

Proseminar Sprachgebrauch M03,3: Einführung in die Computerlinguistik (CL)

Leitung:

Paul Rössler (Univ. Wien),

Dieter Merkl (TU Wien),

Gudrun Kellner (Univ.+TU Wien)

WiSe 2008

Proseminar Sprachgebrauch M03,3: Einführung in die CL

Einheit: Zeit:

- 1) 17. Okt. 9.30-11.00
- 2) 24. Okt. 9.30-11.00
- 3) 31. Okt. 9.30-11.00
- 4) 7. Nov. 9.30-11.00

Ort:

- (Uni Wien, Ü4)
(Uni Wien, Ü4)
(TU Wien, EI 5)
(TU Wien, HS 15)

5+6) 1. Block 12. Dez. 9.30-12.45

(Uni Wien, Ü4)

7+8) 2. Block 9. Jän. '09 9.30-12.45

(TU Wien, EI 4)

9+10) 3. Block 16. Jän. '09 9.30-12.45

(Uni Wien, Ü4)

[evtl. 11+12) 23. Jän. '09 4.Block 9.30-12.45

(TU Wien, EI 5)]

Im 1. Teil des PS (Okt. + Anfang Nov. 08) erfolgt eine allg. Einführung in die Computerlinguistik. Im 2. Teil, nach 4 Einheiten (Dez. 08 + Jän. 09), beginnt die Referatsphase. Die TeilnehmerInnen sollen eigene Themen auswählen und diese in Referaten bzw. Kleinprojekten realisieren und präsentieren.

1. – 4. Einheit: Einführung in die Computerlinguistik (17., 24., 31. Okt., 7. Nov. 2008)

4. Einheit (7. Nov.): Vorschläge und Vergabe der Referatsthemen und Referatstermine

5. – 12. Einheit (12. Dez. 2008; 9., 16., 23. Jän. 2009): Referate

Abmeldung vom PS bis 31.10. möglich! Ab 1.11. wird die Leistung des/r TeilnehmersIn auf jeden Fall benotet.

Literaturtipps zum Einstieg

Carstensen, Kai-Uwe (Hg.): Computerlinguistik und Sprachtechnologie. Eine Einführung.- München 2001 (2. Aufl. 2004)

Schmitz, Ulrich: Computerlinguistik.- Opladen 1992

Bátori, István (Hg.): Computational linguistics. An international handbook.- Berlin 1989

Jahrbuch für Computerphilologie (bisher 8 Bde.). 1999-2007

Ufert, Detlef: Einführung in Datenverarbeitung für Linguisten und Literaturwissenschaftler.- Aachen 1994

Zum erfolgreichen Abschluss des Proseminars notwendig:

- **Handout** (1 Woche vor Referatstermin mit LV-Leitern besprechen)
- **Referat (20 Min. + Diskussion 10 Min.) – empfohlen: interdisziplinäre Teams (2-3 TeilnehmerInnen)**
- **pro Referatsthema eine Theorie + bestehende oder vorstellbare Anwendungsszenarien präsentieren**
- **Proseminararbeit (25-30 Seiten pro Team – Abgabefrist: 1. April 2009)**

Was treiben ComputerlinguistInnen?

ComputerlinguistInnen untersuchen
Aufbau, Bedeutung und Verwendung
menschlicher Sprache,

um Computer zu intelligenten Werkzeugen
menschlicher Kommunikation zu machen.

Computerlinguistik vs. Linguistik – unterschiedliche Herangehensweisen

Die klassische Linguistik des 20. Jhs. dachte eher produktorientiert:

Strukturalismus, Systemlinguistik:

Aus welchen Einheiten besteht Sprache und wie ist das Verhältnis dieser Einheiten zueinander?

CL denkt prozedural bzw. prozessual:

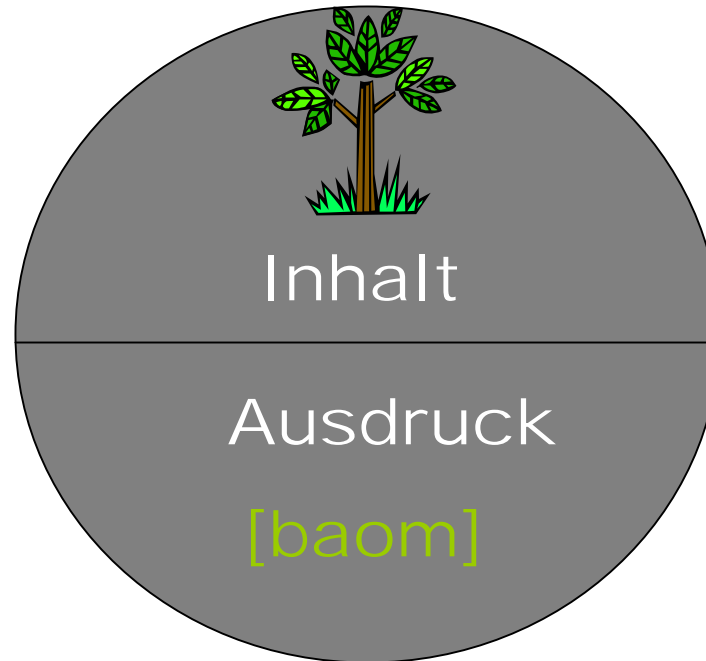
Was muss man wissen, um Sprache sprechen und verstehen zu können? Wie ist der menschl. Geist (Gehirn) organisiert, damit dieses Wissen in der Kommunikation angewendet werden kann?

Wissenschaftsgeschichte: Strukturalismus

- **Ferdinand de Saussure** (1857-1913):
- *Cours de linguistique générale* 1907-1911 Vorlesung in Genf
- ... 1916 Mitschrift v. Charles Bally u. Albert Sechehaye
- ... 1931 dt: Grundfragen der allgemeinen Sprachwissenschaft
- diachron – synchron
- präskriptiv – deskriptiv
- **language** ... allg. menschl. Fähigkeit zur Sprache
- **langue** ... sprachl. Einzelsystem: gemeins. Zeichenvorrat, den alle e. Gemeinschaft kennen
- **parole** ... individuell: persönl. Gebrauch d. Zeichenvorrats

Wissenschaftsgeschichte: Strukturalismus - Saussure: **Sprachliches Zeichen**

Inhalt:
Bezeichnetes
Signifikat
signifié
concept
Begriff



Ausdruck:
Bezeichnendes
Signifikant
signifiant
image acoustique
Lautkörper

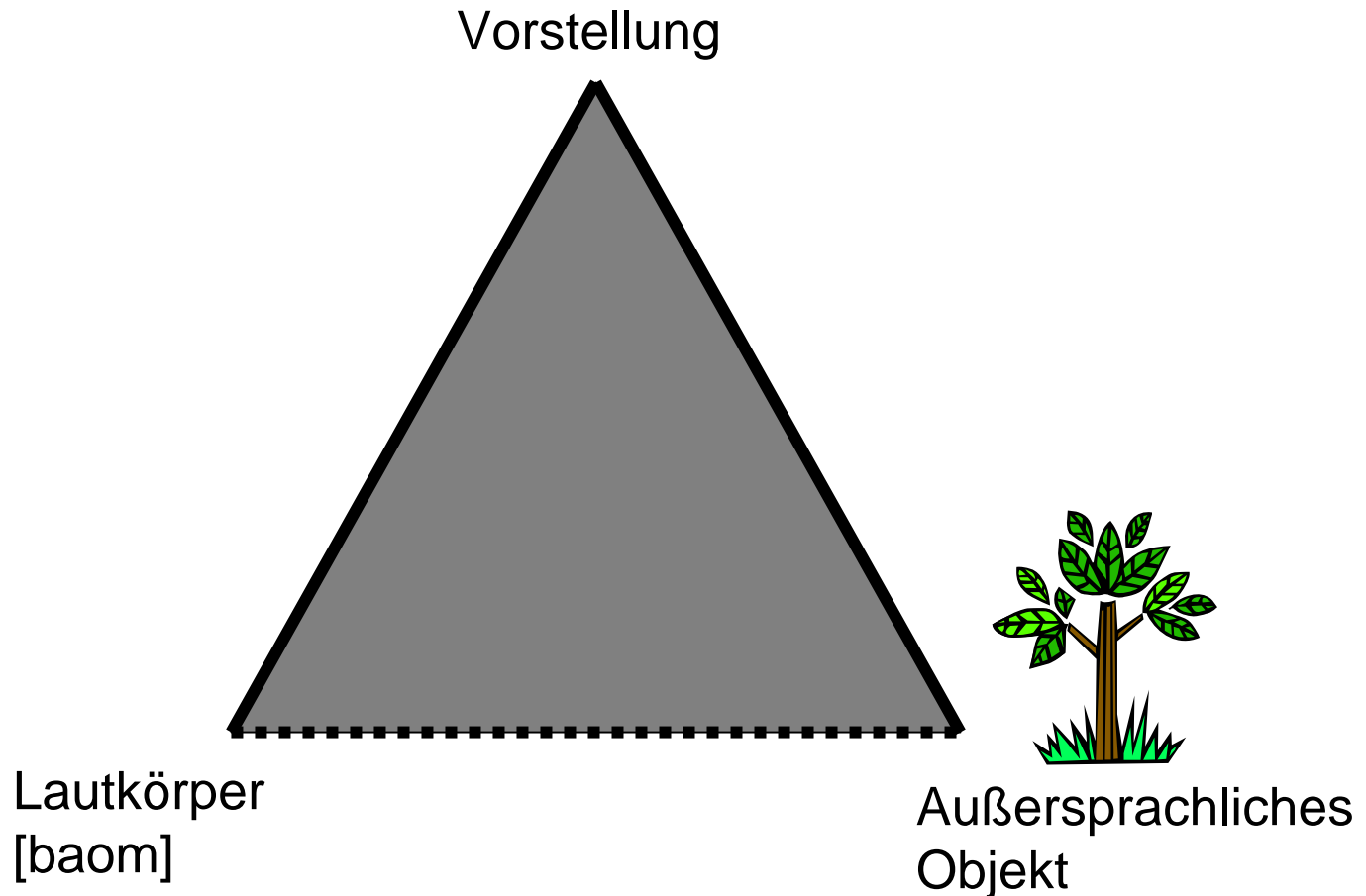
Wissenschaftsgeschichte: Strukturalistische Methodik

- **Segmentierung:** zerteilen in kleinste sprachl. Einheiten
- **Klassifizierung:** Bewertung/Zuordnung der segmentierten Einheiten zu Klassen
- **Syntagma:** Beziehung eines sprachl. Zeichens zum benachbarten sprachl. Zeichen (horizontal)
- **Paradigma:** Beziehung austauschbarer sprachl. Zeichen zueinander (vertikal)

Wissenschaftsgeschichte: Strukturalistische Methodik: Minimalpaare



Wissenschaftsgeschichte:
Strukturalismus – Ogden/Richards:
Semiotisches Dreieck



CL und die Nähe zur Pragmalinguistik

Systemlinguistik

Pragmalinguistik

Laut

Wort

Satz

Textgrammatik

Diskursanalyse

Textpragmatik

Gesprächsforschung

Textlinguistik

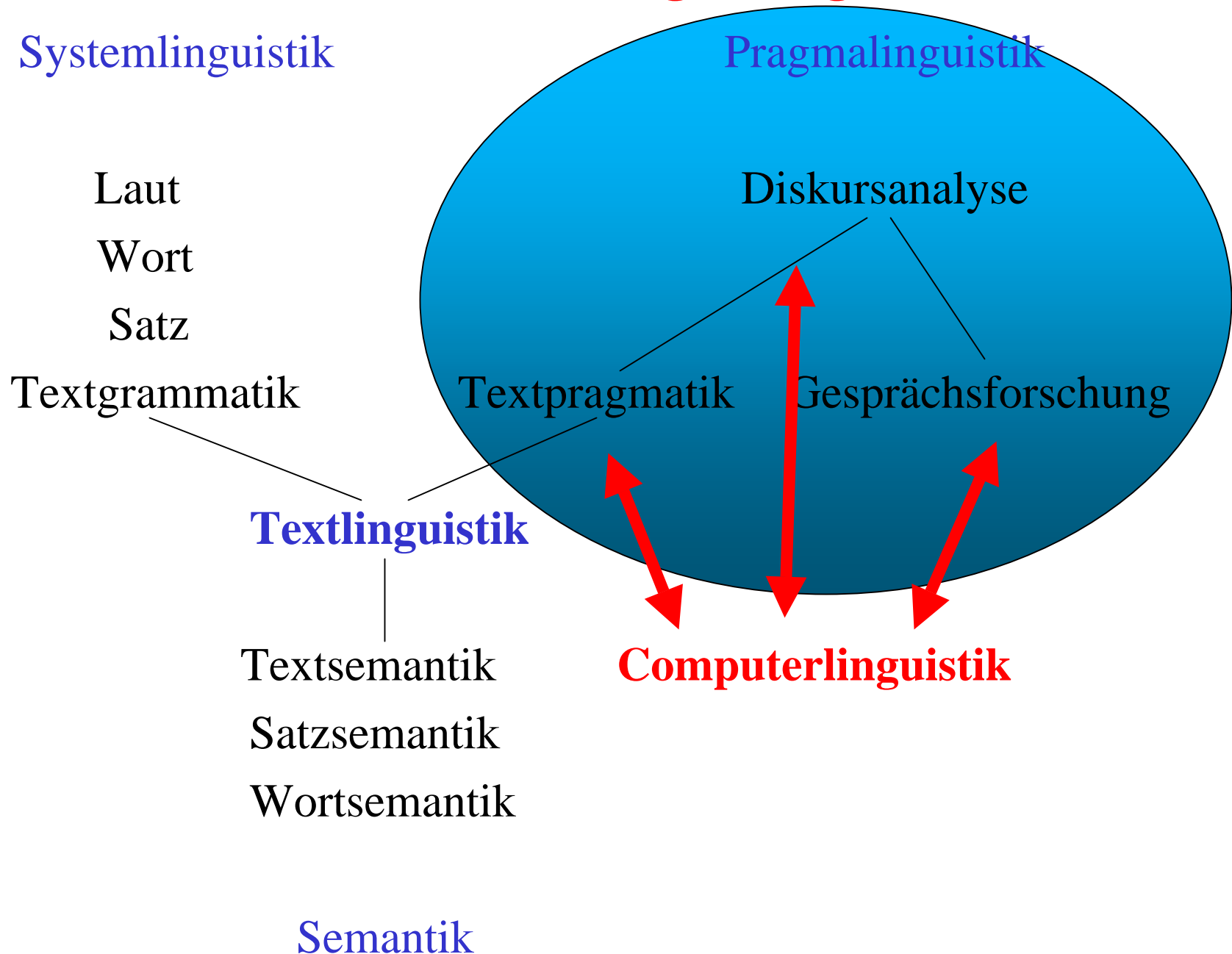
Textsemantik

Satzsemantik

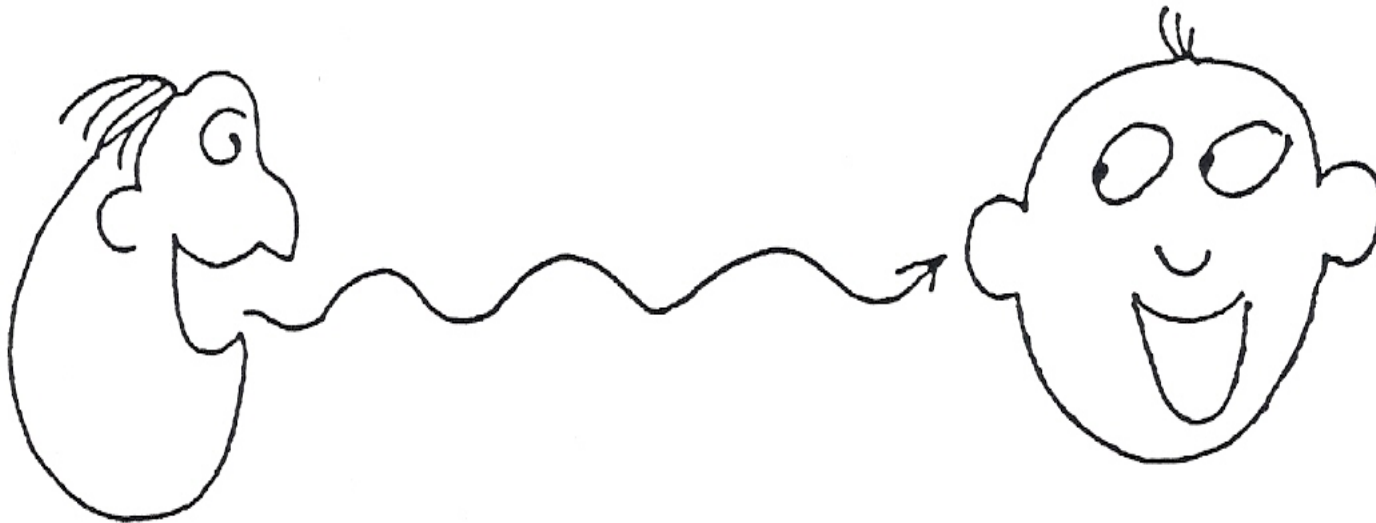
Wortsemantik

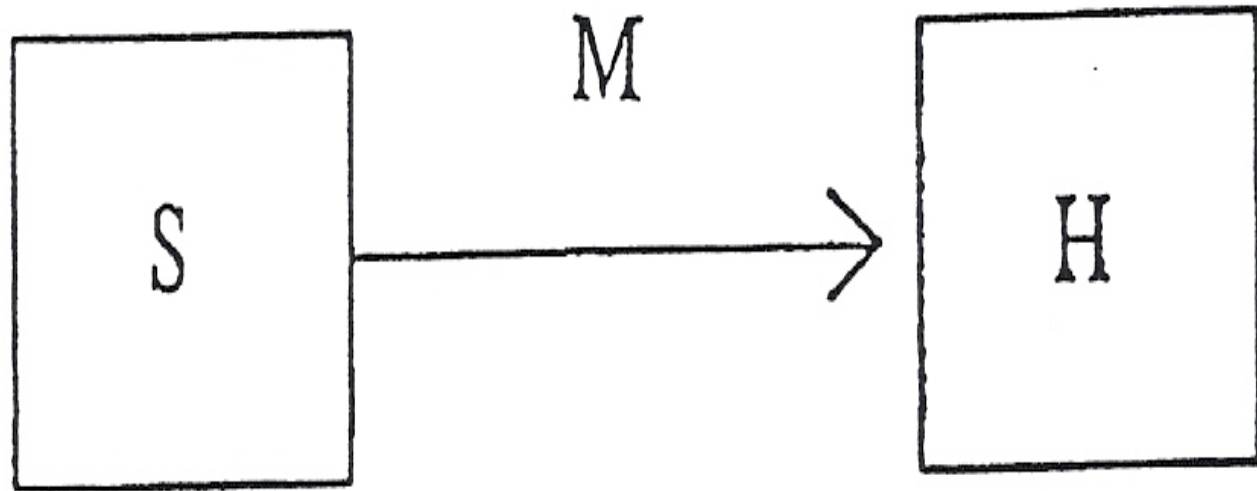
Semantik

Computerlinguistik



Einfaches Kommunikationsmodell

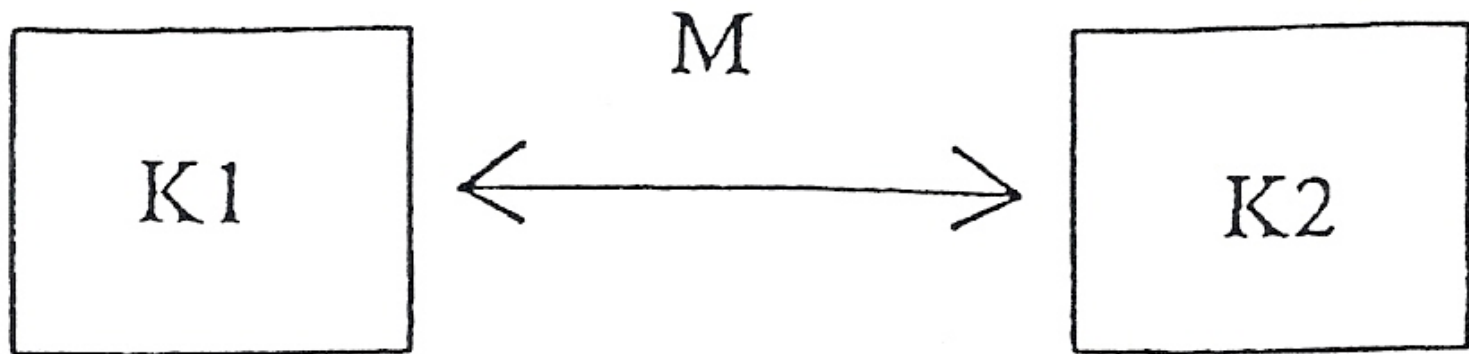




S = Sprecher

H = Hörer

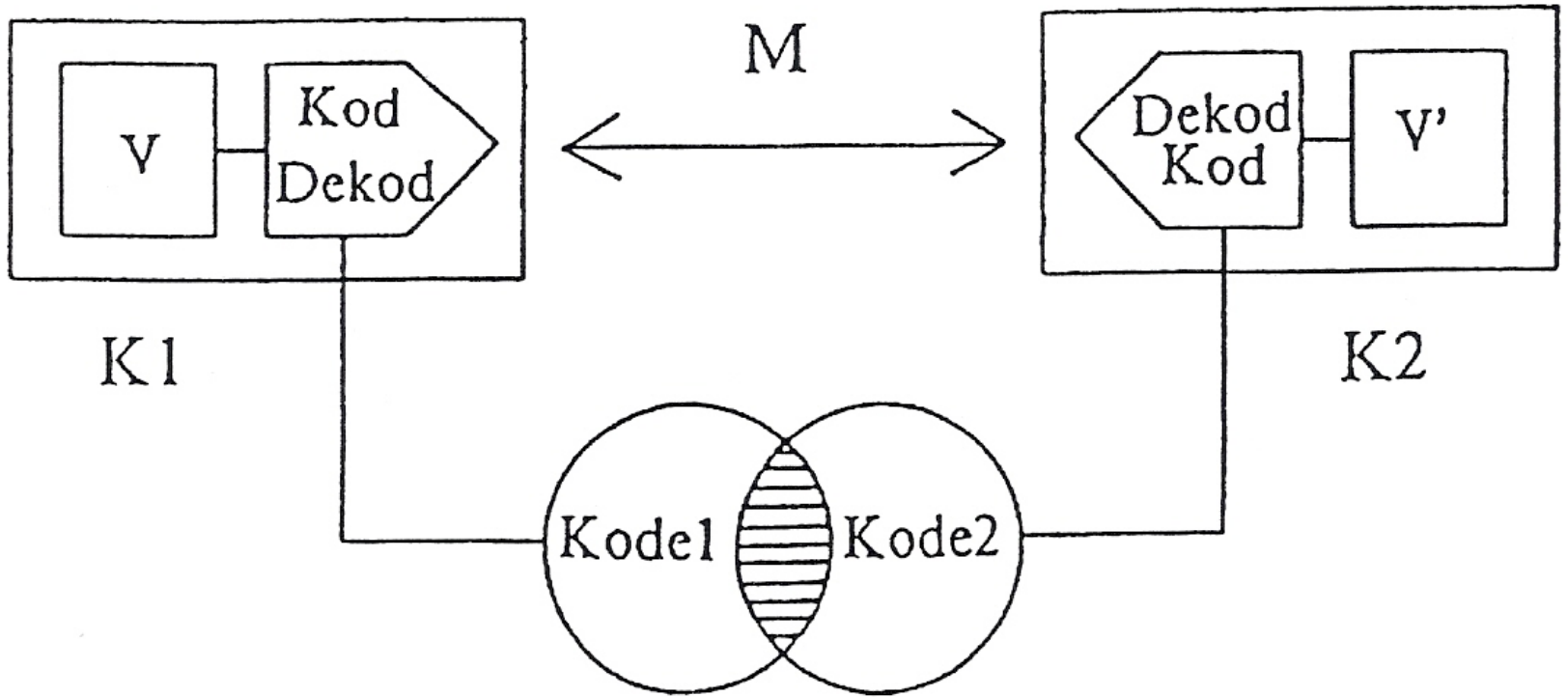
M = Medium



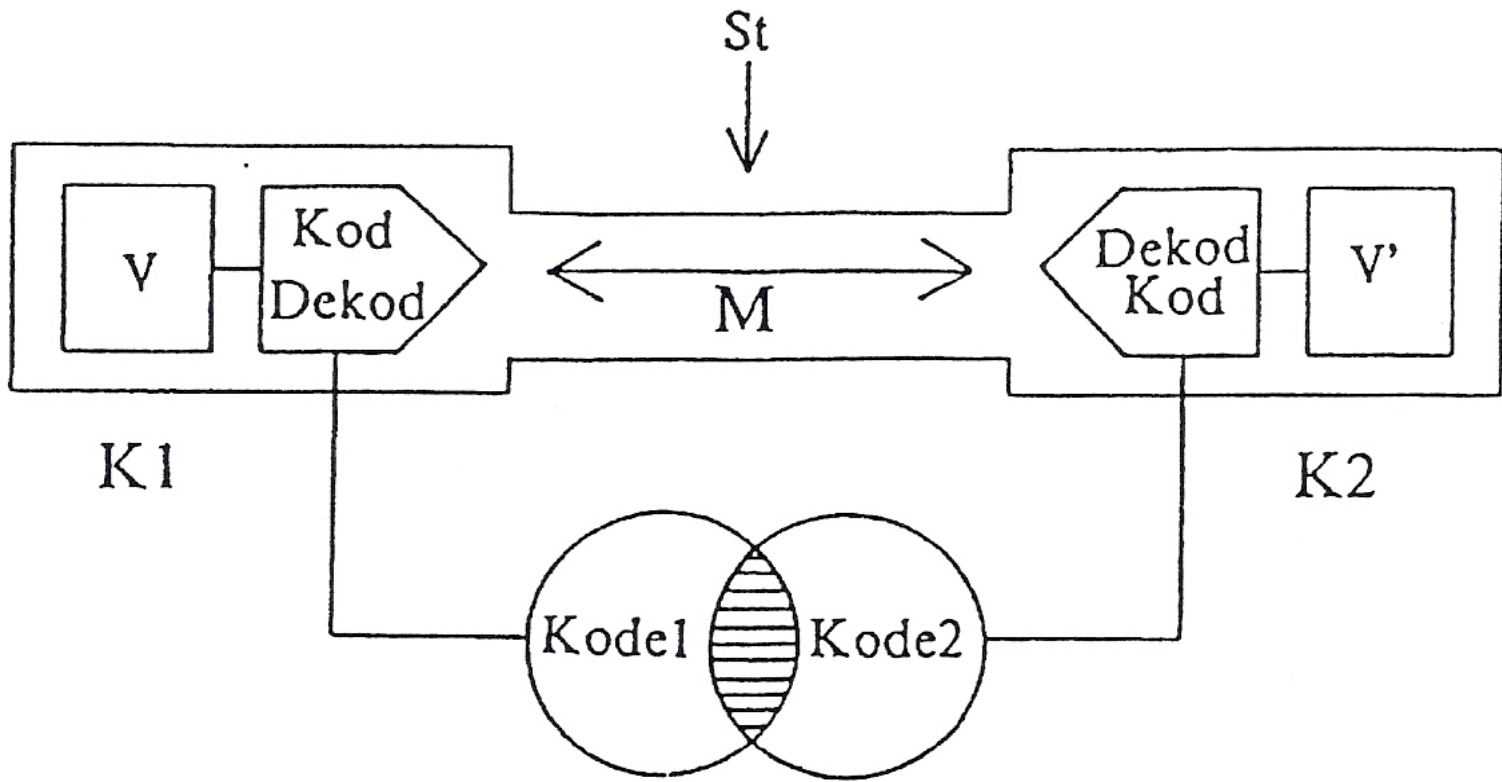
K1 = Kommunikationsteilnehmer 1

K2 = Kommunikationsteilnehmer 2

M = Medium

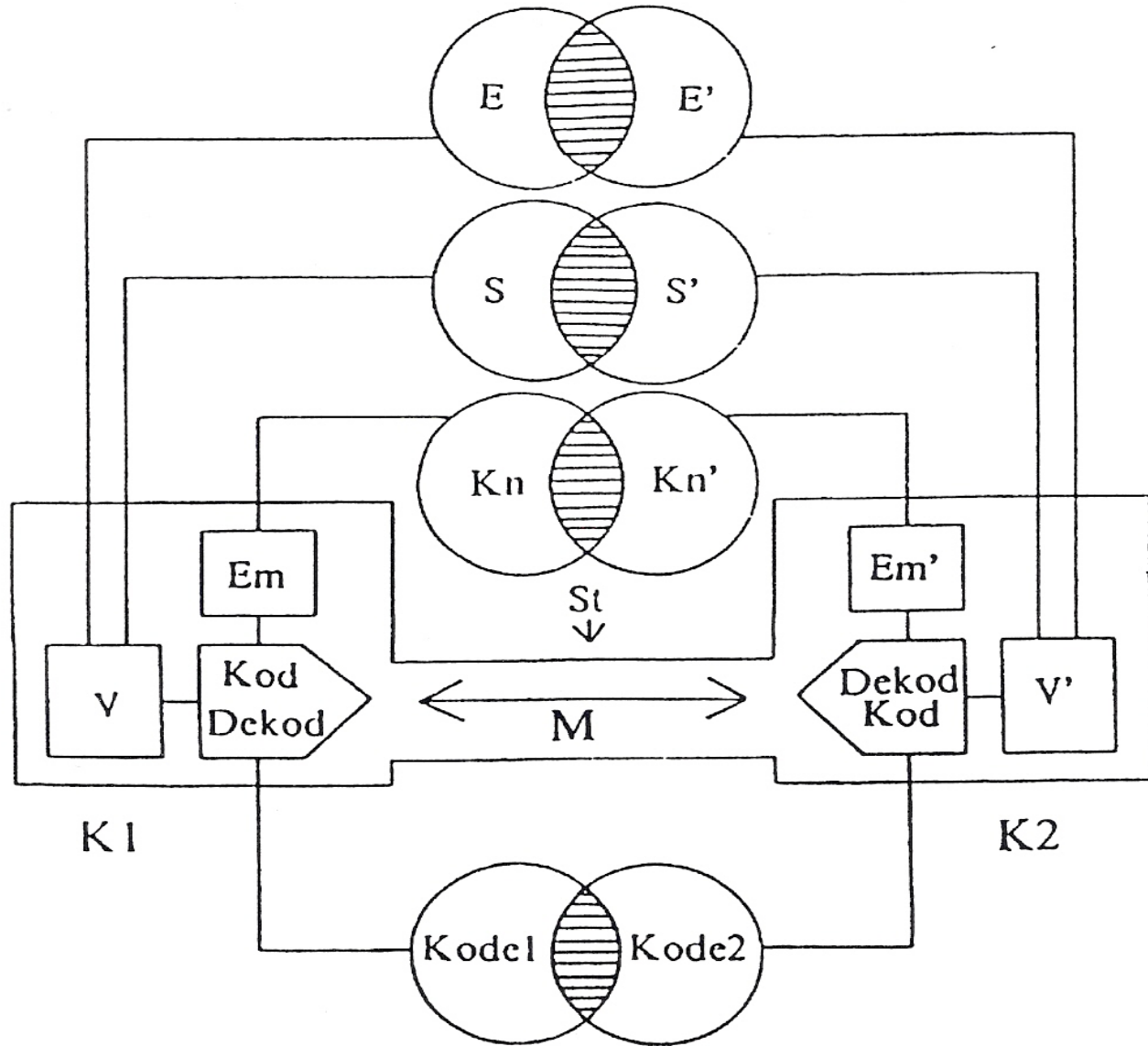


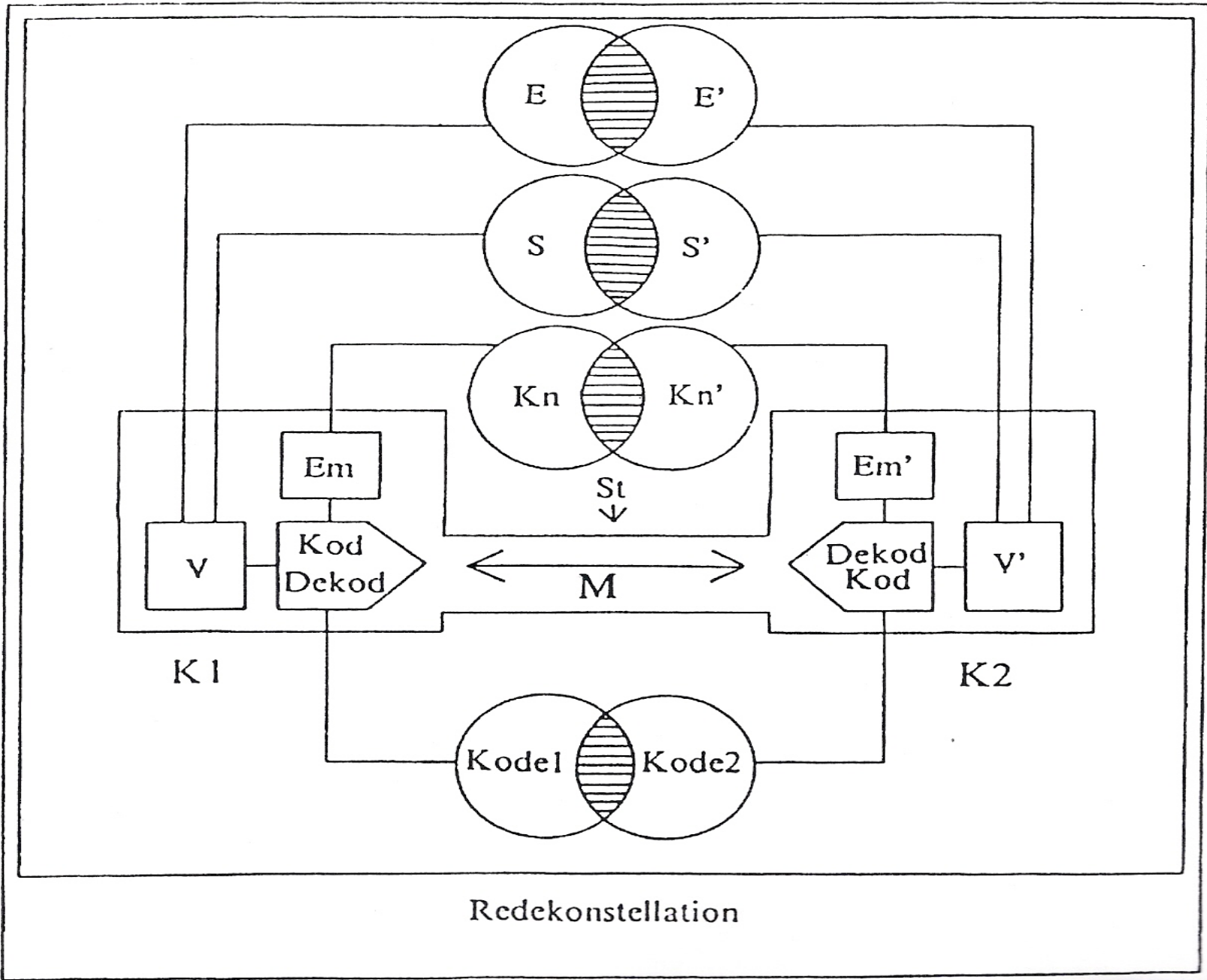
V = Vorstellung
 Kod = Kodierung
 Dekod = Dekodierung



V = Vorstellung
 Kod = Kodierung
 Dekod = Dekodierung
 St = Störung

Em = Emotionalbereich, S = Soziale Sprachschicht, Kn = Konnotat, E = Erfahrungshorizont





Mensch vs. Maschine

Maschinen arbeiten unfreier, beschränkter, eindeutiger und strenger als Menschen. Fehler beispielsweise, schlampiges Sprechen, nachdenkliche Redeweise, Ungenauigkeit, Doppeldeutigkeit, Hintersinn, Ironie, ‚zwischen den Zeilen etwas mitteilen‘ – all das kommt in menschlicher Kommunikation ständig vor.

Für den/die ComputerlinguistIn ist das nicht der Normalfall, sondern eine besondere Schwierigkeit.

Zentrale Fragen der CL in Bezug auf die menschl. Sprache u. Kommunikation

- Was muss ein Mensch können, der eine Sprache ‚kann‘? Welche Teile dieser Fähigkeit können so formalisiert werden, dass ein Computer sie ausführt?
- Woran erkennen Menschen, dass die Schallsignale, die sie wahrnehmen, Sprachlaute und nicht Geräusche sind? Wie kann das ein Computer erkennen?
- Beeinflussen Hören u. Lesen in je anderer Weise das Verstehen von Sprache? Sind beide für maschinelle Verfahren gleich aufwendig?
- Welche einzelnen Aktivitäten werden ausgelöst, wenn Texte verstanden werden? Was müssen Maschinen leisten, die diese Aktivitäten nachvollziehen/übernehmen sollen?
- Welches Vorwissen der Gesprächspartner spielt für eine Mitteilung eine Rolle? Wie kann das im Computer dargestellt werden?

Zentrale Fragen der CL in Bezug auf die menschl. Sprache u. Kommunikation

- Was wird auf welche Weise unmittelbar mitgeteilt? Was wird indirekt, z.B. durch Schlüsse, die der Hörer zieht, mitgeteilt? Wie kann man diese alltäglichen Vorgangsweisen in Regeln fassen?
- Wie erkennen Menschen, um welche sprachlichen Handlungen es geht: z.B. Voraussage, Scheinfrage, Argumentation, Drohung, Befehl, Bitte? Welche Absichten verfolgen Sprecher mit diesen Handlungen? Woran können Computer das ablesen?
- Wie erkennen wir, dass eine sprachliche Äußerung z.B. ein Vorschlag ist? Was wissen wir durch diese sprachl. Handlung vom Menschen, der sie vollzogen hat? Welche Folgen hat es für Menschen, wenn diese sprachl. Handlung (z.B. etw. vorschlagen) ein Computer vollzieht?