

# Use Of Spoken Corpora in Language Learning and Teaching

**Aytkul Nuralievna Djalilova**

Assistant teacher of CSPU

[aytkul.ashirbaeva@mail.ru](mailto:aytkul.ashirbaeva@mail.ru)

**Abstract:** This article discusses the use of spoken corpora to enhance language learners' listening comprehension skills. Spoken corpora provide learners with exposure to authentic spoken language, allowing them to familiarize themselves with natural speech patterns, intonation, and colloquial expressions. Learners can engage in various activities such as transcription, note-taking, and comparative listening to develop their listening accuracy and ability to recognize spoken words and structures. Online platforms and resources like TED Talks, YouTube channels, and linguistic archives such as the International Dialects of English Archive offer valuable spoken corpora for language learners. These resources provide opportunities for listening practice, interactive activities, and targeted learning of specific linguistic features.

**Key words:** spoken corpora, language learners, listening comprehension skills, authentic spoken language, natural speech patterns, intonation, colloquial expressions, transcription, note-taking, online platforms.

## Introduction

In the context of learning and teaching language, the term "corpora" refers to large collections of authentic texts or spoken language samples that are systematically gathered and stored for analysis and research purposes. These collections often consist of written texts, such as books, articles, and websites, as well as transcribed spoken language, such as conversations and speeches.

Corpora are used in language learning and teaching to provide learners and teachers with real-world examples of how language is used in different contexts and by different speakers. By analyzing corpora, researchers and educators can identify patterns, study linguistic features, and gain insights into the structure, usage, and variation of a language.

Corpora can be explored through corpus linguistics, a branch of linguistics that focuses on the analysis of language data contained within corpora. Corpus linguistics allows researchers and educators to investigate various aspects of language, including vocabulary, grammar, collocations, phraseology, discourse patterns, and more. By studying corpora, language learners and teachers can enhance their understanding of the target language and improve their language skills in a more authentic and contextually appropriate manner.

There are a few examples of how corpora are used in language learning and teaching:

- **Vocabulary and Collocation Analysis:** Corpora can be used to analyze the frequency and usage patterns of words and phrases in different contexts. This information helps learners understand which words commonly occur together (collocations) and how they are used in real-world situations. By studying collocations from corpora, learners can improve their vocabulary and develop more natural and accurate language use.
- **Grammar and Sentence Patterns:** Corpora provide a rich source of examples for studying grammar and sentence patterns. Researchers and educators can analyze the frequency and distribution of grammatical structures in corpora, which helps learners identify common patterns and usage rules. This information can be used to teach grammar in a more contextually relevant and meaningful way.
- **Discourse Analysis:** Corpora allow researchers and educators to study how language is used in different types of discourse, such as conversations, academic writing, or business communication. By analyzing corpora, learners can gain insights into the organization of texts, cohesive devices, and strategies for effective communication in specific contexts.
- **Error Analysis:** Corpora can be used to identify common errors or problem areas in language learning. By examining learner language in comparison to native speaker language in corpora, educators can pinpoint recurring errors and design targeted instruction to address those specific areas.
- **Pragmatics and Sociolinguistics:** Corpora provide insights into the pragmatic and sociolinguistic aspects of language use, such as politeness strategies, speech acts, or language variation. Learners can

study corpora to understand how language is influenced by social factors and cultural norms, allowing them to develop appropriate and effective communication skills in different contexts

## Methods And Research

Corpora in language learning can be divided into different types based on their characteristics and purposes. Here are some common types of corpora used in language learning:

1. General Corpora: These corpora consist of a wide range of texts from various genres and sources. They represent the overall usage of a language and provide a broad view of its vocabulary, grammar, and usage patterns. General corpora are often used for studying language in a general context and for developing language resources, such as dictionaries and grammar guides.
2. Specialized Corpora: Specialized corpora focus on specific domains or subject areas, such as medicine, law, business, or technology. These corpora contain texts that are relevant to those particular fields, allowing learners to study and understand the language used in specialized contexts. Specialized corpora are valuable for learners who need to acquire language skills for specific professional or academic purposes.
3. Learner Corpora: Learner corpora consist of language samples produced by second language learners at various proficiency levels. These corpora are used to analyze and understand learner language, including errors, patterns of development, and areas of difficulty. Learner corpora provide insights into the language learning process and help inform language teaching methodologies and materials.
4. Parallel Corpora: Parallel corpora contain texts in multiple languages that are aligned at the sentence or phrase level. These corpora are used for translation and contrastive analysis between languages. Learners can use parallel corpora to study the similarities and differences between their native language and the target language, aiding in vocabulary acquisition, sentence structure, and translation skills.
5. Spoken Corpora: Spoken corpora consist of transcriptions or recordings of spoken language, such as conversations, interviews, or speeches. These corpora provide insights into natural spoken language, including features like intonation, hesitation, and discourse markers. Spoken corpora help learners develop listening and speaking skills, as well as an understanding of conversational patterns and pragmatic aspects of language use.

There are several popular resources and tools available for accessing and analyzing corpora in language learning:

- ✓ Corpus Query Tools: Tools like AntConc, Sketch Engine, and Wordsmith Tools allow users to search and analyze corpora based on specific linguistic criteria. These tools provide features for concordancing, collocation analysis, frequency analysis, and more.
- ✓ Online Corpora: Many online platforms provide access to large corpora for language learning and research. The Corpus of Contemporary American English (COCA), British National Corpus (BNC), Corpus of Historical American English (COHA), and Corpus del Español are some widely used online corpora.
- ✓ Corpus Annotation Tools: Annotation tools like ELAN (EUDICO Linguistic Annotator) and CLAN (Computerized Language Analysis) allow users to annotate and transcribe spoken language data. These tools are useful for analyzing discourse patterns, conversation analysis, and sociolinguistic research.
- ✓ Corpus Linguistics Software: Software such as the Natural Language Toolkit (NLTK), the Stanford NLP (Natural Language Processing) Suite, and the Linguistic Inquiry and Word Count (LIWC) offer a range of functionalities for corpus analysis, including part-of-speech tagging, sentiment analysis, and syntactic parsing.
- ✓ Online Language Resources: Websites like Sketch Engine, Corpus Concordance English (CCE), and BYU Corpora provide access to pre-compiled corpora, concordance tools, and other linguistic resources for language learning and research.
- ✓ Learner Corpora: Corpora of learner language, such as the International Corpus of Learner English (ICLE) and the Cambridge Learner Corpus, are available for researchers and educators to study and analyze second language acquisition patterns and learner errors

## Discussion

Spoken corpora are collections of language samples that capture spoken language in various contexts. These corpora are typically created by recording and transcribing conversations, interviews, lectures, broadcasts, telephone conversations, and other forms of spoken communication. Spoken corpora provide researchers, educators, and language learners with valuable insights into the characteristics of oral language use.

Here are some key aspects and features of spoken corpora:

**Authenticity:** Spoken corpora aim to capture language as it is naturally used in real-life situations. They often include features such as hesitations, false starts, repetitions, and non-standard grammatical constructions that are common in spontaneous speech.

**Multimodality:** Spoken corpora may include not only transcriptions but also audio or video recordings of the spoken interactions. This multimodal aspect allows learners to study not only the linguistic features of spoken language but also non-verbal cues, intonation, and pragmatics.

**Sociolinguistic Variation:** Spoken corpora provide insights into sociolinguistic variation, including regional accents, dialects, social registers, and speech styles. Learners can explore how language use varies based on factors such as age, gender, social background, and context.

**Discourse Analysis:** Spoken corpora are valuable for studying discourse patterns and conversational dynamics. They help learners understand turn-taking, topic management, repair strategies, politeness, and other aspects of spoken interaction.

**Pronunciation and Intonation:** Spoken corpora allow learners to study authentic pronunciation and intonation patterns. By listening to recordings and analyzing the transcriptions, learners can improve their understanding and production of spoken language features.

**Language Teaching and Assessment:** Spoken corpora are used in language teaching and assessment to develop materials, activities, and tests that reflect authentic spoken language use. Learners can use spoken corpora to practice listening comprehension, develop speaking skills, and familiarize themselves with natural language patterns.

There are several online platforms and resources where you can access spoken corpora for language learning. Here are a few examples:

1. **International Dialects of English Archive (IDEA):** IDEA is an online resource that provides audio recordings of native speakers from various English-speaking regions around the world. It offers a collection of accents and dialects for learners to listen to and study.
2. **The Speech Accent Archive:** The Speech Accent Archive is a database of spoken English samples from non-native speakers. It features recordings of individuals from different countries speaking English, allowing learners to compare and analyze various accents and pronunciation patterns.
3. **British Library Sounds:** The British Library Sounds archive provides access to a wide range of audio recordings, including spoken language samples, interviews, oral history collections, and more. It offers a diverse collection of recordings that learners can explore to improve their listening comprehension.
4. **TED Talks:** TED Talks ([www.ted.com](http://www.ted.com)) is a popular platform that hosts a vast collection of video presentations on various topics. It offers a great opportunity for language learners to practice listening skills and enhance comprehension through engaging and informative speeches.
5. **YouTube:** YouTube ([www.youtube.com](http://www.youtube.com)) is a vast video-sharing platform that hosts a wide range of content, including interviews, lectures, panel discussions, and more. Language learners can find channels or playlists that focus on spoken language resources, such as TEDx Talks, language learning channels, or podcasts, to practice listening comprehension.
6. **National Public Radio (NPR):** NPR ([www.npr.org](http://www.npr.org)) is a U.S.-based public radio network that offers a wealth of audio content, including news reports, interviews, podcasts, and storytelling programs. Learners can access NPR's website or use their mobile app to listen to spoken language samples and news reports.
7. **OpenL2Corpora:** OpenL2Corpora ([www.openl2corpora.org](http://www.openl2corpora.org)) is an open-source platform that provides access to various multilingual corpora, including spoken language data. It offers a range of linguistic resources and tools for learners and researchers to study and analyze spoken language samples.

When using these platforms and resources, it's important to consider copyright restrictions and terms of use. Some resources may require free registration or subscription for full access to their content.

Additionally, academic institutions and libraries often provide access to spoken corpora through their online databases or subscriptions, so it may be worth checking with your institution's library as well

## Conclusion

In conclusion, leveraging spoken corpora for language learning can significantly enhance learners' listening comprehension skills. By exposing learners to authentic spoken language, spoken corpora enable them to develop an understanding of natural speech patterns, intonation, and colloquial expressions. Activities such as transcription, note-taking, and comparative listening further enhance learners' listening accuracy and ability to recognize spoken words and structures. Online platforms and resources, including TED Talks, YouTube channels, and linguistic archives, offer valuable spoken corpora for language learners to practice and engage with. Incorporating spoken corpora into language learning activities provides learners with a valuable tool to improve their overall listening comprehension abilities

## List of used literature

1. McEnery, T., & Xiao, R. (Eds.). (2010). *Corpus-based language studies: An advanced resource book*. Routledge.
2. Thompson, P. (2013). *Introducing corpus linguistics*. Routledge.
3. Flowerdew, L., & Tauroza, S. (Eds.). (2015). *Corpus-based approaches to spoken English: Quantitative studies*. Routledge.
4. O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge University Press.
5. Biber, D., & Conrad, S. (2009). *Register, genre, and style*. Cambridge University Press.
6. Seidlhofer, B. (2003). A concept of international English and related issues: From 'real English' to 'realistic English'. *World Englishes*, 22(2), 235-245.
7. Sánchez-Hernández, P. L. (2019). Using spoken corpora for teaching pronunciation in the EFL classroom. *The EuroCALL Review*, 27(1), 56-66.
8. Li, S. (2012). The effectiveness of corpus-based materials in English teaching. *English Language Teaching*, 5(10), 42-48.
9. Wen, Z. (2018). Using spoken corpora to teach English as a second language in China. *Language Learning in Higher Education*, 8(1), 91-107.