**Types of corpora**

A *general corpus* is one which aims to be representative of a particular language, such as the *British National Corpus* (BNC), the *American National Corpus*, the *Turkish National Corpus* (TNC). It may contain written or spoken texts or may comprise texts from both media. The major purpose is to represent texts from different genres, domains, and types in a balanced manner so that it serves as a reference resource for a wide range of linguistic research. A *specialized* *corpus* is relatively small-sized corpus and involves texts that come from a particular field. The tendency in specialized corpus creation is observed in various contexts such as professional and academic domains, business, law, society and internet. Web has been used as a specialized and dynamic corpus such as *WebCorp* or it is used as a source for building corpora as in *Sketch Engine*. In terms of the primary source of data, the text is either written or computer-mediated or spoken. *Written corpora* are easier to compile because of large archives of texts can be found on the internet. However, texts should be arranged in a machine-readable format while building the corpora. Corpora of computer-meditated texts such as emails or websites are easy to gather via data mining software that can store whole website. *Spoken corpora* are relatively smaller due to the difficulties in collecting and transcribing data. While there is always transcription of speech event without its audio recordings in spoken corpora, speech corpora always involve the recordings of the speech events, but not necessarily a written representation of them. In the *multi-modal corpora****,*** apart from audio recording and/or transcripts, the primary data involves additional information on the communicative event, such as video documenting gestures and mimic. Considering the number of language encoded, a corpus may be *monolingual****,*** *bilingual* or *multilingual corpus*. Multilingual corpora comes in two types: *parallel* and *comparable* *corpora*. Parallel corpora are built on the basis of the original group of texts and their translations. The users of parallel corpora can see translation of sentences side by side since they are sentence-aligned along with the tags that act as indicators of which sentences are translations of each other. The first parallel corpus is the *Canadian Hansard Corpus* consisting of government documents in English and in Canadian French. A *comparable corpus* contains original texts rather than translations, gathered in a number of languages and the texts are similar in content or communicative function. When relation to time is considered, there is *synchronic* and *diachronic* *corpora*. Synchronic corpora are built for the purpose of providing a “snapshot” of language use at a certain period of time. In such corpora, all the texts should be selected from the same time period to account for varieties of the language. The *International Corpus of English* (ICE) is designed for the synchronic analysis of the Englishes spoken in Britain, the USA, Australia, Canada, and other countries where En­glish is the first language. Diachronic or historical corpora cover texts representing language use over a particular period of time so that language change is studied. 1.5- million-word *Helsinki Corpus* is one of the most popular multipurpose historical corpus while 400-million-word *Corpus of Historical American English* is published as a larger diachronic corpus. Regarding persistency, a distinction is made between *static* and *dynamic* or *monitor* *corpora*. The contents and size of the static corpora are fixed once they have been compiled and released. Most linguistic corpora are static in that the user can find the same data in the corpus at different times.A *monitor* corpus is the one that it is constantly growing with the addi­tion of new texts from the same variety at regular intervals. It monitors the current changes in language. The *Bank of English* (BoE) and *Corpus of Contemporary American English* (COCA) are well-known corpora of this type for English. Finally*, learner corpora* include texts produced by the learners of that language. Most learner corpora consist of written essays or letters produced in classroom environment. The *International Corpus of Learner English* and the *Longman Learner Corpus* have been a source of research in teaching contexts.