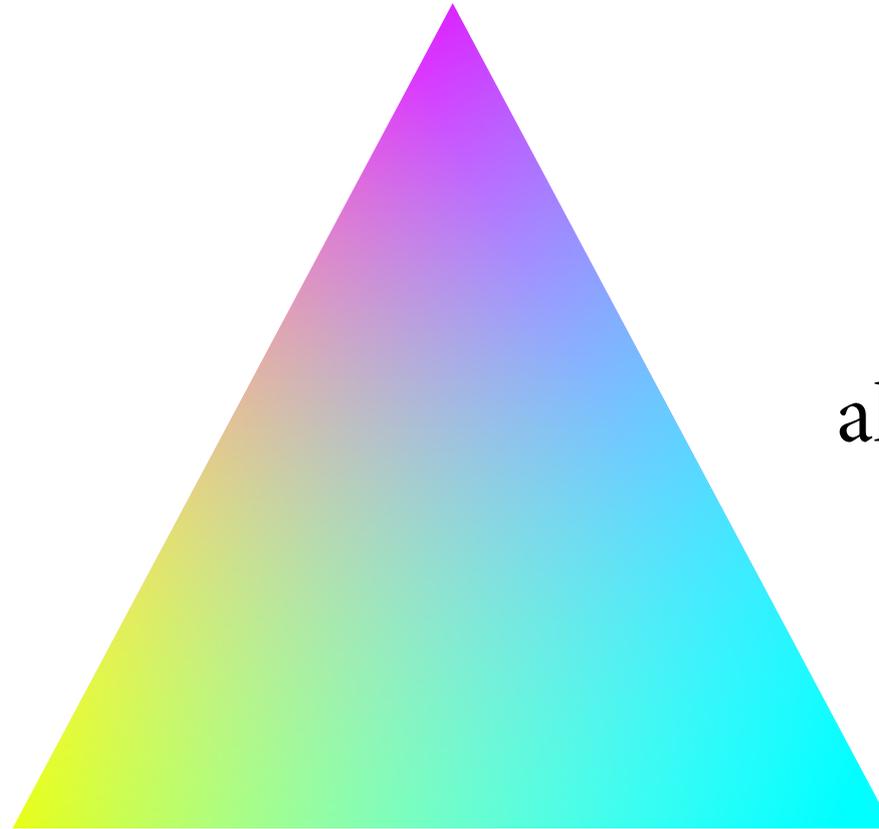
Vertiefung eines
Teilthemas



Survey über Artikel
zu einem Thema

Review eines
wiss. Papers

Vertiefung eines
Teilthemas



Survey über Artikel
zu einem Thema

Review eines
wiss. Papers

→ es entsteht
also etwas Neues!

Grundsatz für die Inhalte

- kein reines Nacherzählen/Erlebnisbericht, sondern Mehrwert durch neue Perspektive:
 - Systematisierung und Einordnung des Themas in den Proseminarkontext
 - subjektive Bewertung auf Grundlage von (erarbeiteten) objektiven Kriterien
 - für unsere Themen: substanziierte(!) Einordnung des Themas in das Gesamtgebiet Verarbeitung Gesprochener Sprache: in welchen Kontexten, Stärken/Schwächen

Literaturangaben

Literaturangaben

- bisher haben wir diese vernachlässigt
 - warum sind sie wichtig?
- Zweck:
 - Nachvollziehbarkeit
 - Abgrenzung Eigenleistung / Vorleistungen
 - „we're standing on the shoulders of giants“
 - Urheberrecht (und -pflicht...)

Literaturangaben

Literaturangaben

- was muss überhaupt belegt werden?
 - wörtliche Zitate
(immer mit Seitenzahl und Anführungsstrichen!)
 - Belege für Behauptungen
 - JEDE Behauptung wird belegt, KEINE unbelegten Behauptungen!!
 - Existenz von beispielhaft genannten Anwendungen, ...
 - auch Strukturübernahmen müssen kenntlich gemacht werden

Literaturangaben

- im Text wird an der entsprechenden Stelle mit einer Kurzangabe (=Schlüssel) auf die Quelle verwiesen
- im Literaturverzeichnis stehen die vollständigen Angaben
- *wie* Kurzangabe und Literaturverzeichnis zu formatieren sind, ist üblicherweise durch den Verleger festgelegt

Kurzangabe

- drei wesentliche Varianten für die Kurzangabe:
 - Der Kuckuck und der Esel hatten einen Streit (Fallersleben 1835).
 - Der Kuckuck und der Esel hatten einen Streit. [1]
 - Der Kuckuck und der Esel hatten einen Streit. [FAL35]
 - Der Kuckuck und der Esel hatten einen Streit.¹
- suchen Sie sich eine aus
 - Eindeutigkeit der Schlüssel: (Fallersleben 1835a, 1835b)
 - mehrere Angaben zusammenfassen: [1-5,7]
 - Seitenzahlen: [2, p. 5], (McTear 2002, p. 105).
 - „... wie schon McTear (2002, p. 105) schrieb ...“
 - „... wie schon McTear [2] schrieb ...“

¹Hoffmann von Fallersleben, August, Heinrich: *Kinderlieder*, Rororo, Reinbek, 1835, Nachdruck 1987.

Literaturverzeichnis

- am besten mithilfe einer Literaturverwaltungssoftware erzeugen lassen

! auf vollständige Angaben achten !

Exkurs: Literaturverwaltung

- am Beispiel JabRef
 - verwaltet Angaben in einer Tabelle
 - (zugrundeliegende Textdatei)
 - Primärschlüssel der Tabelle können von LaTeX aus referenziert werden:
 - `\cite{zhu1997software}` → Kurzangabe je nach gewähltem Stil
 - alle **zitierte** Literatur (und nur diese) ins Literaturverzeichnis:
 - `\printbibliography`

komplexe Anwendungen
aus separaten Modulen

Themen

Themen

- Sprachsignalrepräsentation
 - Spektrogramme, f_0 , Merkmalsvektoren
- Sprachsignalverarbeitung
 - PSOLA, VAD, Geräuschunterdrückung
- Worte und Laute
 - G2P, Morphologie, N-Gramme
- Suprasegmentalia
 - Prosodie, Emotion
- Sprachsynthese
 - Unit Selection (+PSOLA), HMM-basiert, Training
- Spracherkennung
 - Stand und Fortschritte, HMMs, Token-Passing, Training

Themen

- Sprachsignalrepräsentation
 - Spektrogramme, f_0 , Merkmalsvektoren
- Sprachsignalverarbeitung
 - PSOLA, VAD, Geräuschunterdrückung
- Worte und Laute
 - G2P, Morphologie, N-Gramme
- Suprasegmentalia
 - Prosodie, Emotion
- Sprachsynthese
 - Unit Selection (+PSOLA), HMM-basiert, Training
- Spracherkennung
 - Stand und Fortschritte, HMMs, Token-Passing, Training

Themen

- Sprachsignalrepräsentation
 - Spektrogramme, f_0 , Merkmalsvektoren
- Sprachsignalverarbeitung
 - PSOLA, VAD, Geräuschunterdrückung
- Worte und Laute
 - G2P, Morphologie, N-Gramme
- Suprasegmentalia
 - Prosodie, Emotion
- Sprachsynthese
 - Unit Selection (+PSOLA), HMM-basiert, Training
- Spracherkennung
 - Stand und Fortschritte, HMMs, Token-Passing, Training

Themen

- Sprachsignalrepräsentation
 - Spektrogramme, f_0 , Merkmalsvektoren
- Sprachsignalverarbeitung
 - PSOLA, VAD, Geräuschunterdrückung
- Worte und Laute
 - G2P, Morphologie, N-Gramme
- Suprasegmentalia
 - Prosodie, Emotion
- Sprachsynthese
 - Unit Selection (+PSOLA), HMM-basiert, Training
- Spracherkennung
 - Stand und Fortschritte, HMMs, Token-Passing, Training

Themen

- Sprachsignalrepräsentation
 - Spektrogramme, f_0 , Merkmalsvektoren
- Sprachsignalverarbeitung
 - PSOLA, VAD, Geräuschunterdrückung
- Worte und Laute
 - G2P, Morphologie, N-Gramme
- Suprasegmentalia
 - Prosodie, Emotion
- Sprachsynthese
 - Unit Selection (+PSOLA), HMM-basiert, Training
- Spracherkennung
 - Stand und Fortschritte, HMMs, Token-Passing, Training

Anwendungsbereiche

Kleingruppenarbeit:

Wählt einen der Anwendungsbereiche, die wir am Anfang des Seminars erarbeitet haben (vgl. Etherpad).

Welche Module braucht Euer Anwendungsbereich?
Was braucht es noch?

Wie werden sie verbunden? Welche
Herausforderungen/Probleme sehr ihr an den Schnittstellen?

Welche weiteren Herausforderungen gibt es mit Bezug auf die
Verknüpfung von Modulen?

typische Module zur Sprachverarbeitung

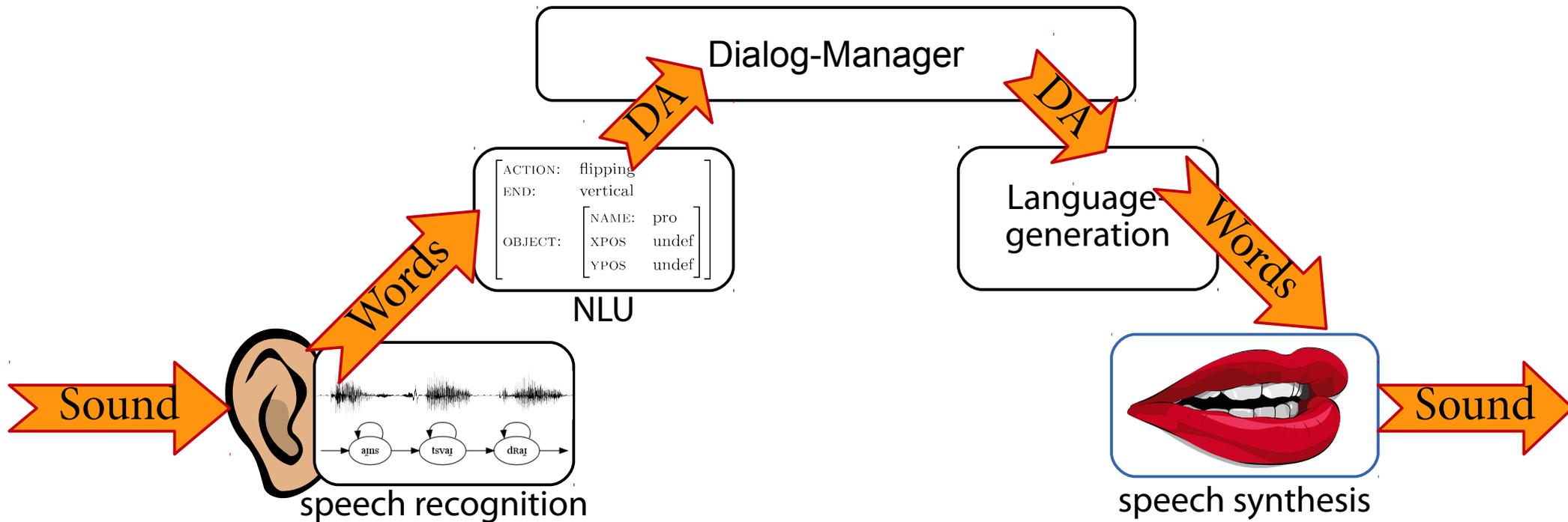
typische Module zur Sprachverarbeitung

- Spracherkennung (Audio → Wortsequenz)
- Sprachverstehen
(Wortsequenz → aussagen?logische Formel)
- Sprachgenerierung (Formel → Wörter (mit Satzzeichen))
- Sprachsynthese (Satz → Audio)

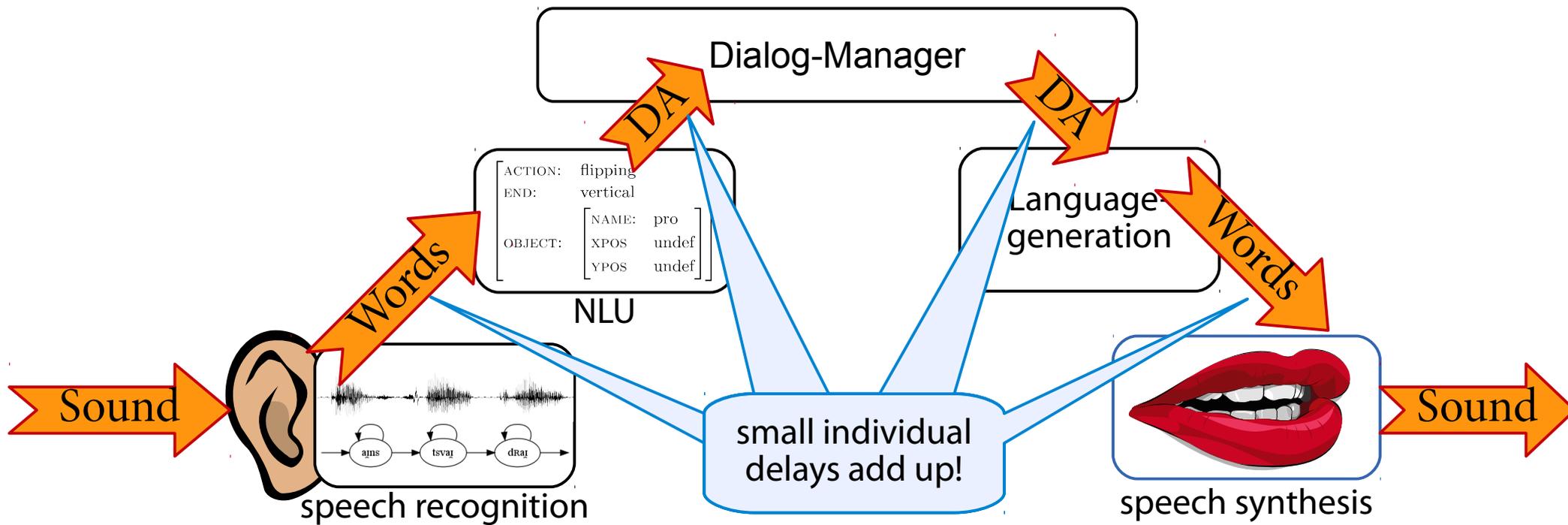
- Module
 - haben meist sehr beschränkte Schnittstellen
 - berechnen erneut, was andere Module schon wussten
 - müssen raten, obwohl Wissen vorhanden wäre

Zusammenarbeit von Modulen in interaktiven Anwendungsfällen

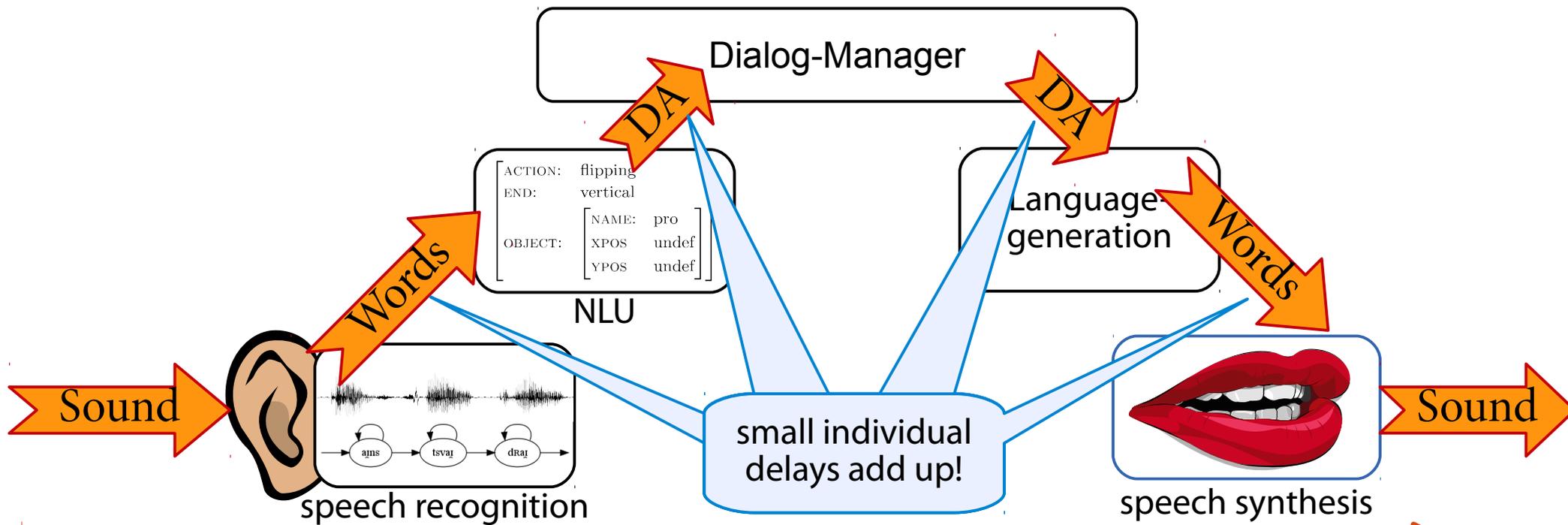
Beispiel: Dialogsystem



ein modulares Dialogsystem



ein modulares Dialogsystem



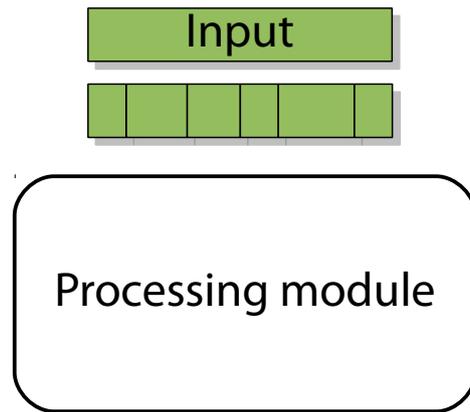
Outputerzeugung beginnt erst nach Ende des Inputs

Incremental Processing

Input

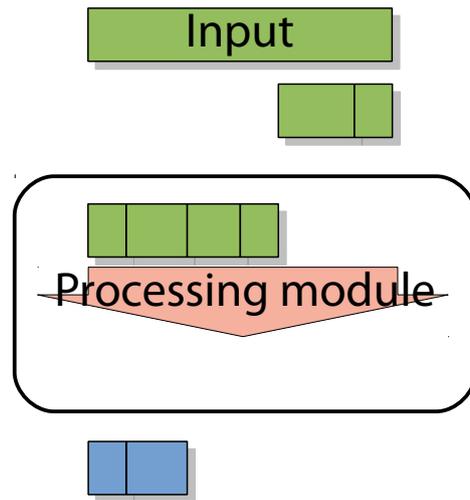
Processing module

Incremental Processing



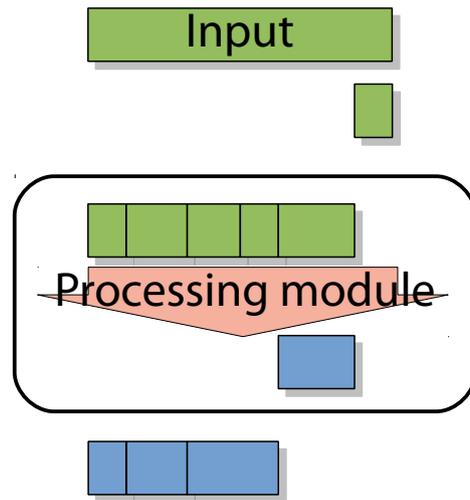
- input consists of individual units that are consumed one-by-one (e.g. speech audio, words, ideas, ...)
- input is consumed unit-by-unit, and output is generated
- input units may be aggregated to larger units

Incremental Processing



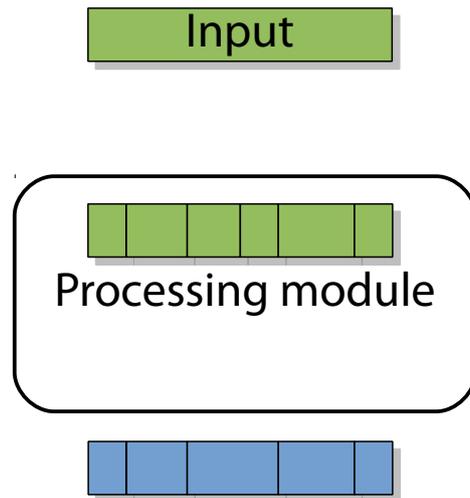
- input consists of individual units that are consumed one-by-one (e.g. speech audio, words, ideas, ...)
- input is consumed unit-by-unit, and output is generated
- input units may be aggregated to larger units

Incremental Processing



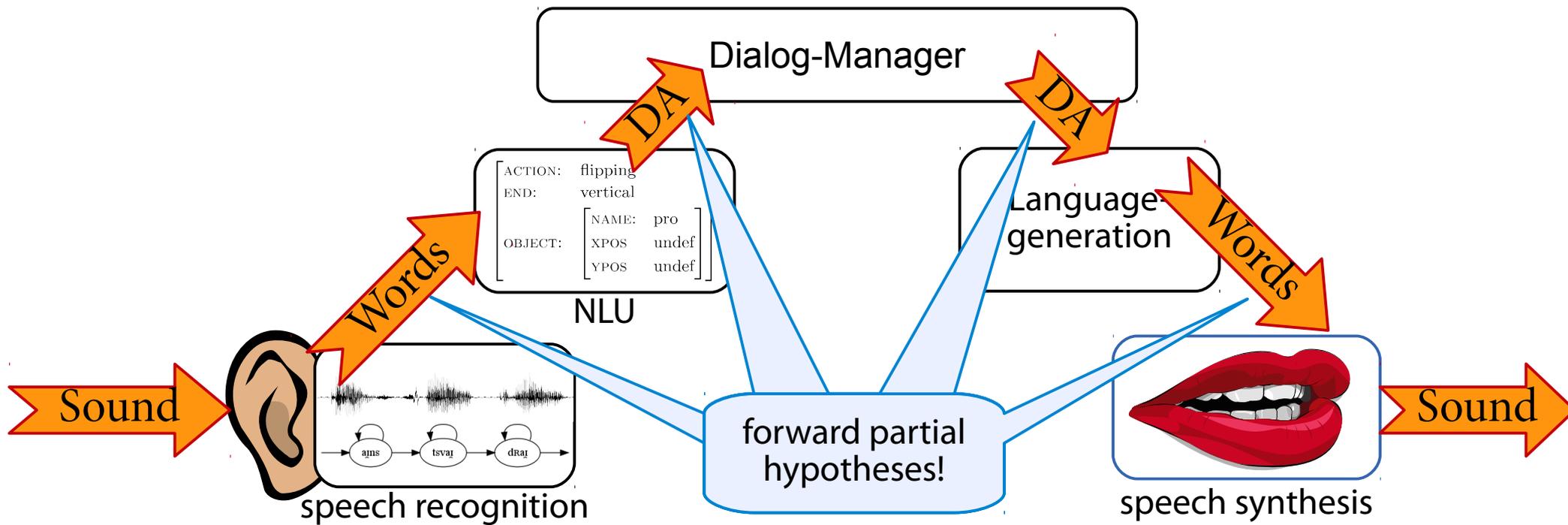
- input consists of individual units that are consumed one-by-one (e.g. speech audio, words, ideas, ...)
- input is consumed unit-by-unit, and output is generated
- input units may be aggregated to larger units

Incremental Processing

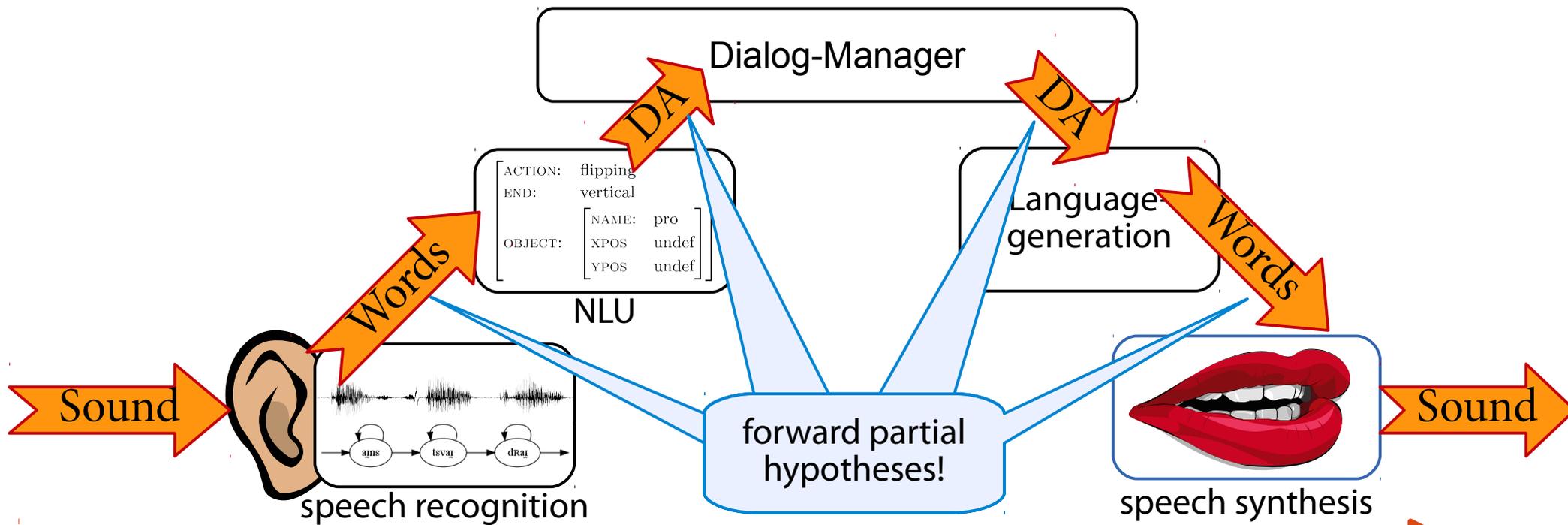


- input consists of individual units that are consumed one-by-one (e.g. speech audio, words, ideas, ...)
- input is consumed unit-by-unit, and output is generated
- input units may be aggregated to larger units

ein modulares Dialogsystem



ein modulares Dialogsystem



Output kann vor Abschluss der Inputverarbeitung beginnen

Incremental Processing: Limitations

- hypotheses are based on *what has been seen so far*
 - later input may result in changes
- example speech recognition:
 - input: [f O 6] → this sounds like “four”!
 - addition of [t i:] → together, this sounds like “fourty”!
 - what happens if [n] is next? then [EI dZ 6 z]?
- ***limited context as future input is not considered***
 - either, results will deteriorate, or:
 - allow to **revise previous hypotheses**
 - as a result, the input of following modules is revised, which will then also have to reconsider their output and so on

IU Model

- Incremental Units (IUs)
 - encapsulate minimal amounts of information at the current level of abstraction (phones, words, ideas, ...)
 - linked to other units on the *same level* to form hypotheses
 - linked to units they are based on to track dependencies
 - network of units stores information states
- Updates to the network reflect changes in understanding:
 - add units when new information becomes available
 - *revoke* units if they turned out to be wrong
 - notify about degree of commitment/certainty to a unit

Datenmodell für inkrementelle Just-in-Time-Verarbeitung

DM reasoning/decision: need to grab to be able to put → confirm

put(cross,Y)

put

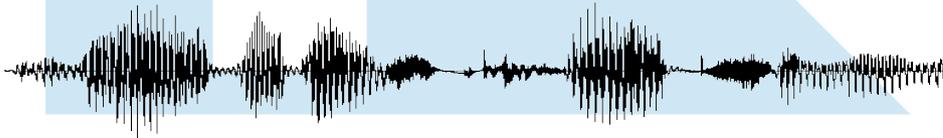
piece:cross

lege

das

kreuz

in



Analyseseseite des Systems

ack(take(X),put(X,Y))

ack

take

X=cross

okay

ich

nehm



Syntheseseite des Systems

Modularisierung

- notwendig um das Problem handhabbar zu machen
- behindert das Finden der „optimalen“ Lösung
 - insbesondere: Schnittstellenproblematik

... kannst nicht mit / kannst nicht ohne ...

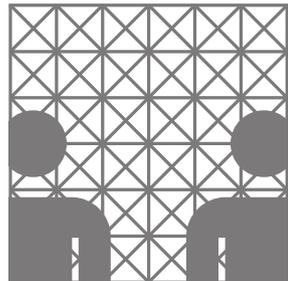
- einzige Lösung: WISSEN um die „Nachbarmodule“ der eigenen Forschung, ihrer Eigenschaften, Einschränkungen, und der üblichen Schnittstellen

Wie geht es weiter im Studium?

- Vorlesung Dialogsysteme im Wintersemester (yours truly)
- diverse Veranstaltungen im Master zu
 - language technology
 - signal and speech processing
 - statistical methods
 - deep learning
 - einschlägige Projekte
- vielfältige Möglichkeiten für Abschlussarbeiten

Vielen Dank.

baumann@informatik.uni-hamburg.de



Lernziele

- was muss belegt werden:
 - (wörtliche) Zitate, Angaben, Struktur
- was ist geeignete Literatur:
 - Fachaufsätze (Zeitschriften, Konferenzbände, ...)
 - Fachbücher / Lehrbücher
 - bedingt: Wikipedia, Webseiten
 - nicht: Zeitungen, Lexika, ...
- wie zitieren: es gibt dafür Standards, sowas sollte man nicht (mehr) von Hand machen
- Kombination von Modulen zu Systemen

Notizen

- Gruppenarbeit war nicht konkret genug: hätte ein paar Module zur Auswahl benennen sollen, darüber nachdenken lassen, wie gut die wohl zusammen passen, was bedacht werden muss. DeepLearning+f0+TokenPassing ist für eine sinnvolle Diskussion eine zu absurde Kombination.