

Converging evidence: linguistic typology and corpus linguistics

Holger Diessel & Kasten Schmidtke
University of Jena

holger.diessel@uni-jena.de
<http://www.holger-diessel.de>

Language System

Phonology

Morphology

Syntax

Semantic

In order to study the language system you have to abstract away from aspects of language use.

Language system



Competence

Language use



Performance

Usage-based linguistics

The linguistic system is grounded in language use.

- Linguistic structure emerges from language use.
- Grammar is a dynamic system that is shaped by the psychological mechanisms involved in language use.

Theoretical vs. empirical linguistics

- Theory-building is grounded in empirical research
- Empirical research involves quantitative data
- Methods of quantitative data analysis

Synchronic vs. diachronic linguistics

- If grammar is a dynamic system we need to know how it emerged and develops.
- Historical linguistics (e.g. grammaticalization) and L1 acquisition have become cornerstones of usage-based linguistics.

Linguistic subfields

- The usage-based approach takes a more holistic perspective on linguistic phenomena.
- It often combines evidence from different subfields -> converging evidence.
- The study of grammar involves the study of language use.

Converging evidence from LT and CL

Linguistic typology is mainly concerned with grammatical conventions, whereas corpus linguistics is concerned with usage data.

If grammar is grounded in usage, we would expect to find the same underlying tendencies/principles in linguistic typology and corpus linguistics.

Class schedule

- The usage-based model / converging evidence
- The linear structuring of complex sentences in cross-linguistic perspective
- The positional patterns of causal, conditional and temporal clauses: Converging evidence from linguistic typology and corpus linguistics
- Purpose clauses: Converging evidence from linguistic typology and corpus linguistics

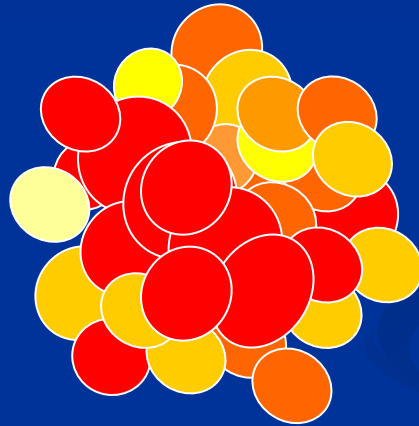
What are the psychological mechanisms that affect grammar?

How does frequency affect the emergence of grammar?

Entrenchment



Entrenchment



exemplar-based
category

Exemplar theory/view

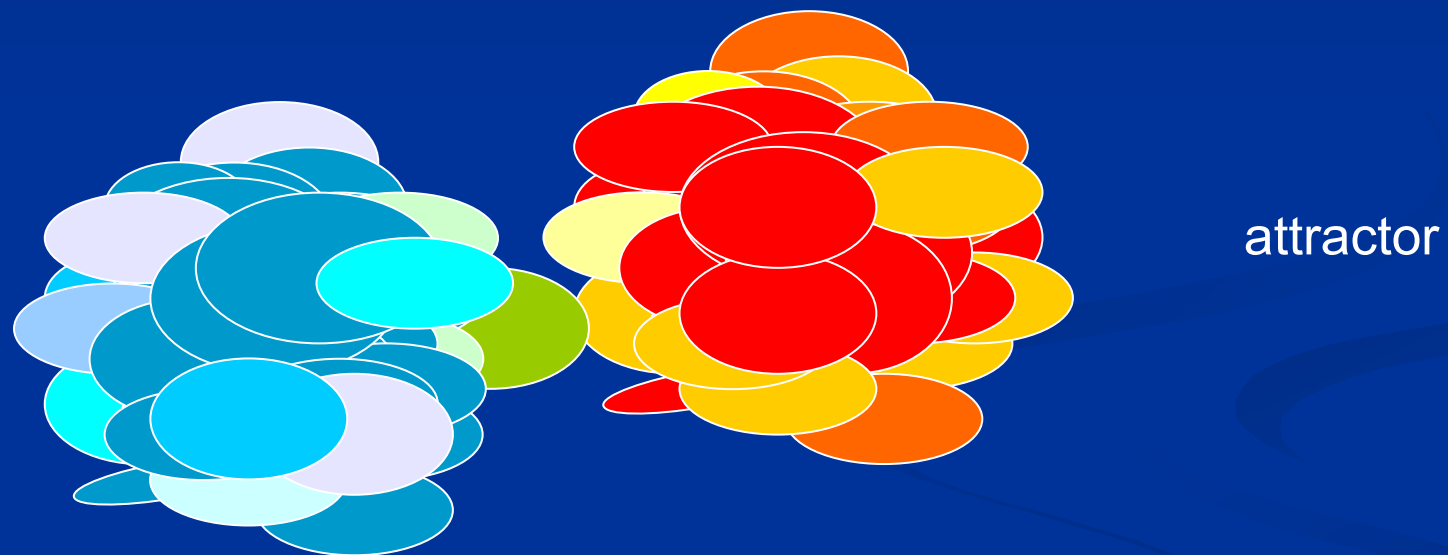


attractor

Exemplar theory/view



Exemplar theory/view



Automatization

$W_1 W_2 W_3 W_4 W_5 \dots$



Automatization involves:

1. emergence of chunks of prefabricated structure
2. loss of linguistic boundaries
3. reduction

Analogy

Analogy is a cognitive process of transferring information from a particular subject, the source, to another particular subject, the target. (Gentner 2004)

Old English

climb	clomb
step	stope
laugh	low

Modern English

climb	climbed
step	stepped
laugh	laughed

Analogy

buy	→	buyed
hit	→	hitted
bring	→	bringed

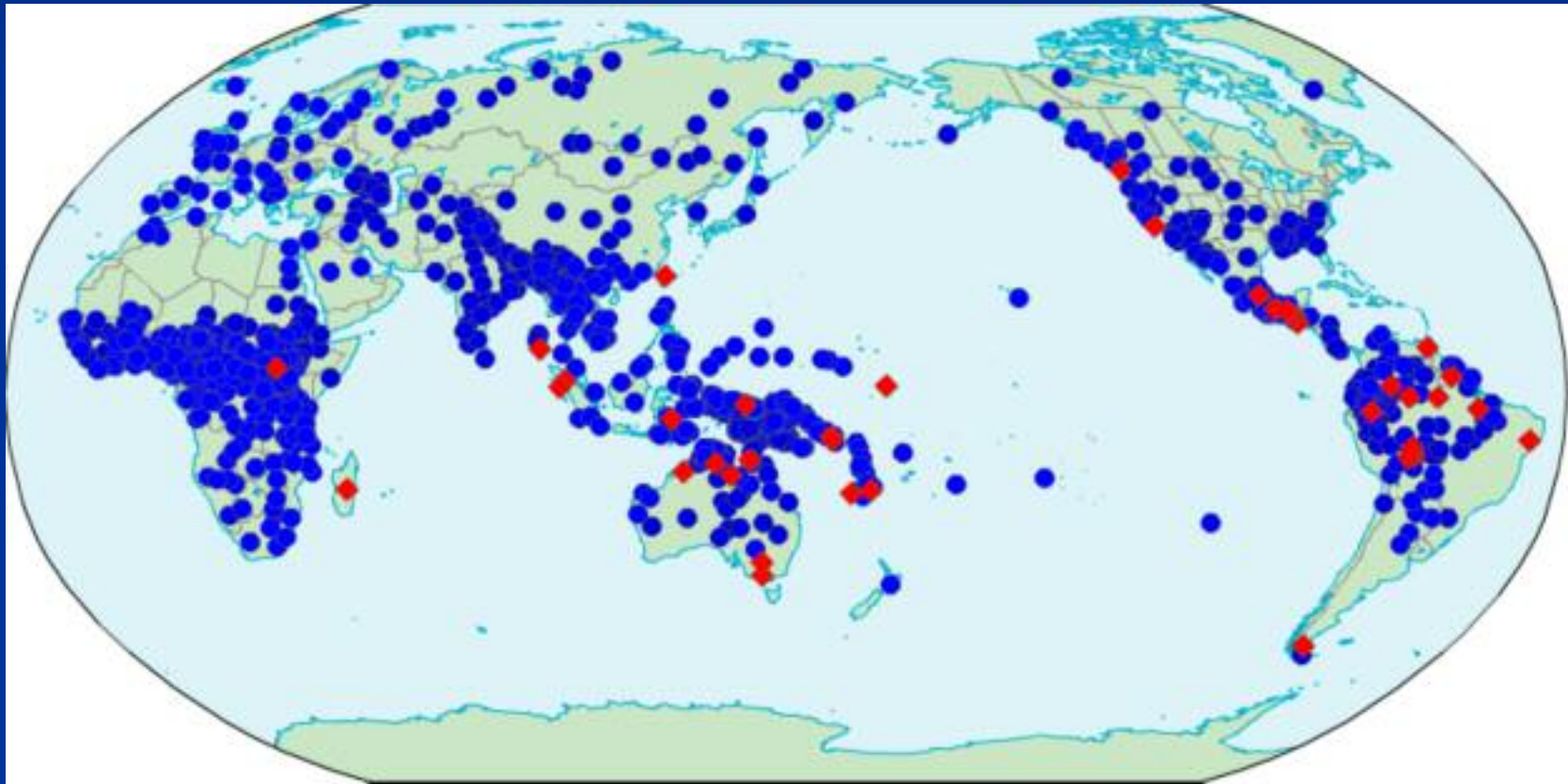
foot	→	foots (feets)
child(ren)	→	childrens

Analogy

Why are there are still irregular verbs in English?

→ Competition between entrenchment and analogy

Information processing



Parsing

Minimize domains (Hawkins 2004):

- The human parser prefers linguistic structures with a short Dependency Domain.
- The dependency domain is defined as the string of linguistic elements that must be parsed (and kept in working memory) in order to access the Mother Node Constructing Categories (MNCCs) of a phrase once the first MNCC has been accessed.
- A MNCC is a linguistic element that allows the parser to uniquely identify the mother node of a phrase.

Minimize domains (Hawkins)

- (1) Mary [gave]_{VP} [the book that she had been searching for since Christmas]_{NP} [to Bill]_{PP}.
- (2) Mary [gave]_{VP} [to Bill]_{PP} [the book that she had been searching for since Christmas]_{NP}.