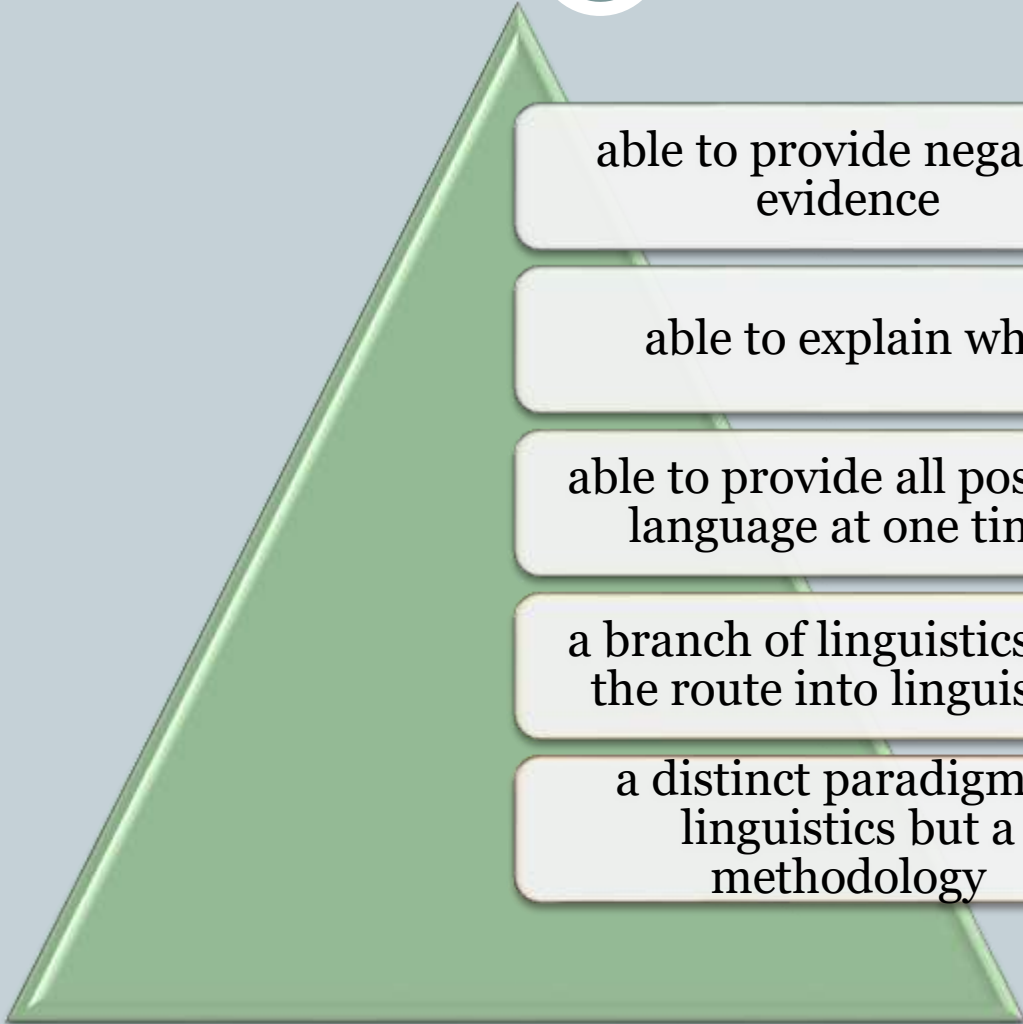


Introduction to Corpus Linguistics



LECTURE 5

Corpus Linguistics is **NOT**



able to provide negative evidence

able to explain why

able to provide all possible language at one time

a branch of linguistics, but the route into linguistics

a distinct paradigm in linguistics but a methodology

Corpus Linguistics is NOT

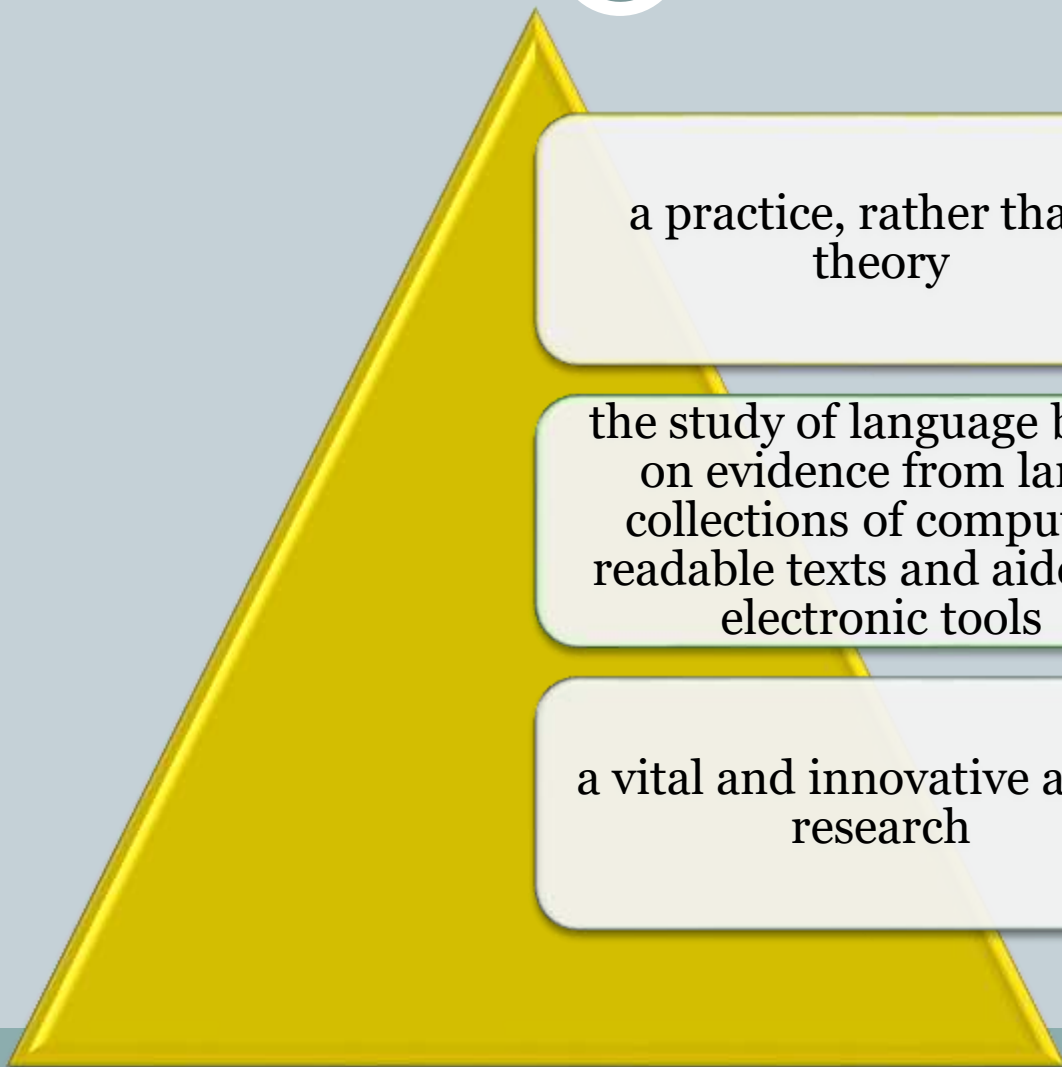


a linguistic theory but rather
a methodology

quite a revolt against an
authoritarian ideology, it is
nonetheless an argument for
greater reliance on evidence

purely observational or
descriptive in its goals, but
also has theoretical
implications

Corpus Linguistics IS



a practice, rather than a theory

the study of language based on evidence from large collections of computer-readable texts and aided by electronic tools

a vital and innovative area of research

Corpus Linguistics IS

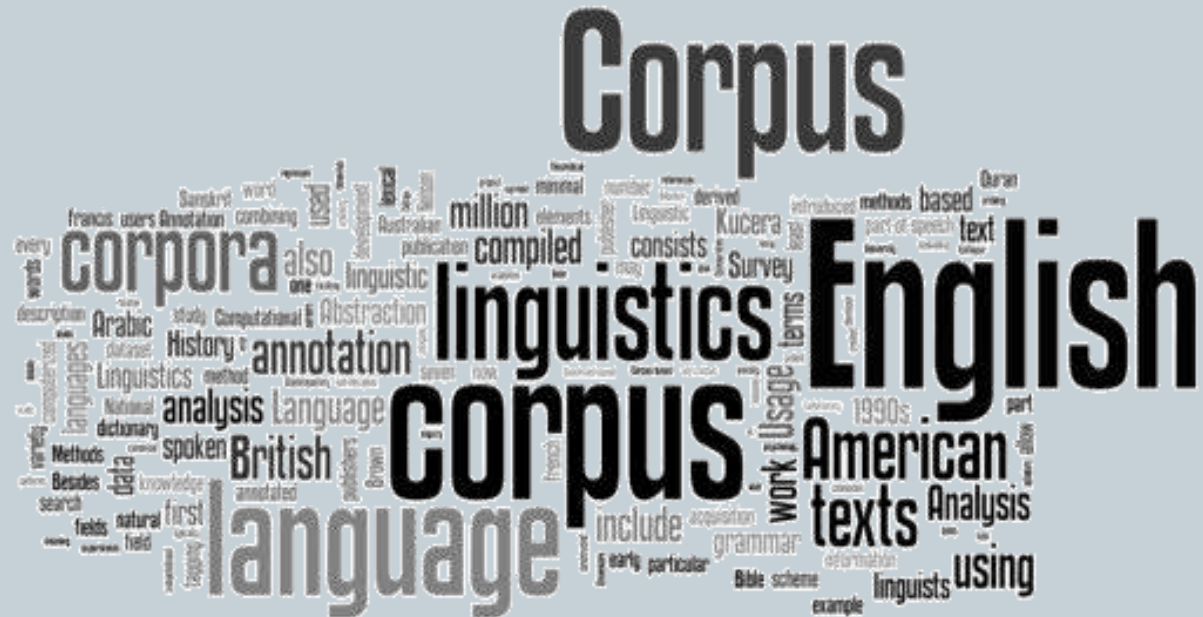


a newly emerging empirical framework that combines a firm commitment to rigorous statistical methods with a linguistically sophisticated perspective on language structure and use

a relatively modern term used to refer to a methodology, which is based on examples of “real life” language use

System and Corpus: Exploring Connections(Equinox, 2005):

- Corpus Linguistics “*is essentially a method for investigating language*”, and is “*almost perversely ‘theory-light’*”.



Corpus Linguistics



- **Corpus linguistics** is the study of language based on large collections of "real life" language use stored in **corpora** (or corpuses) – computerized databases created for linguistic research.
- It is also known as **corpus-based studies** [[Nordquist](#), 2019].



Corpus Linguistics finds out



What are the most frequent words and phrases?

What are the differences between written and spoken language?

What tenses do people use most frequently?

What prepositions follow particular verbs?

How often do people use idiomatic expressions?

Corpus Linguistics Peculiarities

Corpus linguistics is concerned with meaning, with symbolic content. People are not interested in grammatical constructions; they want to know the meaning of what has been said.

Corpus linguistics looks at language from a social, not a psychological perspective. Language is verbal communication between people, is the discourse of what is actually being said (written) and listened to (read).

Corpus linguistics is diachronic. Whatever is said is a reaction to things that have been said before. We can only fully understand utterances if we know what they refer to. The discourse has, of necessity, a diachronic dimension.

Corpus Linguistics Peculiarities



Corpus linguistics uses frequency to arrive at generalisations. The generalisations that corpus linguistics arrives at are not interpreted as laws or rules, but as plausible ways to group similar things together.

Corpus linguistics can also make specific claims concerning unique events of language phenomena by showing in which aspects this event differs from all other occurrences of the same type of phenomenon .

Corpus Linguistics



Chafe

- corpus linguistics as proceeding from understanding language to understanding mind

Tognini-Bonelli

- explores the corpus-driven approach in greater depth, calling it a qualitative revolution

De Beaugrande

- asserts that corpora reveal not the disorder of language use, but its different modes of order

Corpus Linguistics



Sinclair

- focuses on meaning, posits the lexical item as the principal unit of meaning, and asserts that it is monosemous.

Fillmore

- concedes the benefits of corpus data to an armchair (theoretical) linguist.

Aarts

- suggests that corpus linguists should focus more on “the qualitative data that corpora can furnish”

Corpus Approach (Biber, Conrad&Reppen, 1998)



It is empirical, analyzing the actual patterns of language use in natural texts.

It utilizes a large and principled collection of natural texts as the basis for analysis.

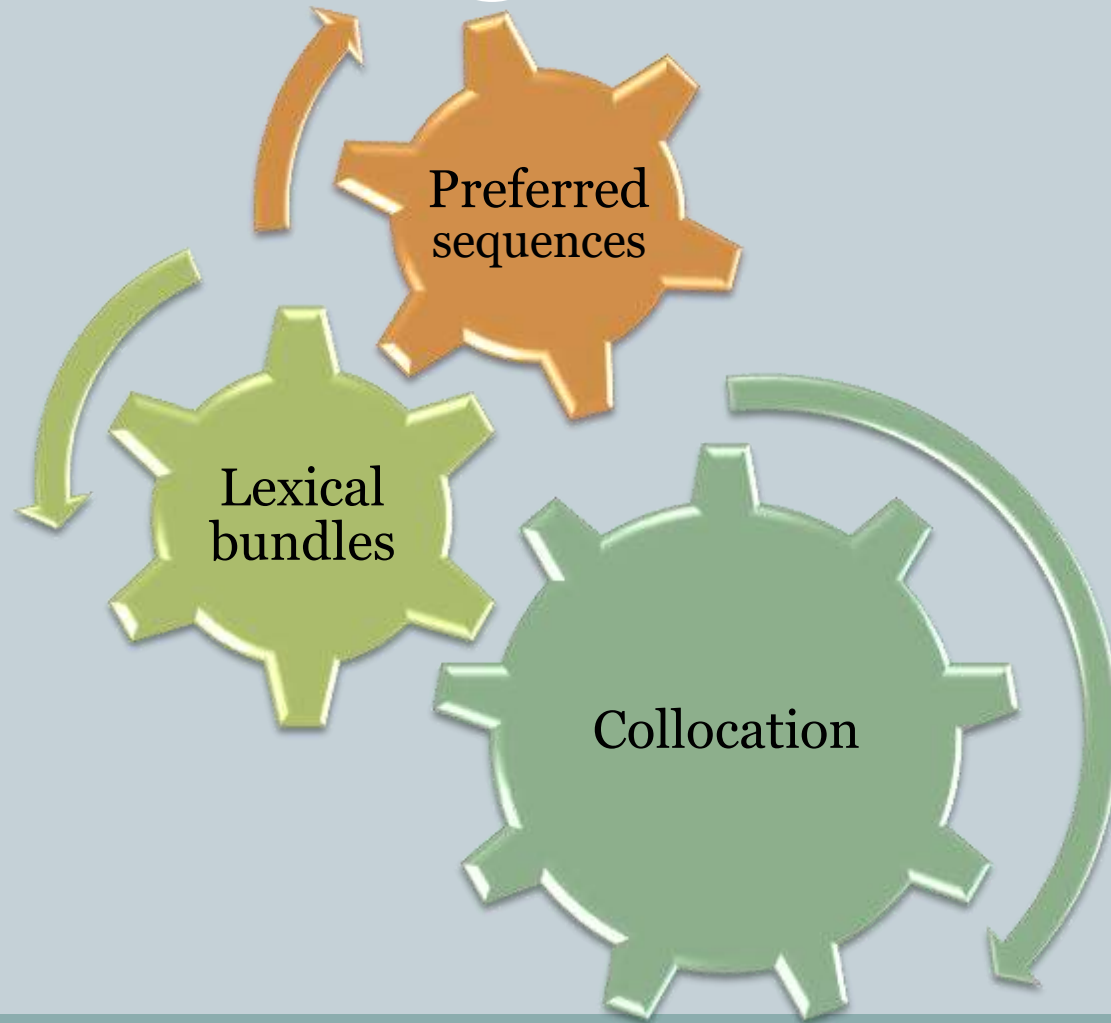
It makes extensive use of computers for analysis.

It depends on both quantitative and qualitative analytical techniques.

Corpus Linguistics Targets



Phraseology



Lexico grammar



- **Sinclair (1991):** “... there is no difference between lexis and grammar, or that lexis and grammar are so closely intertwined that they cannot be productively studied separately. Certain lexical items fall in certain patterns and certain patterns contain certain lexical items ...”



Register



- Register is viewed as a situation of use.



English for Specific Purposes (ESP)



- The Academic Word List (AWL):
<https://www.wgtn.ac.nz/lals/resources/academicwordlist>
- The Business Letters Corpus (BLC):
<http://www.someya-net.com/concordancer/>

Corpus



- A **corpus** is viewed as a principled collection of authentic texts stored electronically that can be used to discover information about language that may not have been noticed through intuition alone.



Types of Corpora



generalized corpora

specialized corpora

learner corpora

pedagogic corpora

References



- An Introduction to Applied Linguistics :
<https://www.press.umich.edu/pdf/9780472033850-part1.pdf>
- *Introduction to Corpus Linguistics: Critical Concepts in Linguistics (6 volumes)*. Available from:
https://www.researchgate.net/publication/282649636_Introduction_to_Corpus_Linguistics_Critical_Concepts_in_Linguistics_6_volumes