

Weighing ethical considerations in corpus research based on online data sources

This paper discusses ethical concerns related to compilation of a so-called 'big-data' corpus (roughly 47.5 million words) harvested from a publically archived online discussion forum under no password protection, where users use pseudonyms to post about their experiences related to a specific (named) type of trauma. Such corpus data would provide unique empirical data illustrating how these trauma survivors use metaphor about various aspects of their experience (sense of self, sense of other(s), road to recovery, etc.), and in this way shed light upon their underlying conceptualizations of both trauma and post-traumatic growth. This research would contribute towards increased scientific knowledge about how metaphor functions as a cognitive tool when making sense of a tumultuous situation, as well as towards raising general awareness about the effects of the traumatic experience in focus. The research methodology would parallel that of the *Metaphor in end-of-life care* project (<http://ucrel.lancs.ac.uk/melc/>), which used semantic tagging to identify metaphors in online forum discourse of terminal patients and their caretakers.

The field of internet research ethics is young, having developed during the last 15-20 years. Results include numerous internet ethical guidelines, by groups ranging from scholarly associations to national research ethics committees. Although there is no single standard, guidelines concerning ethical viability of internet research generally address the following: intrusiveness, public/private, vulnerability, potential harm, confidentiality, intellectual property rights, and informed consent. Such considerations must be balanced against other important ethical values such as scientific and societal value, i.e. the 'greater good'. There is no one-size-fits-all solution. This paper weighs the 'for' and 'against' cases for allowing the proposed research to proceed, viewed in the context of stringent Norwegian ethical internet research guidelines and law – thus illustrating the sometimes thorny ethical issues that corpus linguists must consider when using online sources as data.