Researchers at the ESRC Centre for Corpus Approaches to Social Science (CASS) at Lancaster University are world-leading experts in studying language. They use what is known as the ‘corpus linguistics approach’ to study texts (corpus is Latin for body). This means they harness the power of computers to analyse huge bodies of texts, allowing studies on a scale and depth that would be impossible if done by humans alone. The corpus linguistics approach can include both quantitative and qualitative methods. Quantitative methods are more objective. They involve collecting objective measurements and numerical data that can be transformed into usable statistics to reveal patterns and trends. For example researchers may want to count the number of times a word appears in newspapers in 2010 and compare it to the number of times it appears in 2017. Qualitative methods are sometimes considered to be more subjective, but allow researchers to look in depth at the underlying causes of an issue, such as considering the contexts that words appear in.

Over the last few years the centre’s work has generated insights into a range of important social issues. One CASS study looked at how Islam and Muslims are written about in the British press, and how this changed between 1998 and 2009. The researchers analysed 200,000 news articles from this period containing the words Muslim(s), Islam, Islamic, Islamist, and Islamist. The articles contained almost 143 million words. They found that the number of articles peaked at the time of the 9/11 and 7/7 suicide attacks. In all of the articles, words relating to terrorism appeared more frequently than words relating to Islam, and about half of the topic words in the newspaper articles referred to the concept of conflict. The Sun was the most likely of all the papers to refer to the concept of evil in its articles about Islam, whilst the Daily Express, more than other papers, focused on stories about immigration and Sharia law. About one in 20 references to Muslims occurred next to a word referring to an extremist belief, such as ‘fanatic’ or ‘militant’, whilst Muslims were described as moderate only about one time in every 200 cases. Although the way language was used to describe Islam and Muslims did not change significantly over the time period, there was a steady increase in press interest in Islam as time went by, with more articles being written.

Another study analysed how British press coverage of immigration changed in the decade leading up to the May 2015 General Election.

The researchers looked for adjectives and verbs associated with the terms ‘immigration’ and ‘migration’, taken from 171,401 newspaper articles published between January 2006 and May 2015. They found that the most common adjectives were ‘mass’, ‘net’, and ‘illegal’. Closer examination of the top 50 most common adjectives revealed that many related to the scale or pace of migration. The study showed that these terms were increasingly used over the time period. They made up nearly 40% of all associated adjectives.
Changing language

in 2006, but over 60% in the five months of 2015. Meanwhile, the use of the words ‘illegal’, ‘legal’, ‘unlawful’, or ‘irregular’ declined from 22% in 2006 to less than 10% in January-May 2015.

The most common verbs associated with ‘immigration’ or ‘migration’, on the other hand were words expressing efforts to limit or control movement. These included ‘control’, ‘tackle’, ‘reduce’, and ‘cap’. The overall frequency of these verbs rose by about five times between 2006 and the high point in 2014. Most of this increase was from 2013 onwards. As a share of all verbs expressing some action towards ‘immigration’, this category was consistently making up 30-40% of verbs from 2010 onwards. This reveals a change in the way newspapers portrayed immigration, with attitudes hardening over time.

A further study looked at how the language used to talk about homosexuality and gay rights has changed over time, reflecting changes in societal attitudes. CASS academics compared the language used in speeches by politicians opposed to the Marriage (Same Sex Couples) Bill in 2013, which sought to allow gay people to marry, with that used by politicians opposed to the Sexual Offences Bill, debated in parliament between 1998 and 2000 and sought to ‘equalise the ages at which people can lawfully consent to homosexual and heterosexual sexual activity’. Both sets of legislation were eventually passed in parliament and gave increasing rights and equality to Lesbian, Gay, Bisexual and Transgender people (LGBT).

The researchers used computer software to analyse the most common language patterns in the speeches. They found that in the 90’s debates on consent, the most common words used by MPS opposed to equal rights were ‘moral’, ‘homosexual’, ‘boys’, ‘health’ and ‘vulnerable’. These words were used to make the argument that homosexuality is immoral, unnatural, unhealthy and boys need protection from predatory gay men. The words most commonly found next to the word homosexual were: ‘adoption’, ‘acts’, ‘heterosexuality’, ‘heterosexuals’, ‘marriages’, ‘lobby’, ‘consenting’, ‘heterosexual’, ‘homosexualism’ and, ‘activity’. The words most commonly found next to the word gay on the other hand were: ‘clubs’, ‘lesbian’, ‘community’, ‘rights’, ‘gay’, ‘men’, ‘against’, ‘young’, ‘people’ and ‘sex’. The language used often focused on negative descriptions of sexual activity and the differences between homosexuality and heterosexuality. Homosexuality was also described as being a behavioural choice almost twice as much as it was a part of someone’s identity.

The debates in 2013 against same-sex marriage, however used very different language. Rather than making arguments personally against gay people, and using less outwardly homophobic language, MPs used much more measured, and less inflammatory words. For example the pronoun ‘I’ was used much less than in the 90’s age of consent debate, where it was used in phrases like ‘I think’, or ‘I feel. This suggests that politicians were less confident in openly expressing anti-equality positions. MPs were much more likely to describe homosexuality as being part of one’s identity, rather than as a chosen behaviour. Arguments shifted away from criticising gay people, and instead they attempt to portray LGBT people in a positive light. The changes reflect a seismic shift in attitudes towards LGBT people since the late 1990’s. By the time of the recent same-sex marriage debates, explicit homophobia had become unacceptable.

It’s not just modern societal issues that can be uncovered by studying language. A CASS study
on the portrayal of beggars and poor people in 17th century England found that, whilst in medieval times being poor was not seen as shameful, as the 17th century progressed hostile attitudes to poor people became more and more prominent, with beggars and vagrants described in increasingly negative terms. These included words like 'roguery', 'vagrancy' and vagabondage. Beggars were described as being 'bold', 'proud' and 'foolish', 'vile' and 'idle'. They were also tarnished with connotations of sexual immorality and vice.

Corpus Linguistics allows large amounts of text to be analysed very quickly, revealing patterns that humans would otherwise find difficult to spot. This approach is still relatively new and there is the potential for many fascinating discoveries about language and society to be made.

Further information
CASS ran a free Corpus Linguistics MOOC, which started in September 2017 - https://www.futurelearn.com/courses/corpus-linguistics