Approaching Grammar:
Towards an Empirical Linguistic Research Programme

Marc Kupietz

18 March 2008
GLOBAL COE INTERNATIONAL WORKSHOP

Corpus Linguistics Programme Area
at the Institute for the German Language

- Cyril Belica
- Rainer Perkuhn
- Holger Keibel
- Marc Kupietz

Overview

I. terms and definitions

II. basic decisions for empirical linguistics

III. an empirical linguistic research programme

Theory

A set of assertions about “something”
with (roughly) the following properties:

- precisely formulated (unambiguous)
- rationally justified (not metaphysically)
- logically consistent (no obvious contradictions)

Kinds of Theories

distinction between:

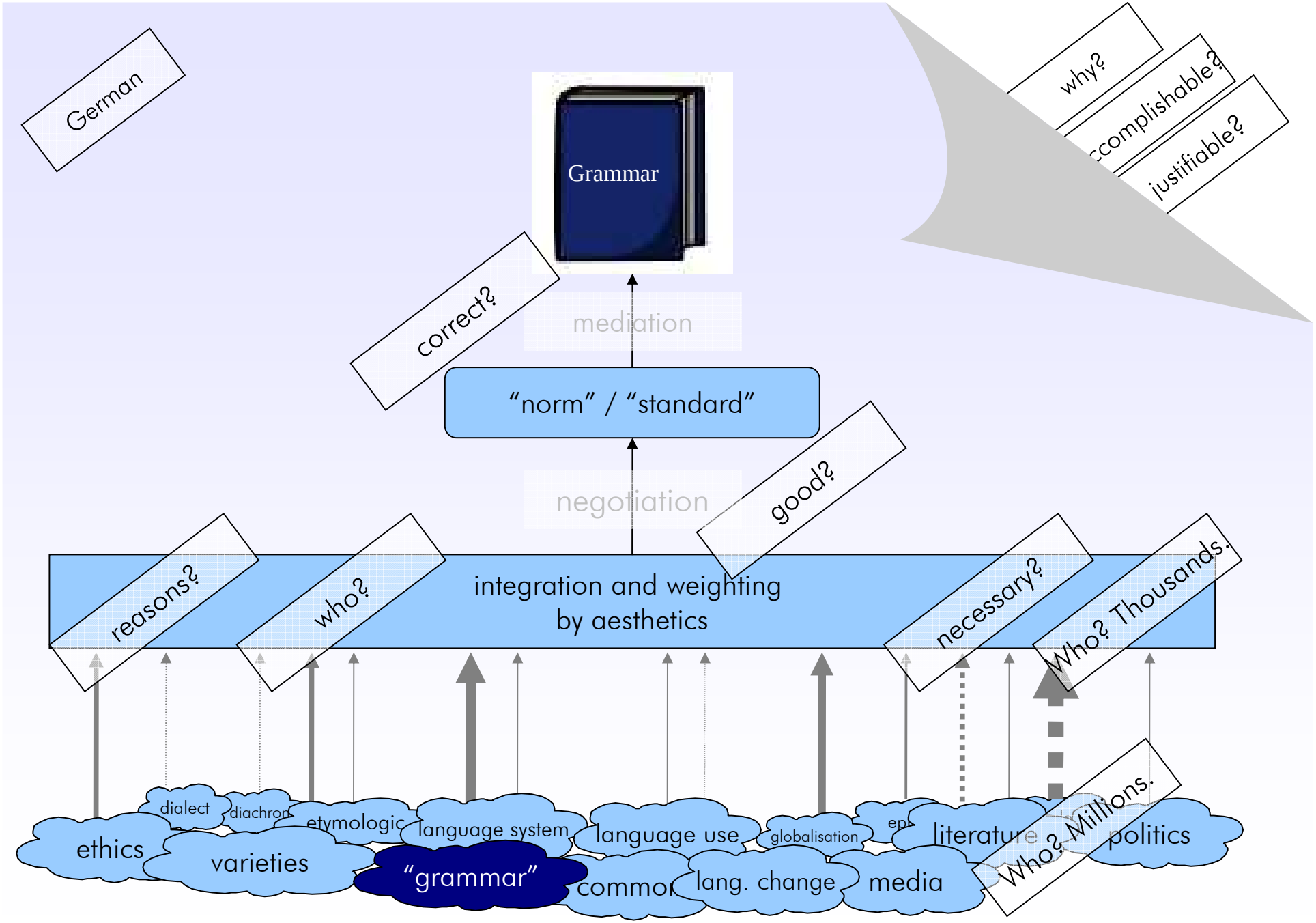
- explanatory theories
- descriptive theories
- normative theories
- …

Explanatory Theories

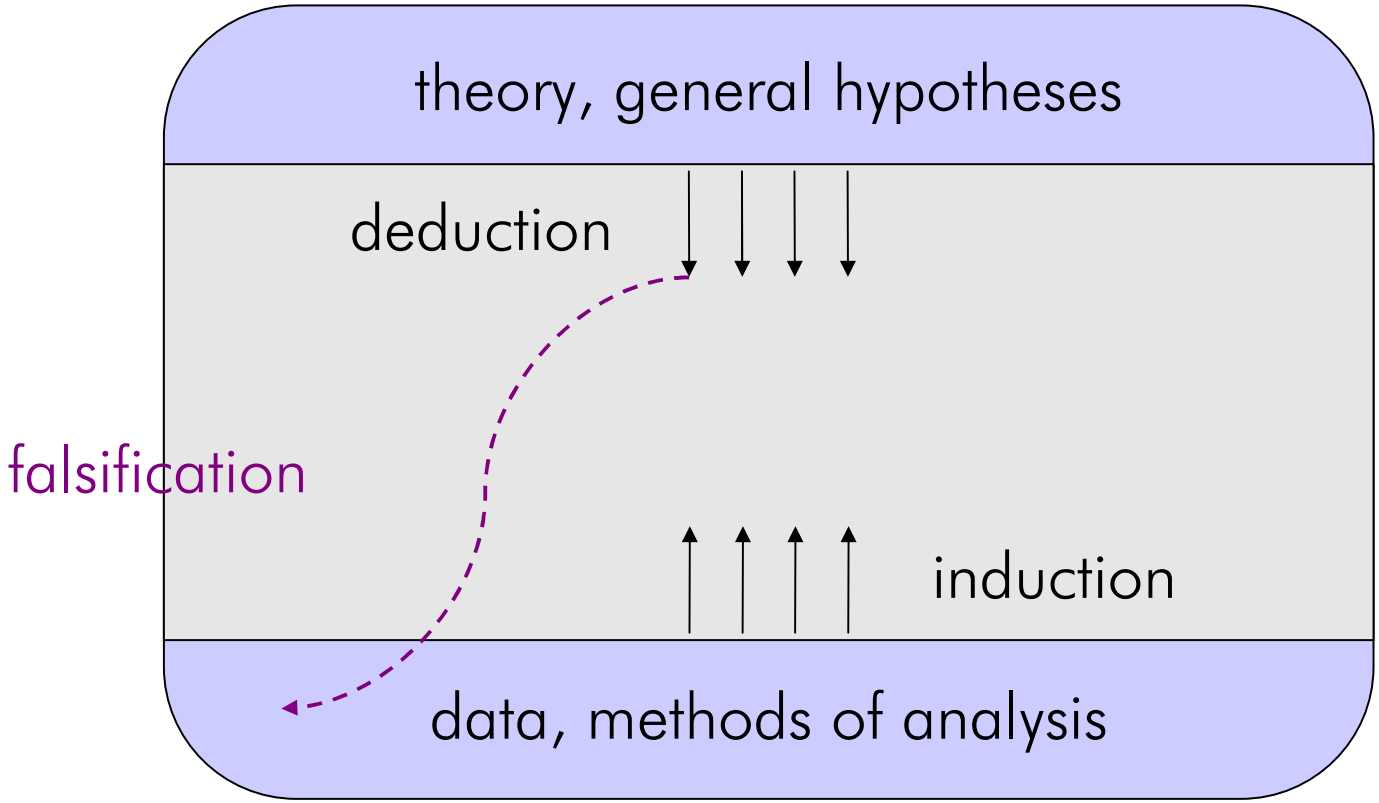
- subject: things from the *real* world
- empirically testable (*falsifiable*)
- propositions and hypotheses connected by more general principles
- motivation:
  - gain of knowledge
  - explanation

Normative Theories

- subject: things from an *ideal world*
- not falsifiable
- motivation/justification:
  - aesthetic
  - ethic
- examples:
  - theories of society
  - theories of science
  - in areas of law
  - theories of individual languages (normative grammars)


German
Grammar
why?
accomplishable?
justifiable?
correct?
mediation
"norm" / "standard"
negotiation
good?
reasons?
who?
integration and weighting
by aesthetics
necessary?
Who? Thousands.
Who? Millions.
dialect
diachron
etymologic
language system
language use
globalisation
literature
politics
ethics
varieties
"grammar"
lang. change
media

Explanatory Gap



Empirical-Linguistic Approach

- empirical-linguistic research aimed at explanatory theories
- decision against:
  - language conceivable as a formal system
  - decomposition universally applicable as a method of explanation
  - completeness as a goal
  - competence as the subject of research

Language as a Formal System

- assumption that *natural* language is explanatorically adquate conceivable as a formal system
  (→ assumption of formal rules)
- not sufficiently justified because …
  - natural language is a dynamic cultural artefact
  - and the producers of this artefact (speakers) are not (in any explanatorically adequate way) conceivable as a formal system (e.g. as a symbolically programmed computer)

Principle of Decomposition

- prerequisites for an analytical-explanatory methodology of decomposition are not sufficiently given:
  - this means:
    the real existence of perfect hierarchical structures cannot be assumed a priori (neither with respect to language nor with respect to the mind)
  - therefore:
    it is an empirical question whether and to what extent decomposition can be used as a methodology for explanation (e.g, in modules, constituents, categories ...)

Completeness as a Goal

- completeness as a goal of a theory of an individual language
  - completeness ≈ descriptive adequacy
- not *achievable*:
  - language = dynamical artefact
    → *every formal rule can be violated, without leaving the language.*
- not *helpful*:
  - theory becomes increasingly complicated
    → loses its explanatory power
    → incompleteness no argument

Theories of Competence


- linguistic theories that take the competence of an ideal speaker/hearer as their subject
- more like normative theories:
  - create the image of an *ideal* speaker/hearer (not a thing from the real world!)
  - *not* empirically falsifiable
  - aesthetic motivation?
    - idol: turing machine: *ideal* computer
    - oriented towards mathematical elegance
    - "a technical fantasy" (Stetter, 1999)

language is language use

- usage-based perspective on language!
- language system (*langue*) is
  - no indepently existing object
    with definite and fixed properties
  - but an *emergent* phenomenon of language use (*parole*)
- see *usage-based framework:*
  - *emergent grammar* (Hopper 1987, 1998)
  - *emergent lexicon* (Bybee 1998)
  - *Cognitive Grammar* (Langacker 1987)
  - *Construction Grammar* (e.g. Goldberg 2006)
  - lang. acquisition (e.g. Elman et al. 1999, Tomasello 2003)

How does emergence arise?

the individual language experience …

→ forms procedural language knowledge (*entrenchment* etc.)

= cognitive routines for production and processing of language material

(also see Sinclair's *idiom principle*, 1991 and Hoeys *lexical priming*, 2005)

→ leads to individual "grammar" *in the head*

→ influences speech productions of the individuum, and thereby also contributes to the language experience of others

How does emergence arise? (ctd.)

- regularities/the "rule-like"
  = conventions of the language community,
  with varying degree of distribution among language users
  and with varying stability
- emergence: continuous negotiation of conventions
  - negotiation potentially in every single language interaction
  - language community: "negotiates" conventions distributed over many such interactions
    (*"coordination without coordinator"*, Keller 2006)

Main Reasons for High Degree of Conventionality

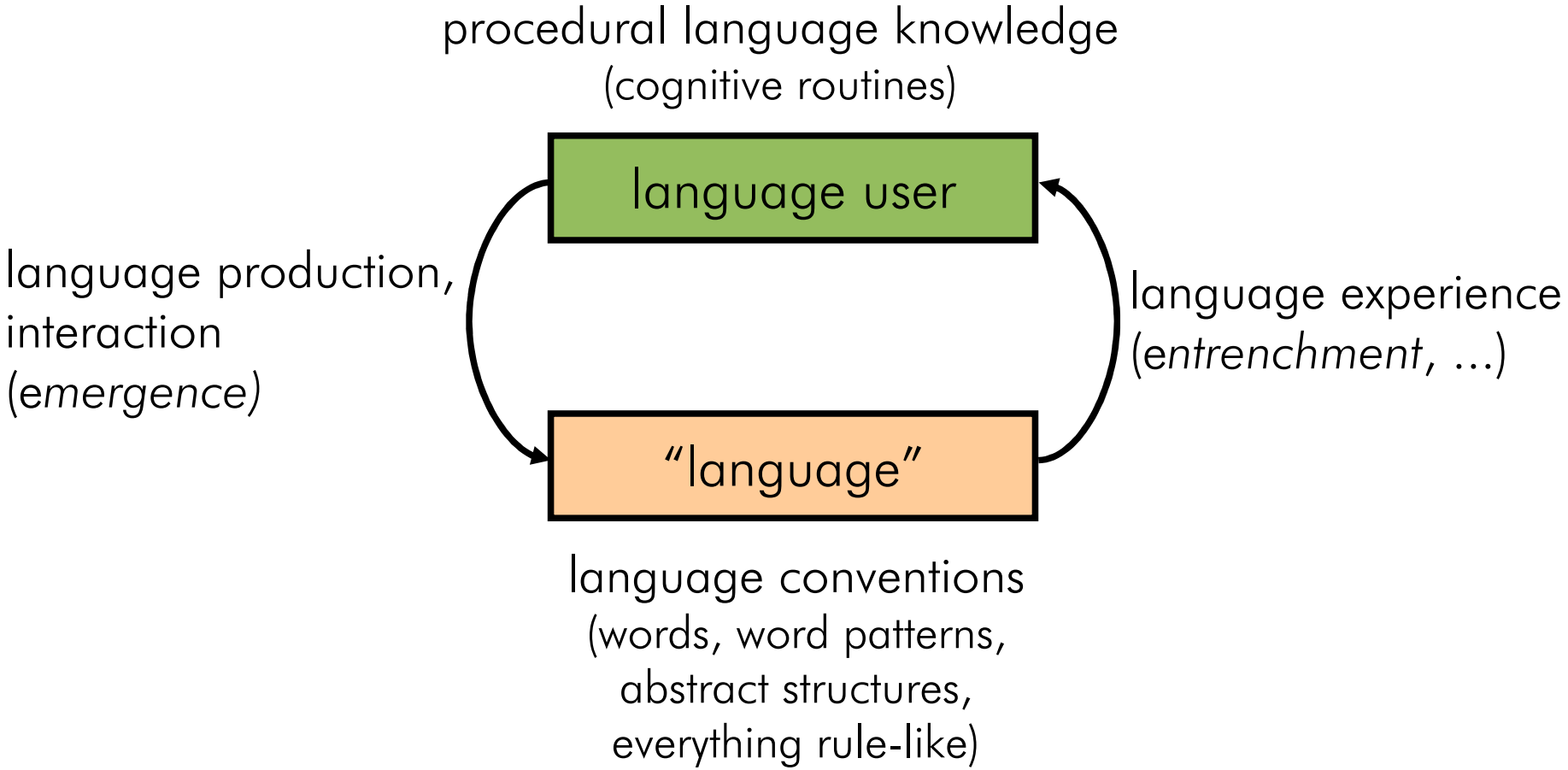
- same species
- same world

} ⇨ language universals (see e.g. Tomasello 2003)

- main stabilising factor ("attractor") for individual languages: dependency on communicative success

- other reasons
  (but already included in language use !!!):
  - media
  - school
  - normative grammars, …

How does Emergence arise? (schema)



Rule vs. "Rule-like"/Regularity

What is the status of grammatical rules
and how do they arise?

– *rules* as stable, authoritative entities do not exist!

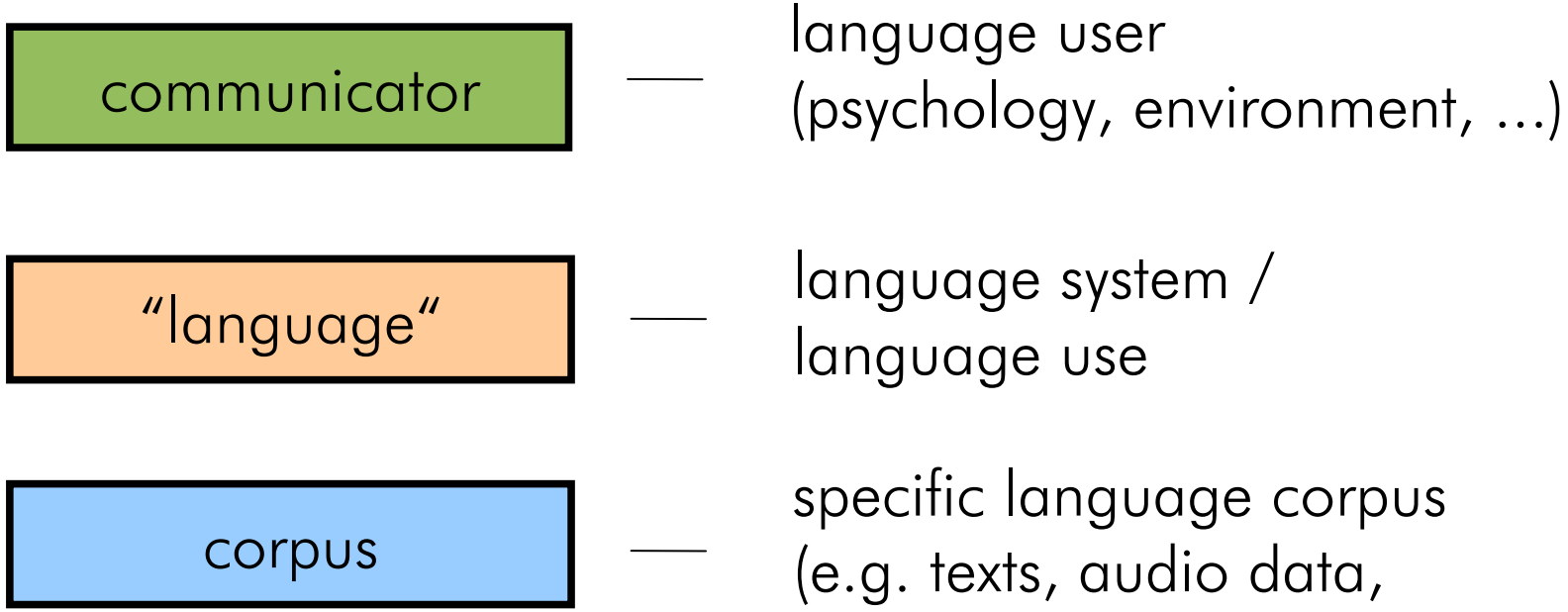
– but instead:

- regularities in language use
- emergent structural conventions
  - conventionalised to different degrees
  - always instable, dynamic, ***dependent on context***, adaptive:
    - may (dependent on context) always be "violated" in a reasonable/meaningful way
    - may change over many of such "violations"

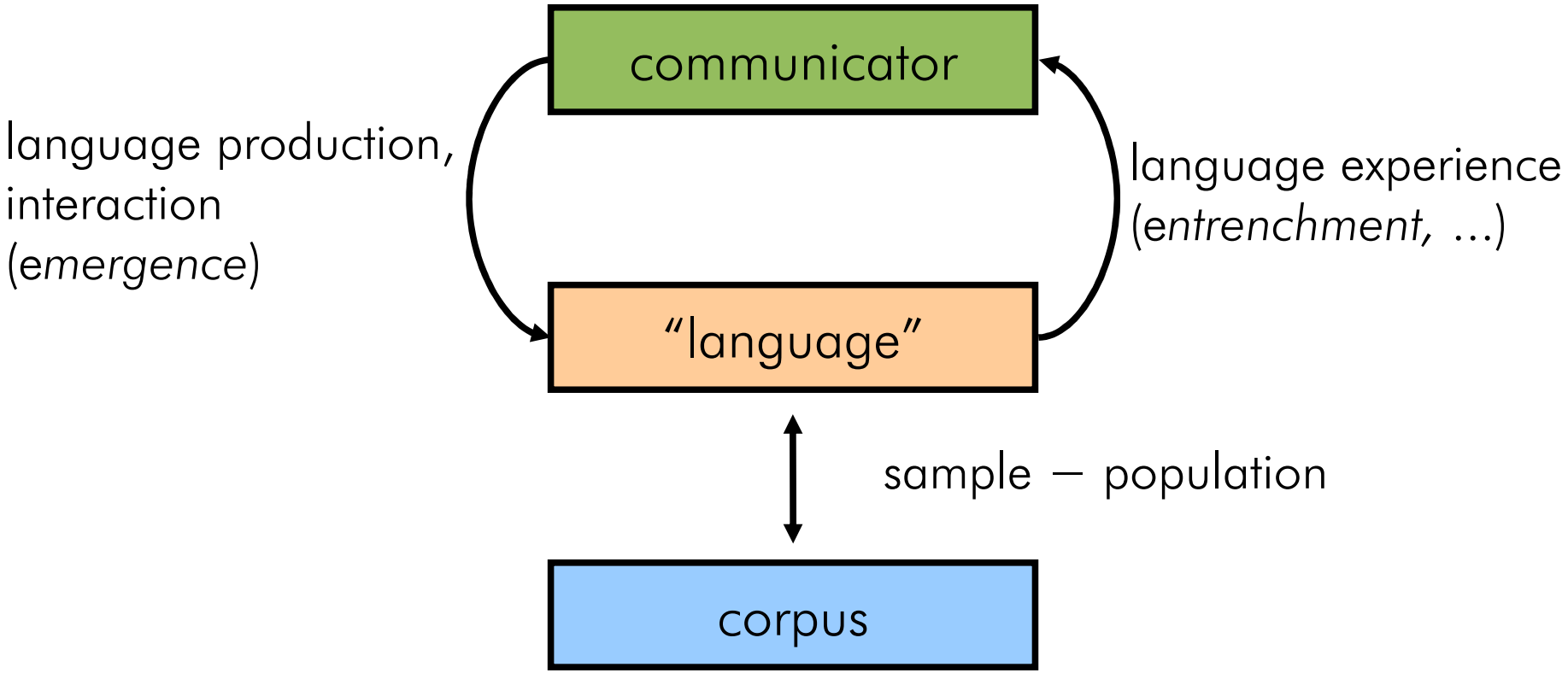
Objects(s) of Research

an empirical-linguistic research programme

- goal: explanatory theories
- objects of examination at three levels



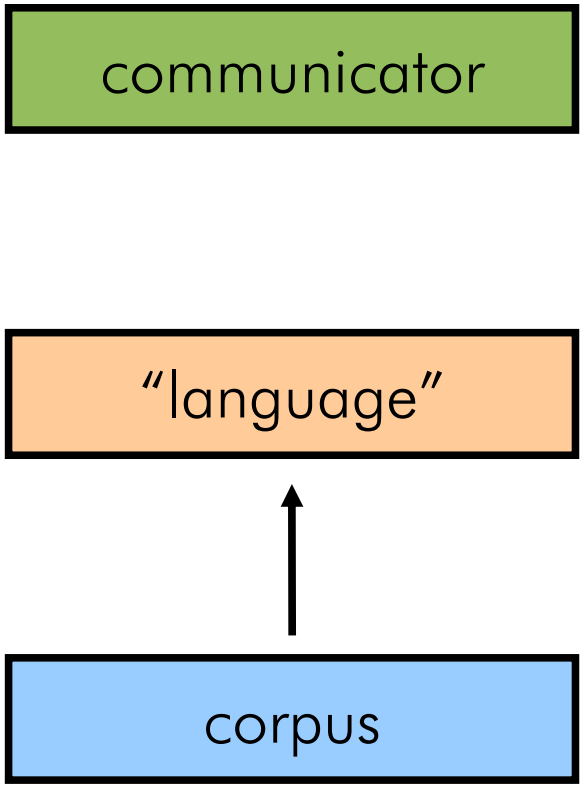
Relations between the Levels



Verification of Hypotheses

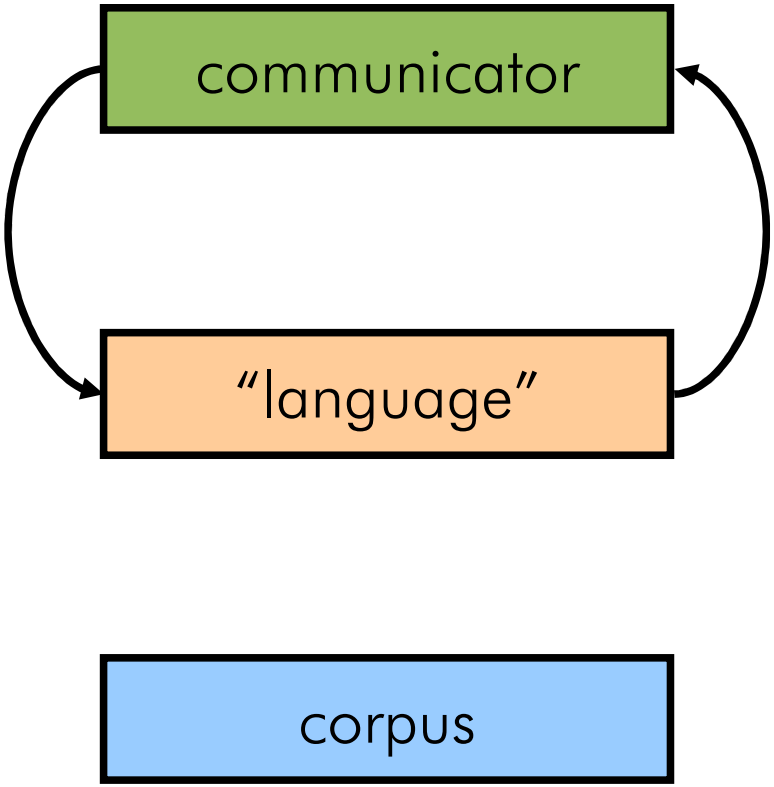
communicator	—	experimental psychology (outside of empirical linguistics)
"language"	—	by means of other suitable corpora
corpus	—	directly by the given corpus

How to get to Hypotheses?

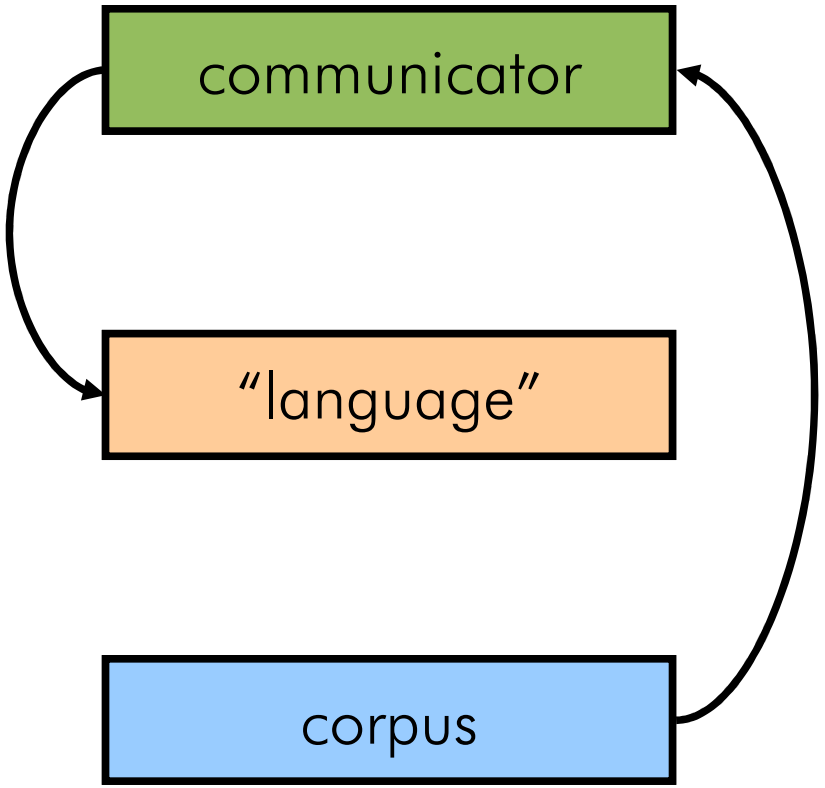


Methodological Access

(making use of the loop of mutual influence)

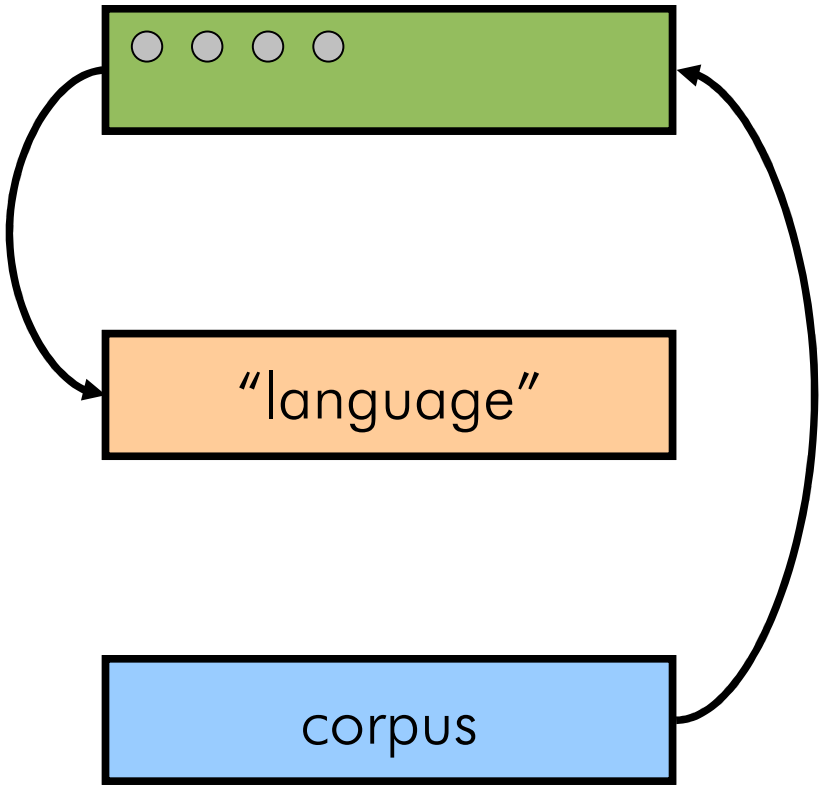


Methodological Access
(trace of generalisations)



Methodological Access

(summary: collect premises, operationalise, apply)



Methodological Access

- justification:
  - psychological facts, that play an important role in processes of generalisations also have a strong influence on the resulting generalisations
    - in the individuum
    - and hence also in the language system/language use
- problem only shifted to psychology?
  - not really: psychological premises can be…
    - selected conservatively
    - tested (psychological experiments)
- stepwise generalisations from corpus to language

Psychological Premises

- first general premise:
  the ability to recognize *similiarities*
  - more precisely: the unability to ignore similarities
  - necessary for living
  - fundamental role in all processes of learning:
    - situations/contexts are never identical
    - without sensitivity to similarity, former experiences would never be usable for future behaviour,
      in short: not generalisable
- aspects of similarity have an influence …
  - on *all* generalisations in the individuum,
  - and hence also on generalisations in the language system

Similarity

- very broad notion:
  gradual counter-concept to i*dentity*
- aspects of similarity:
  - perceptual and conceptual similarity
  - analogies
  - associations (co-occurrence = similarity on time dimension)
  - similarity of entities and similarity of contexts
    (again: very broad notion of context!)
  - ...
  - which aspect are most important for language is an open question

Preference Relations

- generalisations over similarity relations
- between pairs of "objects" of *every* kind
  - objects inside and outside language (context)
  - elementary and inferred objects
  - concrete and abstract objects
  - sets of objects
  - preference relations
  - formations of preference relations
  - ...

Preference Relations – Rules

- interest of linguistics in general: discovery of structures = relations, formations of relations
- formal rules: hard, categorial relations
- preference relations
  - soft, gradual, context-dependent relations
  - quantifiable (e.g. conditional probabilities)
  - include categorial relations (the extreme case)
- norm and variation
  - same status
  - variation is not just *noise* – language **is** variation

Locality

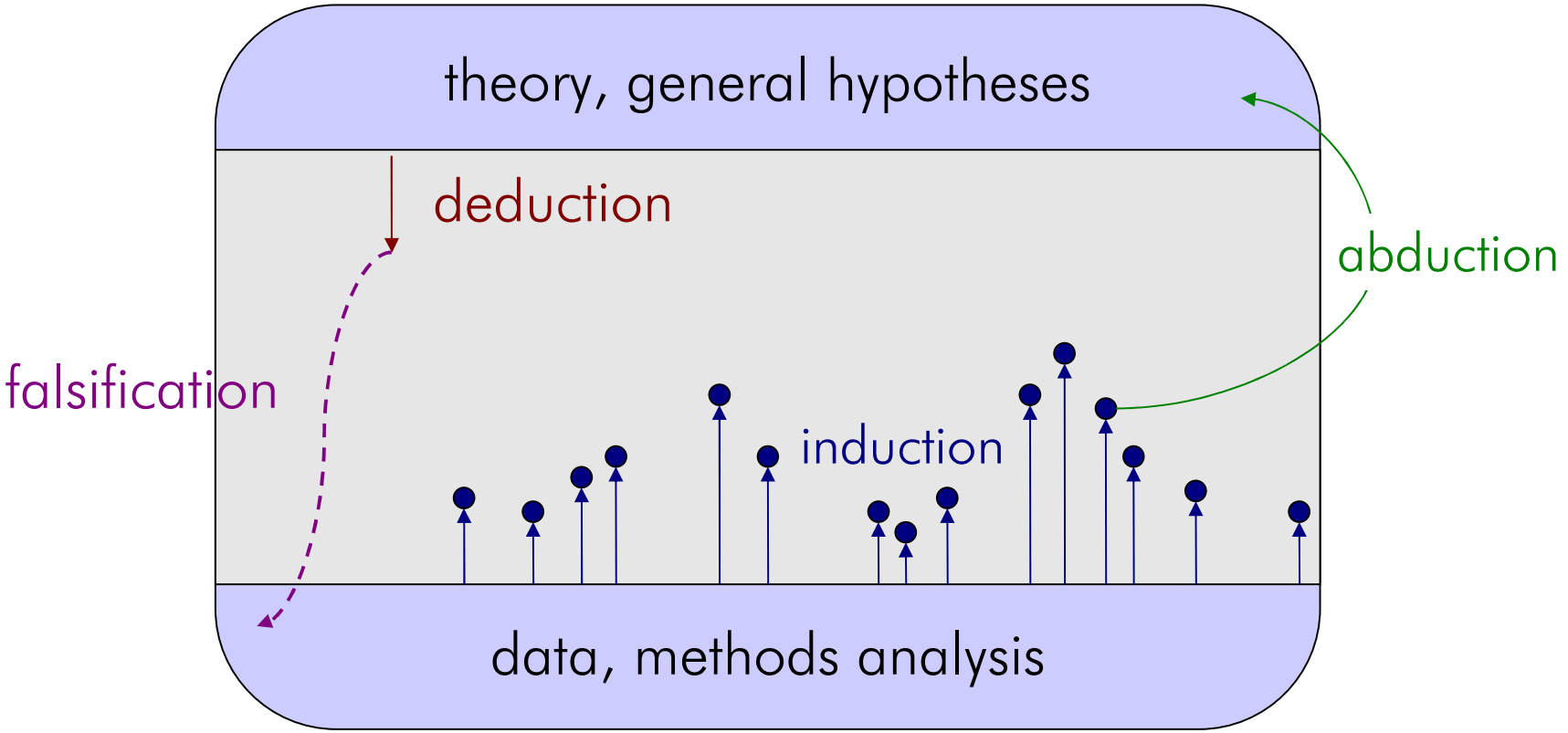
"Locality is arguably the most powerful simplifying assumption in all of physics. I suspect the physics community has been far too cavalier in abandoning this principle in favor of the elegance of their mathematical models."

(Paul Budnik 2007)

Local Modelling

- as methodology of simplifying the object of research
- local models
  - "local": only valid under certain conditions
  - potentially overlapping: redundant or contradictory
  - potentially not unifiable to larger models
- example:
  - every preference relation
    together with its underlying objects

Local Models and the Explanatory Gap



Thank you!

korpuslinguistik@ids-mannheim.de