

## The corpus method in language teaching

Traditionally, language teaching focused on vocabulary and grammar as two separate components of linguistic skill. It was believed that once learners acquire lexical items and internalise grammatical rules, they will be able to combine these components and apply them in communicative situations. However, there is a growing body of evidence that shows that in order to communicate successfully, learners also need to acquire expressions and structures that lie between lexis and grammar, so called lexico-grammar (collocations are one example of a lexico-grammatical feature).

In addition, there is another crucial area of linguistic development that deserves pedagogical attention. This area is language use in context or pragmatics. Pragmatics focuses on the importance of being able to communicate in real-life situations and adjust linguistic means (expressions) according to the shared expectations between interlocutors. For example, very different language is appropriate when talking to friends than when giving an academic presentation at university.

The crucial term in pragmatics is not correctness but appropriateness. While from the grammatical perspective a linguistic structure can be used either correctly or incorrectly, pragmatics focuses on whether the structure is used more or less appropriately in different situations. For example, while the expression '*this is rubbish*' might clearly signal disagreement, it is not appropriate for a constructive debate in the classroom. Yet, the same expression might be entirely appropriate in a very informal style of conversation.

The corpus method encourages students and teaches to notice lexico-grammatical patterns in multiple examples of real language data and to discuss the appropriateness of the use of linguistic expressions in different situations. Through this, it raises the awareness of the importance of contextualised learning and promotes independent learning from exposure to the target language.

### What are corpora?

Corpora are large collections of texts or transcribed speech searchable by a computer. Corpora enable us to see important patterns in language production. For example, corpora can show typical combinations of words (collocations), meaning differences between near synonyms, lexico-grammatical structures as well as pragmatic expressions and their use.

The following materials are based on the Trinity Lancaster Corpus, a three-million-word sample of spoken learner production based on the Trinity College London examination of spoken English. The corpus includes over 1,500 speakers from nine different linguistic backgrounds and covers B1 to C2 proficiency levels of the Common European Framework of Reference (CEFR).

### What is the corpus method?

The corpus method draws on the evidence from language corpora. It focuses on inductive, data-driven learning, which emphasises the importance of the exposure to multiple examples of the same linguistic structure and discovering patterns by learners. Corpus-based exercises provide examples of authentic language use and stimulate discussion about usefulness and appropriateness of different expressions in a variety of contexts.

The key features of the corpus method are:

- ▶ exposure to multiple examples of authentic language use
- ▶ attention to lexico-grammatical structures
- ▶ information about which phrases and lexico-grammatical structures are frequent
- ▶ contextualised learning
- ▶ raising awareness of the appropriateness aspect of language use (pragmatics)
- ▶ showing typical challenges language learners face at various stages of their development
- ▶ empowering learners to become successful communicators.