

Natural Language Processing

Timo Baumann

Heute

- Überblick über die Struktur und Formalia der Veranstaltung
- Wie studiere ich in einer Wahlveranstaltung?
- erster Überblick über Themen der Veranstaltung: Kompositionalität, Mehrdeutigkeit, Variabilität
- in der Übung: reguläre Ausdrücke

im Moodle:

- up-to-date Informationen, Aktualisierungen, ...
- Lernziele
- Aufschlüsselung der erwarteten Arbeitsbelastung
- session breakdown:
 - hier finden Sie (kurz vor oder nach) der Vorlesung die Folien / sonstigen Medien
 - Aufgaben und Materialien für die Übungen
- relevante Termine
- später: **Ihre** Vorschläge für gute Klausurfragen

über mich

- Timo Baumann (timo.baumann@oth-regensburg.de, K-222 und online)
 - seit 2022 Professor an der OTH
 - bis 2022 Universität Hamburg (Professurvertretung Maschinelles Lernen, PostDoc Language Technology, Wimi Natürlichsprachliche Systeme, ...)
 - 2017–2018: Systems Scientist am Language Technologies Institute der Carnegie Mellon University: responsive virtuelle Agenten
 - 2007–2013: Promotion in Computerlinguistik (Potsdam, Stockholm, Bielefeld). Thema: Incremental Spoken Dialogue Processing
 - 2001–2007: Studium der Informatik und Phonetik in Hamburg, Genf und Granada
- Hauptinteressenbereiche:
 - Sprachtechnologie für interaktive (gesprochensprachliche) Anwendungen
 - multimodale und ebenenübergreifende Sprachdatenverarbeitung (z.B. Dubbing in Filmen)
 - holistische Lehre der Informatik am Beispiel Sprachtechnologie

how about you:

Wo sehen Sie sich in 3–5 Jahren?

Was müssen Sie tun um dort hin zu kommen
(und sich dort zu halten)?

Wie kann dieser Kurs dazu beitragen?
→ formulieren Sie 5 konkrete Anforderungen!

how about you:

Wo sehen Sie sich in 3–5 Jahren?

Was müssen Sie tun um dort hin zu kommen
(und sich dort zu halten)?

Wie kann dieser Kurs dazu beitragen?
→ formulieren Sie 5 konkrete Anforderungen!

how about you:

Wo sehen Sie sich in 3–5 Jahren?

Was müssen Sie tun um dort hin zu kommen
(und sich dort zu halten)?

Wie kann dieser Kurs dazu beitragen?
→ formulieren Sie 5 konkrete Anforderungen!

Dimensionen des Lernens (nach Bloom)

- unterschiedliche Aspekte von Wissen:
Fakten, Konzepte, Prozeduren, Metakognition
- unterschiedliche kognitive Prozesse:
erinnern, verstehen, anwenden, analysieren,
bewerten, kreieren

Dimensionen des Lernens (nach Bloom)

	erinnern	verstehen	anwenden	analysieren	bewerten	kreieren
Fakten						
Konzepte						
Prozeduren						
Metakognition						

Dimensionen des Lernens (nach Bloom)

	erinnern	verstehen	anwenden	analysieren	bewerten	kreieren
Fakten						
Konzepte						
Prozeduren						
Metakognition						

- keine inhärente qualitative Ordnung:
analysieren *wie* man Fakten erinnert ist nicht
immer besser als Fakten zu erinnern

Erhoffte Lehr- /Lernergebnisse (I)

- die Studierenden haben einen Überblick über die Teilbereiche der Sprache und ihre jeweilige Verarbeitung
- die Studierenden können zentrale Probleme bei der Sprachverarbeitung analysieren und mögliche Lösungen gegeneinander abwägen

Erhoffte Lehr- /Lernergebnisse (II)

- die Studierenden können ihr Studienverhalten reflektieren und steuern
- die Studierenden wissen, wie sie sich in weitere ähnliche Anwendungsfelder für Informatik und KI einarbeiten können, und können ihr Verständnis und ihre weiteren Fragen mit Kollegen diskutieren

Das ultimative Studienziel

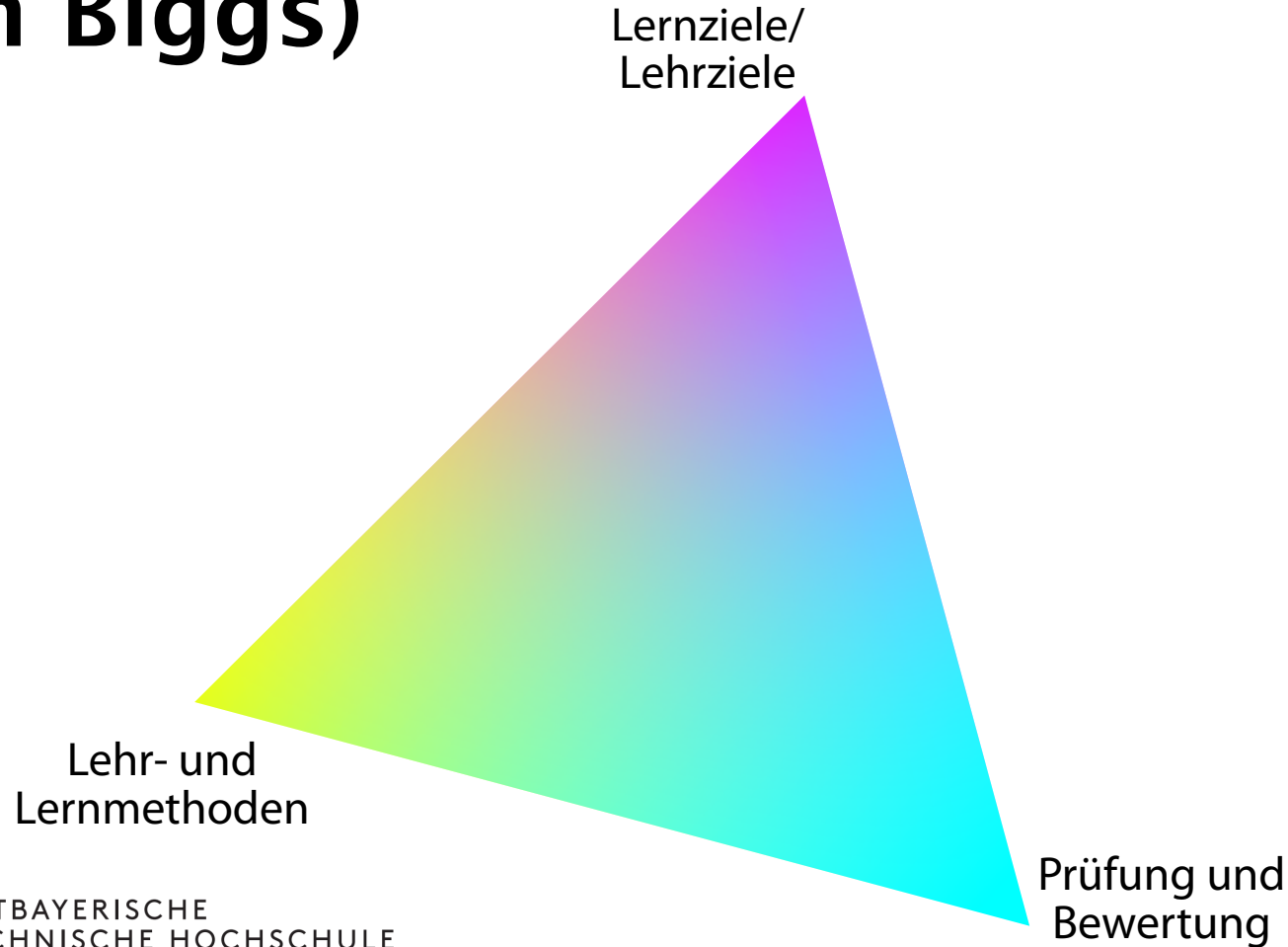
In Ihrem Studium haben sie zwei (?) Wahlmodule,
in denen Sie sich in Teilgebieten spezialisieren.

Mehr benötigen Sie nicht.

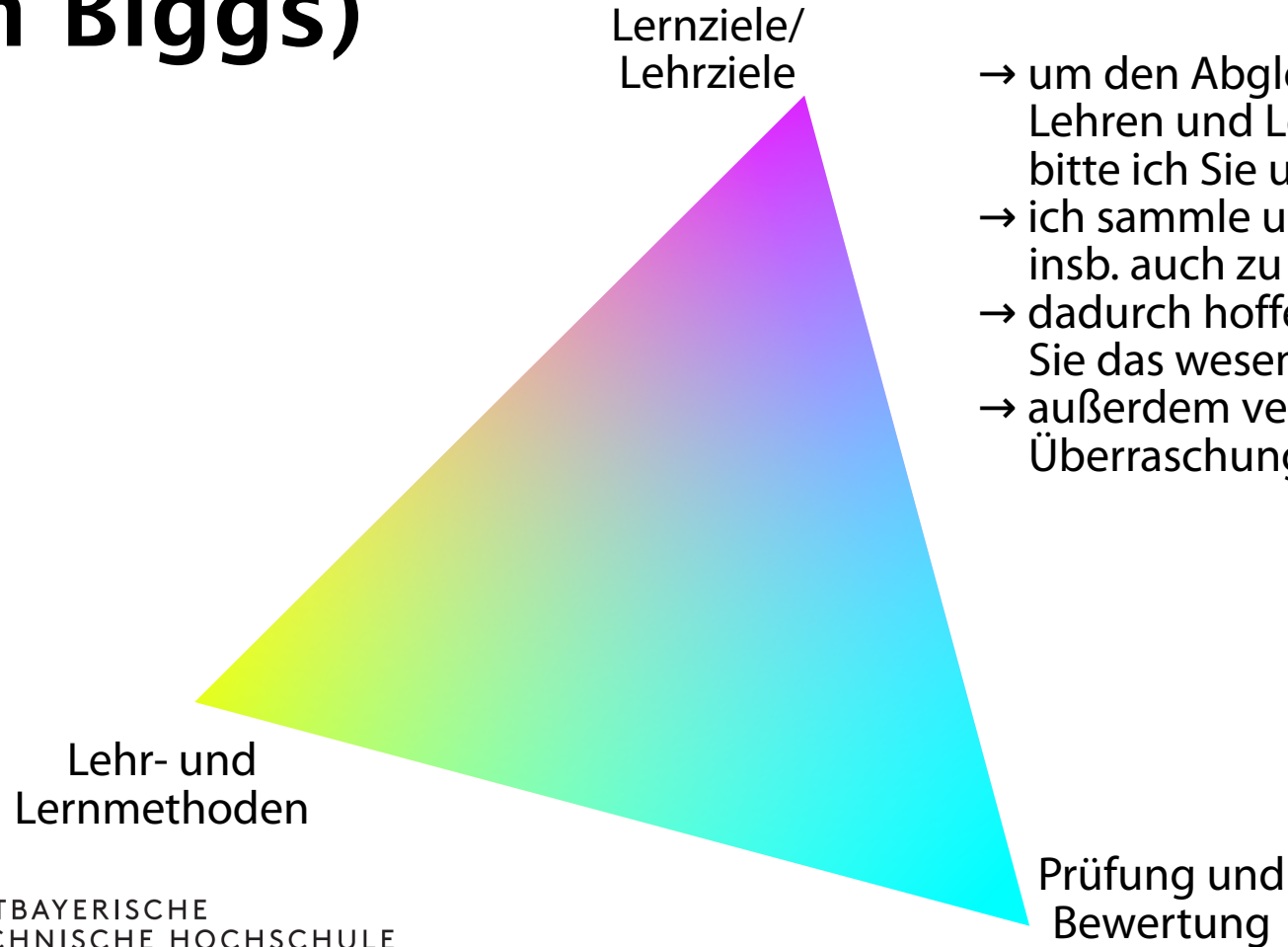
Sie sind dann in der Lage, sich selbständig weiter zu
spezialisieren.

(Studieren Sie so, dass Sie dieses Ziel erreichen.)

Constructive Alignment (nach Biggs)

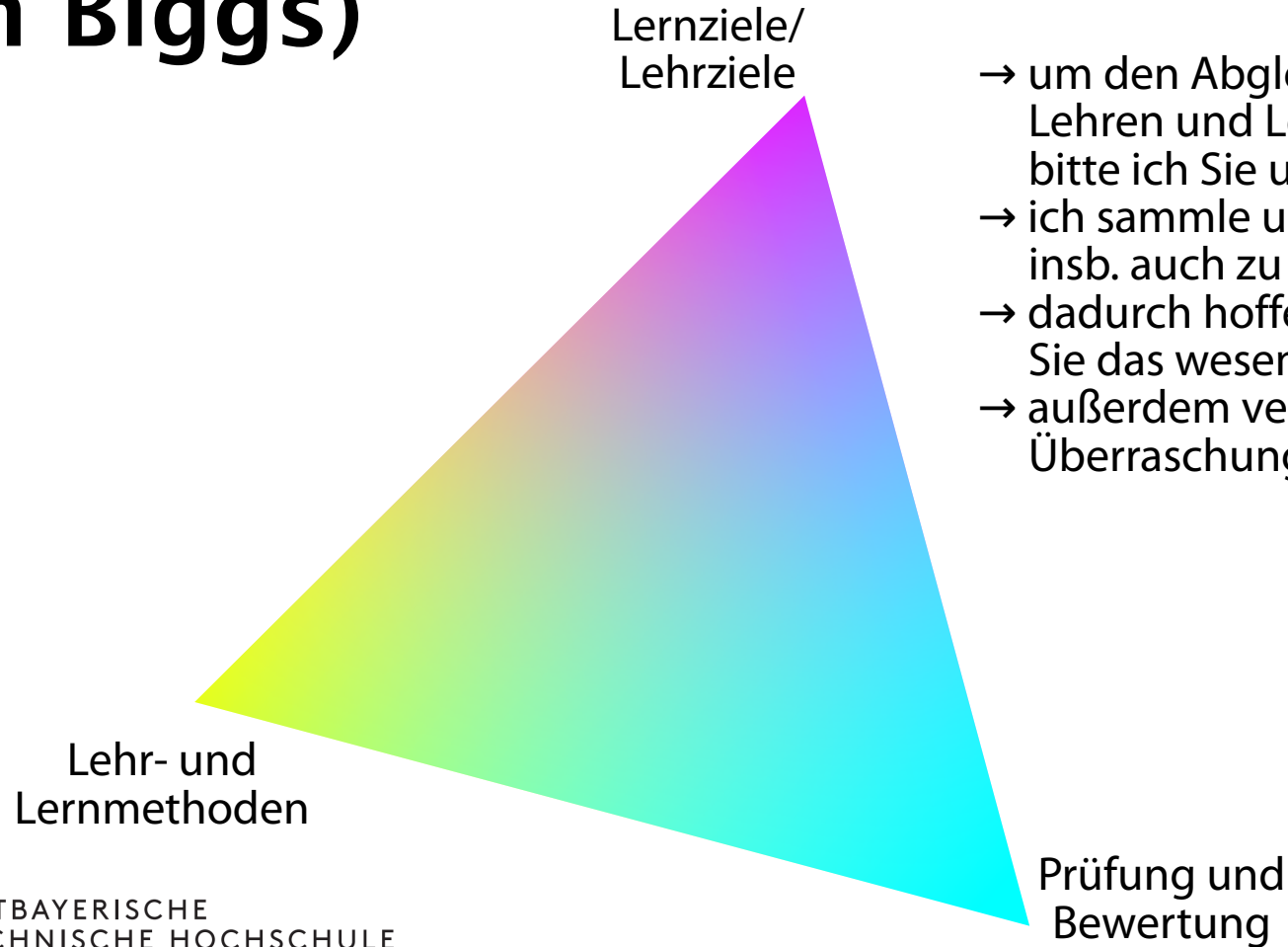


Constructive Alignment (nach Biggs)



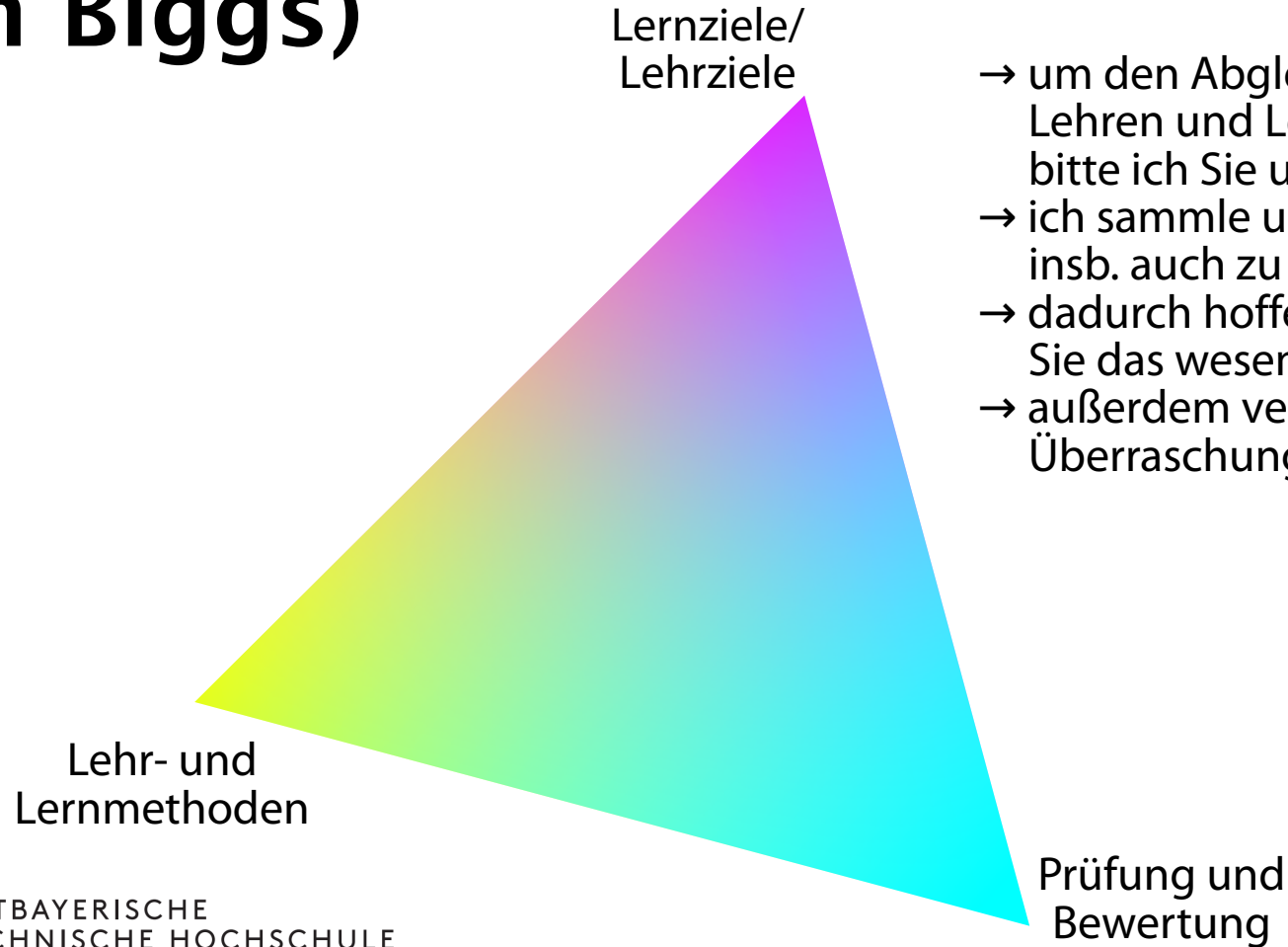
- um den Abgleich zwischen Lehren und Lernen zu sichern, bitte ich Sie um Prüfungsfragen
- ich sammle und kommentiere sie insb. auch zu unserem Alignment
- dadurch hoffe ich so zu lehren, dass Sie das wesentliche lernen
- außerdem vermeiden wir so Überraschungen in der Klausur

Constructive Alignment (nach Biggs)



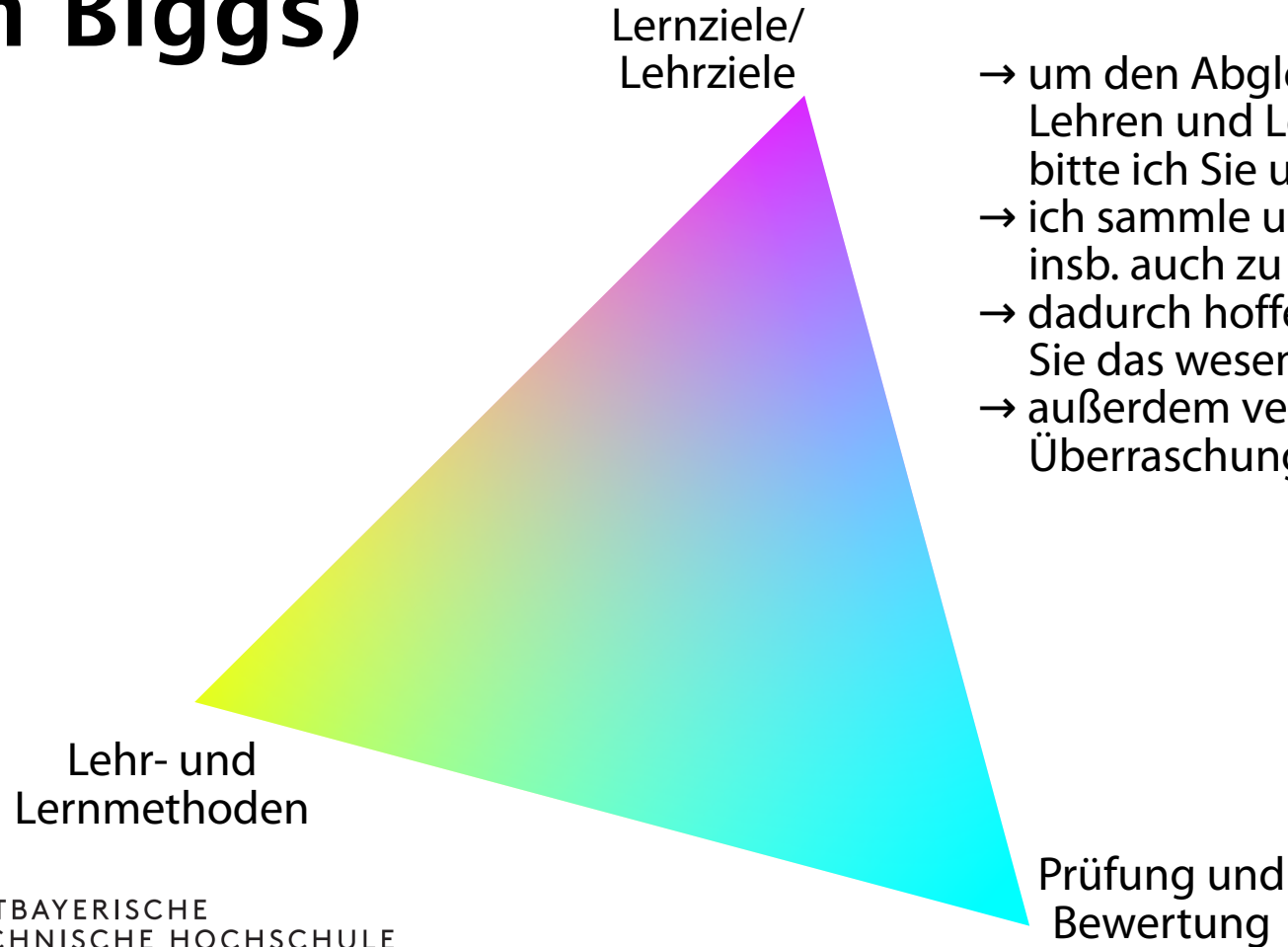
- um den Abgleich zwischen Lehren und Lernen zu sichern, bitte ich Sie um Prüfungsfragen
- ich sammle und kommentiere sie insb. auch zu unserem Alignment
- dadurch hoffe ich so zu lehren, dass Sie das wesentliche lernen
- außerdem vermeiden wir so Überraschungen in der Klausur

Constructive Alignment (nach Biggs)



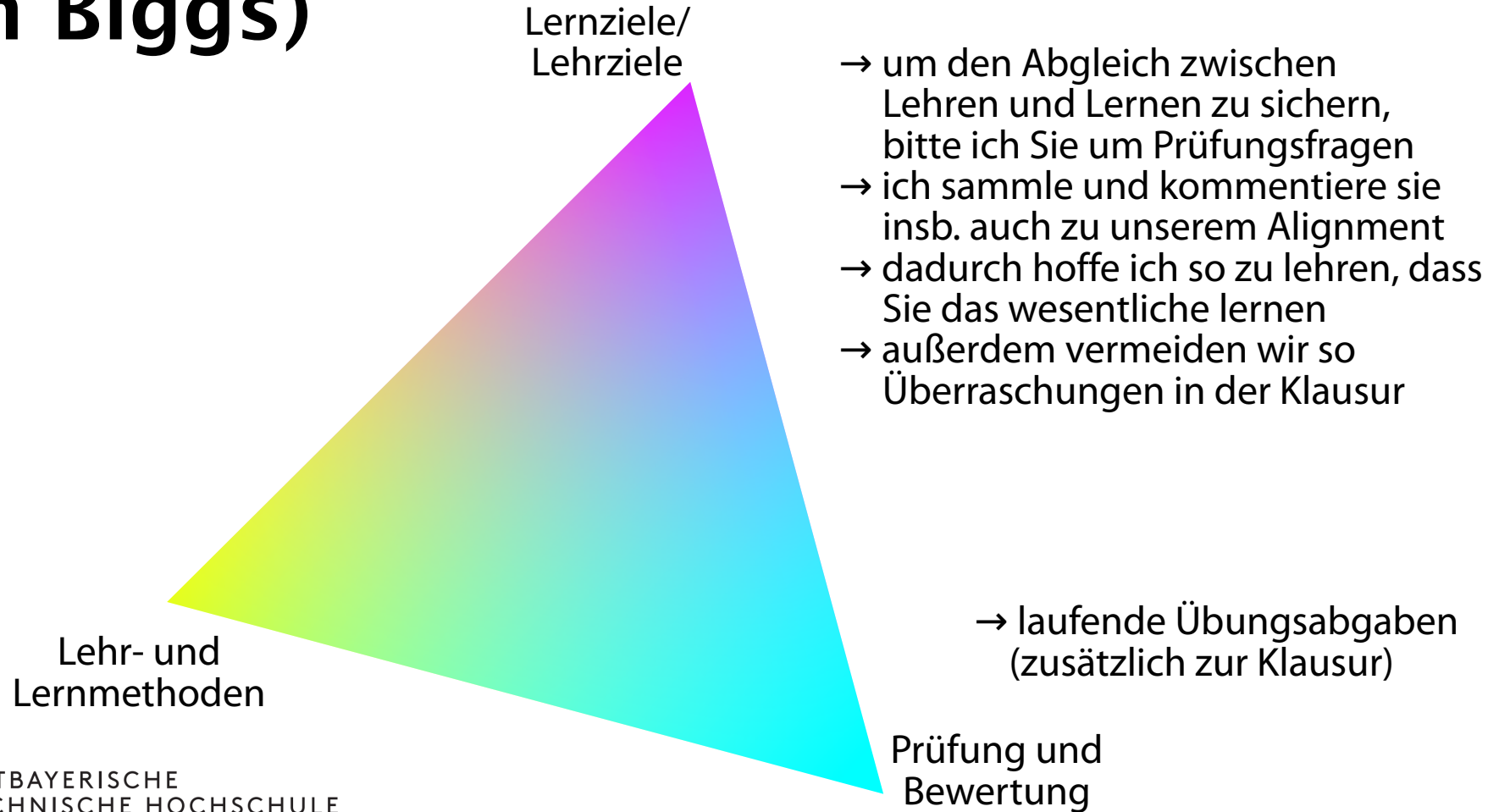
- um den Abgleich zwischen Lehren und Lernen zu sichern, bitte ich Sie um Prüfungsfragen
- ich sammle und kommentiere sie insb. auch zu unserem Alignment
- dadurch hoffe ich so zu lehren, dass Sie das wesentliche lernen
- außerdem vermeiden wir so Überraschungen in der Klausur

Constructive Alignment (nach Biggs)



- um den Abgleich zwischen Lehren und Lernen zu sichern, bitte ich Sie um Prüfungsfragen
- ich sammle und kommentiere sie insb. auch zu unserem Alignment
- dadurch hoffe ich so zu lehren, dass Sie das wesentliche lernen
- außerdem vermeiden wir so Überraschungen in der Klausur

Constructive Alignment (nach Biggs)



Konkrete Umsetzung

- Wissensdomänen:
 - Aspekte nat. Sprache
 - formale Strukturen zu deren Behandlung
 - Algorithmen für NLP
 - trainingsbasierte Verfahren
 - Betrieb existierender Software für NLP-Aufgaben
- Fertigkeiten (→ Übungen)
 - bestehende Software inbetriebnehmen
 - händisch Algorithmen nachvollziehen
 - NLP-Algorithmen implementieren
 - datenbasierte Experimente für NLP durchführen
 - Verknüpfungen per Skript
- pro Übung nur eine Auswahl!

Formalia

- Ihre Aktivität in Vorlesung und Übung ist erforderlich, um Ihre Studienziele zu erreichen
- Bewertung:
 - 50% Klausur (im Prüfungsblock)
 - 50% Übungsabgaben (=kumulative Studienarbeit)

Brainstorming Anwendungsbereiche

Wo steckt überall Sprachtechnologie /
Natural Language Processing?

Kleingruppenarbeit (3–4 Personen)

Strukturieren und ergänzen
Sie die gesammelten Beispiele.

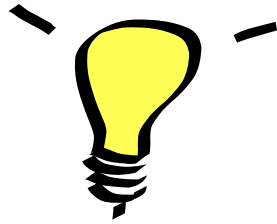
Anhand welcher Kriterien
können Sie strukturieren?

Möglichkeiten zur Strukturierung

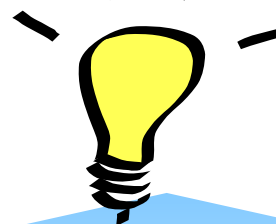
Möglichkeiten zur Strukturierung

z.B. entlang der Kommunikationskette

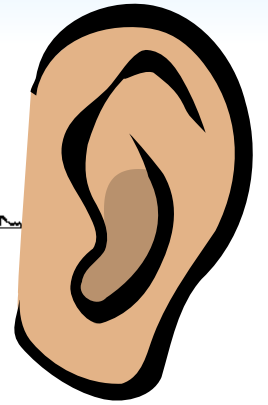
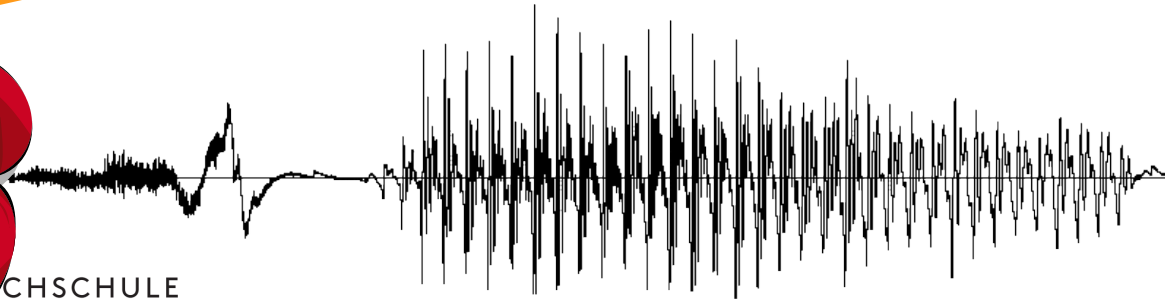
Sprachliche Kommunikationskette



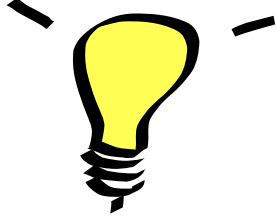
find message that describes idea
determine structure to convey meaning
sequentialize structure to word stream
represent words through sounds



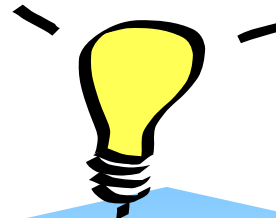
recover idea described by message
determine meaning of structure
recover structure of sequence
recombine sounds to words



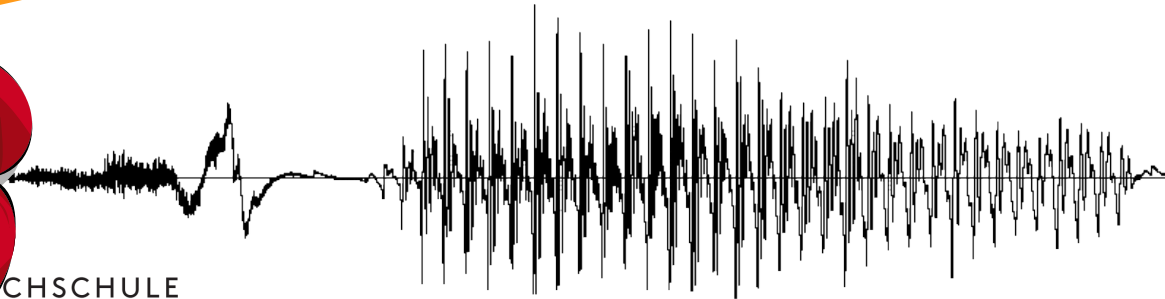
Sprachliche Kommunikationskette



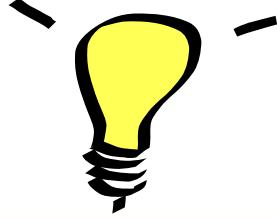
find message that describes idea
determine structure to convey meaning
sequentialize structure to word stream
represent words through sounds



recover idea described by message
determine meaning of structure
recover structure of sequence
recombine sounds to words



Sprachliche Kommunikationskette

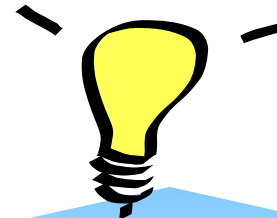


find message that describes idea

determine structure to convey meaning

sequentialize structure to word stream

represent words through sounds

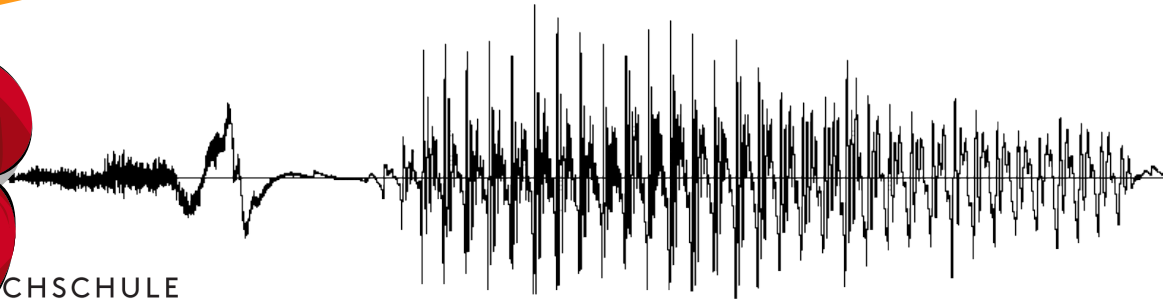


recover idea described by message

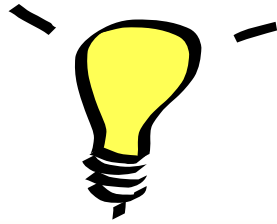
determine meaning of structure

recover structure of sequence

recombine sounds to words



Sprachliche Kommunikationskette

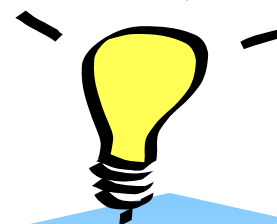


find message that describes idea

determine structure to convey meaning

sequentialize structure to word stream

represent words through sounds

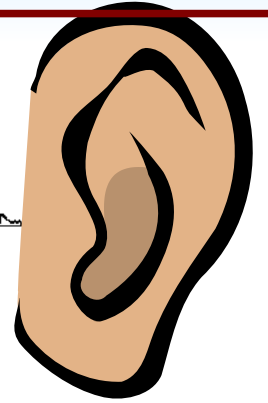
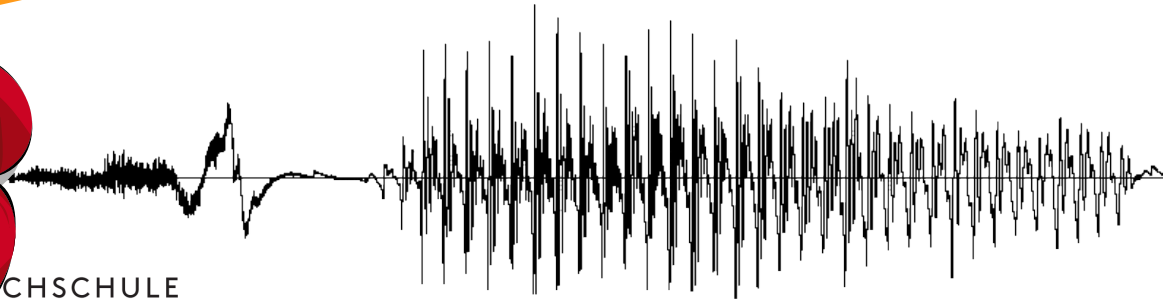


recover idea described by message

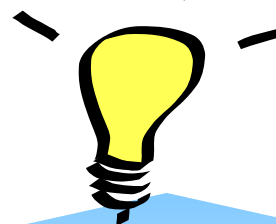
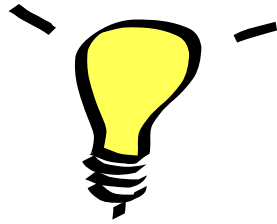
determine meaning of structure

recover structure of sequence

recombine sounds to words



Sprachliche Kommunikationskette



find message that describes idea

Pragmatik

recover idea described by message

determine structure to convey meaning

**Semantik/
Lexikologie**

determine meaning of structure

sequentialize structure to word stream

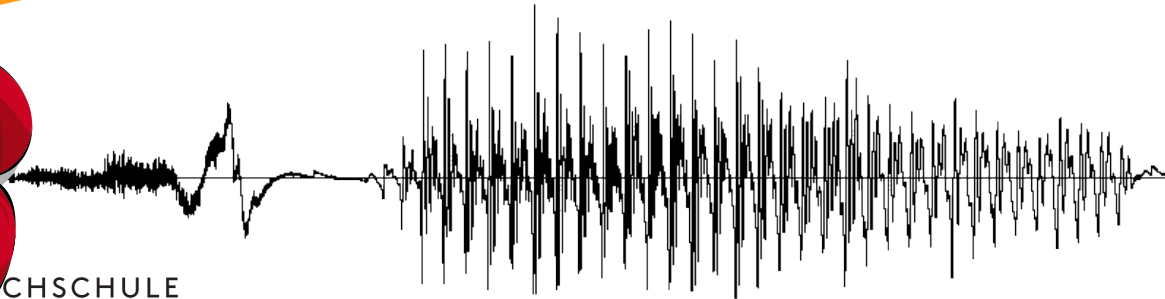
**Syntax/
Morphologie**

recover structure of sequence

represent words through sounds

**Phonologie/
Phonetik**

recombine sounds to words



Prinzip der Kompositionalität

- Laute (oder Buchstaben) bilden Wörter
 - Wortbildung folgt gewissen Regeln
- Wörter bilden Wortfolgen
- Wortfolgen haben eine innere Struktur (Syntax)
- Bedeutung erschließt sich anhand der Struktur

Mittel der Kompositionalität

- unterscheiden sich über sprachliche Ebenen
 - Wortbildung ist (überwiegend) regulär
 - Syntax nutzt (überwiegend kontextfreie) Grammatik
 - Semantik nutzt logische Zusammenhänge zwischen verknüpften Entitäten (=Graphen)
- diese kompositionale Herangehensweise verfolgen wir im ersten Drittel des Semesters

Mehrdeutigkeit

- "Man hört da viele Vorträge, aber auch viele interessante."
- "Der Schwiegersohn von Versandhaus-Milliardär Michael Otto und Ex-Profiboxer Ismail Özen ist am Donnerstag vor dem Amtsgericht Altona vom Vorwurf der Körperverletzung freigesprochen worden."
- Comicheldin
- Museumshopping

Mehrdeutigkeit (II)

“One morning I shot an elephant in my pajamas.”
--“How he got in my pajamas, I don't know.”

- mehrdeutig: ja. gleich wahrscheinlich? nein.
 - probabilistische Modelle im 2. Semesterdrittel

ONE METHOD TO RULE

THEM ALL

**Drittes Drittel: Neuronale Netze als (semi)uniformes
Werkzeug zur Modellierung von Sprachdaten**

aber vorher

- ganz basic: wir fangen mit regulären Ausdrücken und der Kommandozeile an

Vielen Dank, Ihre Fragen?

`timo.baumann@oth-regensburg.de`

weiterführende Literatur

- ich erwarte nicht von Ihnen, dass Sie den Teil zur Lerntaxonomie und Constructive Alignment nachbereiten.
- reguläre Ausdrücke und endliche Automaten: Ihre Lieblingseinführung in die Formalen Grundlagen der theoretischen Informatik

Lehrziele

- die Studierenden verstehen die Formalia und die Strukturierung des Kurses und die Bewertung ihrer Leistung
- die Studierenden haben ihre eigenen Ziele mit Bezug zum Wahlmodul und das akademische Studium als solche durchdacht
- die Studierenden kennen die unterschiedlichen Arten von Wissen (Fakten ... Metakognition) und Lerntiefen (erinnern ... kreieren)
- die Studierenden vermeiden es, sich auf oberflächliches Lernen zurückzuziehen
- die Studierenden verstehen, dass ihr Feedback unverzichtbar ist für den Abgleich zwischen den Lehrzielen des Dozenten und ihrem eigenen Lernen