

Definition of Corpus

“A corpus is a representative collection of a fairly large number of examples (or, texts) that share similar contextual or situational characteristics. These texts are then analyzed collectively in order to understand how language is used in these different contexts.”

(Crawford and Csomay, 2016, 6)

History of Corpus Linguistics

1950s

1960

1970

1980

1990

...

2021



- Jesuit priest Roberto Busa creates an **electronic lemmatised index** of the complete *Index Thomisticus*, beginning in the 1950s and completing it in the late 1970s
- Structuralists and other linguists want **real language data** instead of random or staged or theoretical material for their studies
- First **computer-generated concordances** using punched-card technology for storage; processing of some 60,000 words takes more than 24 hours
- Reel-to-reel **tape recorders** coming on the market; possibility of recording in a prepared setting

History of Corpus Linguistics



- First **electronic corpus of spoken language** assembled at the University of Edinburgh (1963-1965)
- First large-scale computers requiring a huge mainframe **computer** as big as a room

History of Corpus Linguistics



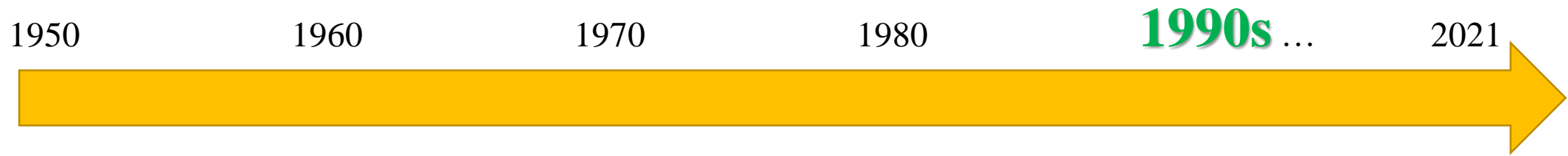
- Library and information scientists develop a keen interest in **Key Word In Context (KWIC) concordances** as a way of replacing catalogue indexing cards and of automating subject analysis
- Computer-generated concordances as a ‘general-purpose working tool **for the study of literature**’ (H. Hill); several concordances of works by well-known authors and philosophers (e.g., T.S.Eliot)

History of Corpus Linguistics



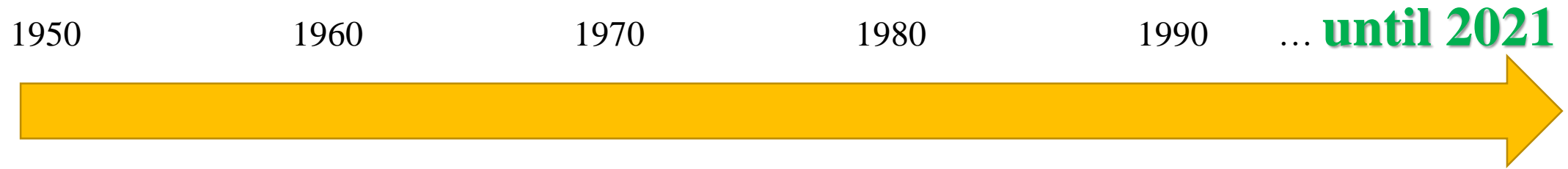
- Analogue portable **cassette recorders** available; possibility of recording spoken interactions in the field; speech events could be transcribed without using shorthand (=Steno)
- Invention of **scanners** for texts, some as big as household chest-freezers

History of Corpus Linguistics



- Miniature **digital recorders** and small but high-powered **video and DVD recorders** available
- **Improvement of scan quality**, fast scanning of texts in editable formats possible
- Late 90s: Massive increase in **desktop computing power**, small teams and individuals are able to take on ambitious corpus projects
- Growth of the **internet** and digitisation; fast download speed enables rapid transfer of information of data

History of Corpus Linguistics



- High-quality audio and video recorders included in each **cell phone**
- very fast and **specialised tools** for corpus analysis freely available on the internet; **accessibility of large digitised corpora** for everybody
- Creation of **multi-modal corpora**: pragmatic elements (e.g., speech, body-language, writing) are taken into account for analysis of corpus material

Three stages or generations of CL (Tognini Bonelli and Sinclair, 2006)

In: Tognini Bonelli, E. (2020). Theoretical overview of the evolution of corpus linguistics. P. 17

- (a) The first twenty years, c. 1960–80; learning how to build and maintain corpora of up to a million words; no material available in electronic form, so everything has to be transliterated on a keyboard.
- (b) The second twenty years, 1980–2000; divisible into two decades:
 - (i) The eighties, the decade of the scanner, where with even the early scanners a target of twenty million words becomes realistic.
 - (ii) The nineties, the First Serendipidity, when text becomes available as the by-product of computer typesetting, allowing another order of magnitude to the target size of corpora.
- (c) The new millennium, and the Second Serendipidity, when text that never had existence as hard copy becomes available in unlimited quantities from the internet.

Group Work: Definitions

Concordancing: listing all words of the corpus in alphabetical order with **information about the context**/sentence in which each word is used

Pragmatics: relationship between natural language and its user: choice of language in **social** settings, how language **affects** others

Lemmatised: reducing the different forms of a word to the **basic form** (e.g., builds, building, built -> lemma: build)

Serendipity: unintentional lucky discovery, finding interesting / valuable things by chance

APA Citation

In-text citation

Direct Citation in the text – APA

Text (Name, Year, p. xx)

"All of that is false" (Lakoff, 2010, p. 72). Therefore, we will explore this topic further.

As Lakoff said, "[a]ll of that is false" (2010, p. 72), and therefore we will explore this topic further.

Find here the official citation guide:

www.mendeley.com/guides/apa-citation-guide

APA Citation

Bibliography

- list full references for all your in-text sources used in paper
- organise it alphabetically by authors' last names
- order multiple works by the same author by date, if the works are from the same year, they are ordered alphabetically by the title and are allocated a letter (a, b, c etc) after the date.

Book in the bibliography

Name, initial(s). (Year). *Title of the Book. Subtitle*. Location of the publisher: Publisher.

Austin, J. L. (1962). *How to do Things With Words. The Williams James Lectures delivered at Harvard University in 1955*. Oxford: Oxford University Press.

Lakoff, G. (2005). *Don't Think of an Elephant! Know Your Values and Frame the Debate*. Carlton North, Vic.: Scribe.

APA Citation

Journal Article in the bibliography

Name, Initial., & Name 2, Initial 2. (Year). Article Title. Subtitle. *Name of Journal*, volume number(issue or season), page numbers. DOI or Retrieved from URL.

Fisher Liu, B.; Iles, I., & Herovic, E. (2019). Leadership under Fire: How Governments Manage Crisis Communication. *Communication Studies* 1(1), 1-20. DOI: 10.1080/10510974.2019.1683593

Harlow, W.F.; Brantley, B.C., & Harlow, R.M. (2011). BP initial image repair strategies after the Horizon spill. *Public Relations Review* 32(1), 80–83. DOI: 10.1016/J.PUBREV.2010.11.005

Olsson, E.-K. (2014). Crisis Communication in Public Organisations: Dimensions of Crisis Communication Revisited. *Journal of Contingencies and Crisis Management* 22(2), 113-125. Retrieved from <https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-5973.12047>.

APA Citation

Attention: every dot / comma / space is there for a reason!

for more authors: comma and then the & ; if there are more than three authors: Name, Initial. et al.

Fisher Liu, B.; Iles, I., & Herovic, E. (2019). Leadership under Fire: How Governments Manage Crisis Communication. *Communication Studies* 1(1), 1-20. DOI: 10.1080/10510974.2019.1683593

title of journal in cursive

volume (=Jahrgang) and in brackets the issue number **without** space

no p. but only page numbers

Olsson, E.-K. (2014). Crisis Communication in Public Organisations: Dimensions of Crisis Communication Revisited. *Journal of Contingencies and Crisis Management* 22(2), 113-125. Retrieved from <https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-5973.12047>.

only initials of first name(s)

if you found it on a website and it is not **downloadable**