Discourse analysis and qualitative analysis of text at scale represent opposite ends of the spectrum of closeness of textual reading, however both offer insights into the way that language shapes the way we view society. Both methods analyse cultural texts to draw out meaning and offer different insight into the ways in which language is used. I argue that when used in tandem these methods are best used in sequence, working from scale toward a more targeted reading of significant texts.

Both methods require a deep understanding of the cultural context of documents. With text at scale a deep understanding of context allows the researcher to use keyword searching or dictionary-based approaches appropriately, or to be able to supervise a machine learning model appropriately. Intertextuality is also an important concept to both methods, unsurprisingly at different scales. Natural language processing looks across a corpus to view patterns and clustering of words in order to ascertain the use of particular words around topics. Intertextuality in this context is quantified and analysed though principal component analysis in order to analyse the overlap between texts. Discourse analysis however looks at intertextuality to understand how discourses draw on existing cultural forms to shape their truth. Conversely, omission of concepts within a document are also consequential. When considering that all discourses are contesting truth, it is important to consider how the counterdiscourse is framed, and what is and isn’t contested. The importance of intertextuality and omission highlight the need for careful reading around the document you are looking to analyse. The researcher's understanding of the broader context and theory is crucial for identifying gaps, inconsistencies, and the contestations present within a text or corpus.

Discourses are the ideas and practices that shape social power and the making of truth within a society. Based firmly in the post-structural tradition, analysis of discourse takes a lot from Michel Foucault’s work. Foucault wrote about discourses in disciplinary structures such as prisons, but the deliberate deployment of social power through language can be seen in many forms, i.e framing of austerity or constructing identities of migrants. The use of language is a social practice through which power operates, the interlocutor, the recipient and the words and form used to express this are all significant within discourse analysis. The strength of computational methods are the scale at which they can be used, this does however limit the closeness of reading that can be had at scale. However, as we know that certain terms and concepts – such as ‘civilization’, ‘Whiteness’, ‘nation’, ‘Disability’ – have implications of power and meaning imbued within them, and as such we can use our wider knowledge of a subject in order to direct our keyword searches or to guide the creation of dictionaries for analysis.

The two methods can be used in conjunction, with computational methods used exploratively to narrow down large corpi of text for identification of significant texts. This can be done with sentiment analysis and named entity recognition.

Keyword searches and dictionary-based approaches allow us to explore large amounts of text relatively simply. Single keyword searches measure the relative frequency of terms’ usage, and the distribution of these key terms. A keyword search in a corpus of newspaper articles could help refine the scope of research to a specific number of articles within the corpus. Dictionary-based approaches are similar but are based on collections of keywords that are associated with particular topics. A specific example of this use is sentiment analysis, where the frequency of words associated with particular sentiments – positive, neutral, negative - are amalgamated to quantify the emotional tone of a token. Sentiment analysis presents a one-dimensional plane of emotion, but its use can be instructive when looking at large amounts of text. This quantification allows for identification of the extremes of emotion at each end. Texts which involve extremes emotive language are more likely to be interwoven with meaning making and other discourses that are ripe for analysis.

The use of named entity recognition can support researchers to find significant texts for subsequent discourse analysis. Due to the structure of language, specific syntactic objects can be drawn out of large bodies of text. Language models, such as spaCy, use dependency parsing (how words relate to each other) in order to analyse specific sections of text. Named Entity Recognition is an example of this, which allows us to parse the names of places, people, products and organisations from large bodies of text, giving the opportunity to investigate the frequency at which particular organisations or sources are mentioned with regard to certain topics. By identifying frequently mentioned organisations who for example put out policy or other text, we can identify significant output that can be viewed more closely with discourse analysis. The methods can also be used in the other direction; Discourse analysis can be used to parse out themes from texts that may appear in clusters of words that are identified through unsupervised machine learning.

While ostensibly these methods, their scale and the closeness of analysis they are capable of can appear to be vastly different, they have a lot in common in their level of analysis and intention. Both are concerned with the use of language, in identifying patterns of meaning making and intertextuality. Both use words as the basis of drawing interpretations from a text but also recognise that words are used deliberately. Both discourse analysis and computational methods for analysing text at scale offer great insight into how language shapes and is shaped by society.