

---

## Beteiligte Projekte

- Ausbau und Pflege der Korpora geschriebener Sprache
- Methoden der Korpusanalyse und -erschließung

## Beteiligte Mitarbeiter

- Dipl.-Ing. Cyril Belica
- Dr. Harald Längen
- Dr. Marc Kupietz
- Dipl.-Inf. Rainer Perkuhn

## Kontakt

Dipl.-Ing. Cyril Belica  
Programmbereich Korpuslinguistik  
Institut für Deutsche Sprache  
Postfach 10 16 21  
D-68016 Mannheim

Telefon: 0621/1581-0  
Fax: 0621/1581-200  
E-Mail: [korpuslinguistik@ids-mannheim.de](mailto:korpuslinguistik@ids-mannheim.de)



 INSTITUT FÜR  
DEUTSCHE SPRACHE

R 5, 6–13  
D–68161 Mannheim  
[www.ids-mannheim.de](http://www.ids-mannheim.de)

Mitglied der  
  
Leibniz-Gemeinschaft

*Das Institut für Deutsche Sprache (IDS) ist die zentrale Einrichtung zur Erforschung und Dokumentation der deutschen Sprache in ihrem Gebrauch und in ihrer neueren Geschichte. Es gehört zu den 86 außeruniversitären Forschungseinrichtungen und Serviceeinrichtungen für die Forschung der Leibniz-Gemeinschaft.*

---

 INSTITUT FÜR  
DEUTSCHE SPRACHE

Programmbereich Korpuslinguistik

# Empirisch- linguistisches Forschungs- programm

*„Auch spielt das Zufällige, Unberechenbare, Inkommensurable eine zu große Rolle. Unsere Gesetze fußen nur auf Wahrscheinlichkeit, auf Statistik, nicht auf Kausalität, treffen nur im allgemeinen zu, nicht im besonderen. Der Einzelne steht außerhalb der Berechnung . . .“*

Friedrich Dürrenmatt: Das Versprechen. Diogenes Verlag, Zürich 1958.

---

Mit seinem empirisch-linguistischen Forschungsprogramm verfolgt der Programmbereich Korpuslinguistik das Ziel, neue Einsichten in Wesen, Eigenschaften und Regelmäßigkeiten von Sprache zu gewinnen. Dabei wird Sprache v.a. als ein (psychologisch und sozial) reales Phänomen verstanden und alles Regel- und Systemhafte an diesem Phänomen im Sinne Paul Hoppers als *emergent* betrachtet. Im Fokus stehen alle linguistischen Dimensionen von Sprache (Semantik, Syntagmatik, Paradigmatik, usw.).

Wie jeder explanatorische Ansatz ist auch dieses Programm mit einer breiten *explanatorischen Kluft* zwischen der angestrebten theoretischen Beschreibungsebene und der phänomenologisch zugänglichen Datenebene konfrontiert. Die zentrale Herausforderung ist daher die wissenschaftstheoretische Frage, wie man angesichts dieser Kluft überhaupt zu hinreichend abgesicherten Erkenntnissen über das Phänomen Sprache gelangen kann.

Im Programm werden diesbezüglich Forschungsstrategien erarbeitet, die sich insbesondere an den folgenden Prinzipien orientieren:

- minimale Annahmen über Sprache
- frühzeitige linguistische Klassifikationen vermeiden (soweit möglich)
- *bottom-up*: Induktion und Abduktion
- *top-down*: Deduktion und Falsifikation
- Vorgehensweise in kleinen Schritten
- Integration verschiedener Datentypen

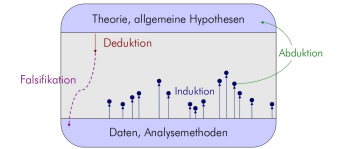
Ein zentrales Instrument ist hierbei unser *gläsernes Labor* CCDB.

Stand: r1944 / r1501 - 2012-04

## Was ist eine geeignete Datengrundlage für die empirische Untersuchung der deutschen Gegenwartssprache?

### Was ist eine ausgewogene Stichprobe der deutschen Sprache?

Wie verhält sich sprachliche Variation zu sprachlicher Norm? Woran kann der Begriff „Standardsprache“ festgemacht werden?  
Sagt man „der“, „die“ oder „das Nutella“?



### Worauf zielen sprachwissenschaftliche Theorien?

Wie ist die Wechselwirkung zwischen dem individuellen Sprachwissen und sprachlichen Konventionen?

### Worauf gründen sich sprachwissenschaftliche Theorien?

Welcher Natur sind sprachliche Daten?

Was sind authentische Sprachproduktionsprozesse?

### Welche Eigenschaften sprachwissenschaftlicher Theorien sind erstrebenswert?

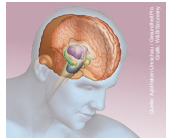
Deskriptiv? Normativ? Präskriptiv? Explanatorisch? Ist empirisch-linguistische Forschung vereinbar mit den Grundprinzipien einer explanatorischen Theoriebildung?  
Vollständig? Global widerspruchsfrei? Falsifizierbar? Wahr?

### Wie entstehen sprachliche Regularitäten und wie entstehen sprachliche Regeln und Normen?

### Ist Sprache ein eigenes, vom Sprachgebrauch unabhängiges System?

Was bedeutet die Unterscheidung zwischen Kompetenz und Performanz?  
Was die Vorstellung eines „idealen Sprechers“?

Beziehen sich linguistische Kategorien auf *psychologisch* reale Inhalte unseres *prozeduralen* sprachlichen Wissens?



### Wie weit ist Sprache kompositionell und regelbasiert?

### Wie weit ist sie assoziativ und präferenzuell?

Basiert das Systemhafte in der Sprache auf harten Regeln über identischen Ereignissen?  
Oder auf weichen Regularitäten über ähnlichen Ereignissen?  
Wie können Regularitäten dieser Art aufgedeckt werden?

### Inwiefern ist eine symbolische Repräsentation von „Bedeutung“ möglich?

Wie lässt sich „Bedeutung“ fassen, wenn sie sich erst über den Gebrauch im Kontext einstellt?

Wie lässt sich der Diskurs erfassen, in dem sich die „Bedeutung“ konstituiert?

### Wie sieht das Lexikon-Syntax-Kontinuum aus?

Ist Idiosynkrasie, Idiomatisches, Kollokation nur ein Randphänomen von Lexis und Syntax?

Oder sind diese Phänomene Ausdruck einer aussagekräftigeren Vorstellung davon, was Sprache in ihrem Innersten zusammenhält?

### Wo kommt unser Sprachgefühl her?

Wie verhält sich das Alter von Ereignissen und die Wiederholung von ähnlichen Ereignissen zu Verfestigung oder Verblässen sprachlich-kognitiver Routinen?



### Woraus leiten wir ab und wie einigen wir uns darüber, wie man spricht?

sind Selbstorganisation und Emergenz die fundamentalen Erklärungsansätze in der Sprachwissenschaft?