Applied Linguistics Chapter Four: Corpus Linguistics

Corpus Linguistics

Exploring actual patterns of language use and use of this exploration in developing material for language classroom instrucion

CL / It provides a powerful tool for natural language analysis and insights on how language use varies in different siruations: spoken/ written, formal / casual

Characteristics of corpus based- analysis of Language

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

Using large and principled collection of natural texts known as corpus/ the basis for analysis

Extensive use of computers for analysis, using both interactive and automatic techniques

Dependent of both quantitavie and qualitattive techniques

- Although computer provides a wide range of sophisticated statistical techniques and acomplish mechanical tasks fastly and accurately , human analysis remains indispensable to decide upon which information to extract from the corpus and giving appropriate interpretation.

Hence, the greatest contribution of CL lies in bringing together aspects of quantitaive and qualitative techniques

1- the quantitative analyses provides an accurate insight on macro-level characteristics.

2- qualitative analyses provide the complementary micro-level perspectives .

Corpus Design and Compilation

-No minimum size for for a text collection to be considered as a collection.
- yet, the larger the corpus is the more valuable it is.
-

Therefore,   it is of great importance to know how corpora are designed and compiled to examine the existing corpora and to understand   what sorts of analyses they are best suitable for.

Types of Corpora

General Corpora:

- aims to present Language in its broadest sense.
- Includes texts that are from different types.
- May include both spoken an written language.
- **Specialized Corpora:**
- Designed with more specific research goals.
- Includes also both and written language.
- It may include historical texts corpora, fiction texts corpora , newspaper writing corpora…
- Learners' corpus: a corpus of spoken and written language samples of non-native speakers ; the most of all is the International Corpus of Learner English- ICLE

Issues in Corpus Design

- **Reliability** of the results: the composition of the corpus should reflect the anticipated research goals. (an intended corpus for exploring lexical questions needs to be very large to allow for accurate presentation of a large number of words of different senses, meanings

Corpus Compilation

- When creating a corpus, data collection involves obtaining or creating electronic versions of the target texts– storing – organizing them.

- Data collection of written corpus means using scanner and optical character recognition software to scan documents into electronic text files.

- Materials for written texts are mostly keyboarded manually.

- Data collection of building a spoken corpus is lengthy and costy.

- Deciding of a transcription system (most spoken corporas use an orthographic transcription system that does not capture prosodic details).

- Choosing the transcription system is deciding how the interactional characteristics of the speech will be represented in the transcrips.

What can a corpus tell us?

- Word counts and basic corpus tools: ranged from simple word lists – catalogues- complex grammatical structures- interactive analysis- linguistic and non-linguistic association patterns- individual linguistic features - identification of features that characterize particular registers- frequency of occurrence information.

Concordancing backages provides additional information about lexical co-occurence patterns. Therefore , the search (word/ phrase is selected , the programme can provide a concordancing listing showing the occurrence of a target word in context. (this display is known as 'Kew word in context')

Working with tagged texts:

carrying more sophisticated types of corpus analyses, it is often necessary to have a tagged corpus. When a corpus is tagged each word in the corpus is given a grammatical label and sometimes it is a complex process:

Can : Model, Verb, Noun

Overview of different types of corpus studies

- Corpora has addressed a number of issues:

    - Language change intrigues researchers, teachers and language learners. Historical changes led to specialized corpora to gain insights into related language development.

    - Exploring differences and similarities across different national and regional varieties of a language.

    - Exploring differences between spoken and written language.

    - Describing sub-registers that provide valuable resources for both teachers and learners.

Corpora and language teaching

- Deciding on which language features and structures are important and how various features and structures are used.

Bringing corpora into language classrrom

- Teachers can shape instruction based on corpus based information. (consulting corpus based studies to gain information about the features they intend to teach ; directing their efforts to the instructions to what grammatical structures the learners are supposed to encounter according to language functions.