The Hong Kong Polytechnic University 香港理工大學

Department of English and Communication 英文及傳意學系

IRCAHC

Corpus Approaches to Digital Health

A Healthcare Symposium

3 June 2026 (Wed)
09:30 - 17:00 (HK Time)
Hybrid mode

Venue: HHB110, 1/F, PolyU Hung Hom Bay Campus or Online via Zoom

Whatever our views on concepts such as GenAI, we now find ourselves inhabiting a digital world, in which health is increasingly represented in ways that have yet to be completely understood. The talks in this year's symposium interrogate the nature of healthcare from perspectives based in corpus-assisted corpus linguistics, digital health, and the role of GenAI in health communication. Our expert speakers and rapporteur will bring their considerable experience in these fields to talks, discussions, and a workshop.
This symposium will provide an overview and a way forward in addressing these areas.

Programme

09:30-09:45
Introduction by Stefano Occhipinti
The Hong Kong Polytechnic University

09:45-10:30
That's your anxiety talking': sketching anxiety in a corpus of forum posts.
Paul Baker
Lancaster University, UK

10:30-11:15 followed by Tea Break
Using Corpora to Analyse Digital Health Discourse
Gavin Brookes
Lancaster University, UK

11:35-12:20
GenAI in Health Communication Research: Method, Data, Discourse
Niall Curry
Manchester Metropolitan University, UK

12:20-12:50 followed by Lunch
Rapporteur for morning talks
Tony McEnery
The Hong Kong Polytechnic University

14:15-15:30: Workshop followed by Tea Break
Using Surprise Words to examine Representations of Obesity in UK News
Paul Baker
Lancaster University, UK

15:45-17:00
Roundtable discussion & closing remarks

REGISTER NOW

https://polyu.hk/DJwAZ

Participants are advised to bring a laptop. Lunch and tea will be provided for all participants. The registration will be on a first-come, first-served basis.

09:45-10:30

'That's your anxiety talking': sketching anxiety in a corpus of forum posts.

Paul Baker, Lancaster University, UK

Anxiety is a growing, worldwide phenomenon with The World Health Organization (2017) estimating that 264 million people live with anxiety, a 14.9% increase since 2005. This talk focusses on a corpus of 23 million words of text posted to the Anxiety Support forum of the social networking service HealthUnlocked between March 2012 and October 2020, comprising 294,082 posts (Collins and Baker 2023).

The ways that we understand and orient towards an emotion (Chen, Chen and Yang 2019) or mental health issue (Chan, Chan and Kwok 2015) can impact on its prevalence or severity. So, the ways that we talk, both to ourselves, and others, about mental health are worthy of investigation, particularly in terms of identifying representations that have the potential to be helpful or unhelpful.

Using the corpus analysis tool Sketch Engine, a technique called Word Sketch, which groups collocates according to grammatical relationships, was used in order to identify a range of oppositional representations around the word anxiety. Across the corpus, a wide range of perspectives of anxiety were identified. For example, some posters viewed it as an acquired medical disorder, to be resolved through therapy or medication. Others saw it as part of themselves, an aspect of their personality, to be accepted and lived with. Some people used metaphors which personified anxiety or conceptualised it in abstract terms, and while some people downplayed its impact, others viewed it through a catastrophising lens, using hyperbolic language to write about it.

The study indicates that people experiencing anxiety do not view the condition through the same lens, but that there is considerable variation, incorporating combinations of different discourses. In terms of applications, this kind of research could enable medical practioners to match their patients' own framings or identify and challenge uses of language which result in representations that may not help a patient to effectively manage their anxiety.

10:30-11:15

Using Corpora to Analyse Digital Health Discourse

Gavin Brookes, Lancaster University, UK

Digital technologies have profoundly shaped the social practices that constitute individuals' experiences and understandings of health and illness within contemporary societies. This includes, for example, the development of life-preserving and life-enhancing technologies and wearable devices and apps that allow individuals to monitor certain health indicators. In this context, discourse continues to play a key role in shaping the very practices that constitute health and illness. Yet at the same time, the discourses that surround health have also been transformed by the affordances of digital (particularly communicative) technologies. Online platforms in particular, while opening up new channels for health-related discourse, have also provided new avenues for linguists and other researchers and practitioners interested in learning more about the discursive dynamics of health and illness. Indeed, platforms such as social media, blogs and online support groups have allowed for better representation of the linguistic routines of non-practitioner/professionals interacting about health, and in turn the diverse perspectives that such language use articulates. In this talk, I will explore the linguistic terrain of what we might broadly term 'digital health discourse'. Taking a corpus linguistic perspective, I will reflect in particular on the various opportunities, but also challenges, that such discourse presents. This includes, inter alia, the methodological challenges inherent in representing within corpus data the mediated nature of such discourse, the ethical challenges associated with gathering sensitive discourse data at scale (as corpus compilers are often wont to do), and the work involved in capturing and analysing health-related discourses whose influence is amplified by their capacity to shift between (online and offline) contexts.

11:35-12:20

GenAI in Health Communication Research: Method, Data, Discourse

Niall Curry, Manchester Metropolitan University, UK

Generative AI (GenAI) has become a matter of critical concern in Applied Linguistics. Amid hype-ridden and exclusionary reactions that shaped initial responses to it, a critical perspective on GenAI has since emerged, situating its affordances and limitations within the various subfields that span Applied Linguistics. Given its diverse data and analytical approaches, and the variety of real-world contexts to which it responds, linguistic research on health communication constitutes one among many important sites of applied linguistic research in which contemporary GenAI and its use remains underexplored. In light of this, in this talk, I argue that GenAI poses not only technical and ethical challenges for health communication research, but also fundamental questions about how language, evidence, and authority are produced and interpreted in the health-related contexts.

To support this argument, I examine GenAI through three interconnected lenses: as method, as data, and as discourse. First, I explore how issues of epistemology, ontology, and ethics can be operationalised within health communication research methodologies, highlighting the affordances and limitations of GenAI in the research process. Second, I consider AI-generated content as a new and consequential form of data, outlining both the risks it introduces and the analytical opportunities it offers for understanding health narratives, misinformation, and public-oriented health communication. Third, I turn to GenAI itself as an object of discourse in health communication, highlighting the role of applied linguists in analysing and unpacking how AI is discursively constructed in public discourse, and here with regard to its role in contemporary healthcare in particular. By bringing these three dimensions together, the talk aims to offer a reflection on GenAI that moves beyond adoption or critique and, instead, that graduates toward a critical, theory-informed, and socially responsible perspective on GenAI in health communication research. In so doing, I aim to clarify what is at stake for health communication research and to identify emerging directions for applied linguistics research at the intersection of AI, language, and health.

14:15-15:30: Workshop

Using Surprise Words to examine Representations of Obesity in UK News

Paul Baker, Lancaster University, UK

In this workshop we will work with a new tool called the S-Words Analysis Tool or SWAT (Baker 2026). The tool introduces a new way of carrying out corpus analysis, based not on comparisons of the frequency of a word but a comparison of its meaning and context across two corpora. I propose the concept of collocationally-divergent words, called surprise words (s-words), defined as words whose collocational profiles differ significantly between two corpora. For example, the collocates of bank might be account, manager, robberyand charge in a corpus containing texts relating to economics but muddy, grassy, river and steep in a corpus about the environment. A related term - core words (c-words) are those which show high collocational similarity across corpora. Divergence scores are calculated by comparing the similarity between the top collocates of each word across two corpora, and all the words in the corpora are then listed in order of their divergence score. S-words can reveal salient differences between two corpora, e.g. in the ways that a topic might be represented, or the ways that argumentation strategies are used.

In this workshop the top s-word and c-words will be identified by comparing two comparable corpora of UK news articles about obesity in The Guardian (a liberal broadsheet) and The Sun (a conservative tabloid). Analysis of s-words will reveal differences in discourse framing, particularly in relation to responsibility, cost and solutions around obesity, while the c-words indicate topics and representations which are consistent to both newspapers.