HAL open science

Honoring our dead: text mining a century of academic obituaries in The Lancet

Dakota Murray, Vincent Larivière, Cassidy R. Sugimoto, Guillaume Cabanac

To cite this version:

Dakota Murray, Vincent Larivière, Cassidy R. Sugimoto, Guillaume Cabanac. Honoring our dead: text mining a century of academic obituaries in The Lancet. IC2S2 2020: 6th International Conference on Computational Social Science, Jul 2020, Cambridge, MA, United States. , 2020. hal-02911615

HAL Id: hal-02911615

https://hal.science/hal-02911615v1

Submitted on 4 Aug 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Honoring our dead: text mining a century of academic obituaries in *The Lancet*

[1]Dakota Murray, [2]Vincent Larivière, [1]Cassidy R Sugimoto, [3]Guillaume Cabanac

[1]Indiana University Bloomington, [2]Université de Montréal, [3]Université de Toulouse,

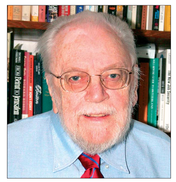
Obituaries as a way to understand science

Obituaries offer a unique window into the values and social dimensions of academic life.

So what do we say when we talk about the dead?

Data is from the full-text of 5,069 obituaries published in *The Lancet* between 1850 and 2019

The Lancet, founded in 1823 in the U.K., is one of the world's oldest and most prestigious medical journals.



Brian MacMahon

Leading epidemiologist. Born on Aug 23, 1923, in Sheffield, UK, he died after a stroke on Dec 5, 2007, in Boston, MA, USA, aged 84 years.

After earning the diplomas of the Royal Colleges of Physicians and Surgeons, in 1948, Brian MacMahon tried his hand as a locum doctor in some of the poorer parts of Birmingham, UK, and as a ship's doctor. At that point, he "became somewhat dispirited about a future in clinical medicine", MacMahon recalled in a 2004 interview. So he enrolled in a public health programme at the University of Birmingham, intending to go into public health. But a trio of epidemiologists—Thomas McKeown, Ronald Lowe, and Reginald Record—soon had him under their spell. They convinced him to study pyloric stenosis, which was the start of a several-decades-long career in pioneering epidemiology.

"I became a card-carrying epidemiologist only when invited in 1958 to be chairman of the department of that name at the Harvard School of Public Health", MacMahon said in a 2004 interview with Walter Willett (*Epidemiology* 2004; **15**: 4). "My research interests remained essentially unchanged, but they were now legally epidemiology. I have been content in that designation." At the time, the department was weighted towards infectious diseases, but MacMahon shifted its focus to chronic diseases.

MacMahon is perhaps best known for his work showing that the risk of breast cancer is substantially affected by the age at which the first full-term pregnancy occurs. "It was exceptionally important, although it may sound simple", said Dimitrios Trichopoulos, who knew MacMahon for 40 years, first as a graduate student and then as the successor to MacMahon's chair at Harvard. The finding was later determined to be due to the fact that a full-term pregnancy, as opposed to an abortion, induces terminal differentiation in the mammary gland.

MacMahon was among the first to try to introduce the study of diets, Trichopoulos said. "In contrast to [Sir] Richard Doll, who was always sceptical that people would find ways to study diet, he was very much pioneering the study of diet and the relation to cancer, although one of his students, Walter Willett, shone more than he did." MacMahon and Trichopoulos were the first to publish a link between lung cancer and passive smoking. The effect was real, but MacMahon made it clear that the effect was 2–3 orders less than that of active smoking.

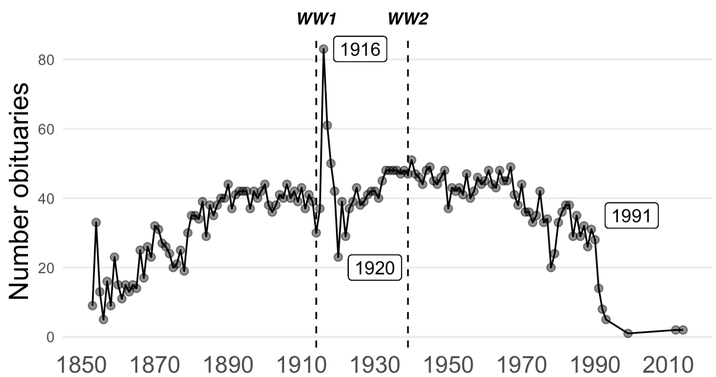
"He got several things right, and some things wrong, like every scientist", Trichopoulos said. In 1981, the two were authors, with colleagues, of a study in *The New England Journal of Medicine* purporting to show a link between coffee drinking and pancreatic cancer. The study was criticised because of its choice of controls, however, and was not confirmed in subsequent work. Asked a few months after publication about the choice of his control group, MacMahon said: "I'm satisfied with it." He acknowledged, however, that the control group had a large number of patients with gastrointestinal disorders. "I think that if that point had occurred to me before publication I would have added information showing that the relationship holds whether you exclude gastrointestinal disorders or not, although it is true that the relative risks are reduced by excluding the patients with gastrointestinal disorders."

"He was a person with outstanding and unusual integrity, and exceptional generosity. He was the first one to be critical of himself. But he was also stubborn when he had his own views", Trichopoulos said. "He had a view that too much statistics took away all the pleasure and the message of epidemiology." MacMahon was a terrific mentor to students who would go on to lead many epidemiology departments in the USA. "Even when I was a student, he would invite me to his house, which was a $40 cab ride, and he would drive me back to my dormitory himself", recalled Trichopoulos. He was also caring—just as an example, when he hit an animal on the road he would take it home to take care of it. "If you visited his house, it was full of animals", Trichopoulos said.

MacMahon remained chair of the Harvard School of Public Health's department of epidemiology until 1988, when he retired. In March, 2006, more than 50 years after he had first published on the subject, MacMahon published a paper on pyloric stenosis (*Epidemiology* 2006; **17**: 195–201). In what Trichopoulos said was a bit of his career coming full-circle, He is survived by two sons, Michael and Kevin; and two daughters, Kathleen and Mary. His wife died in 2001.

Ivan Oransky
ivan-oransky@erols.com

www.thelancet.com Vol 371 January 12, 2008

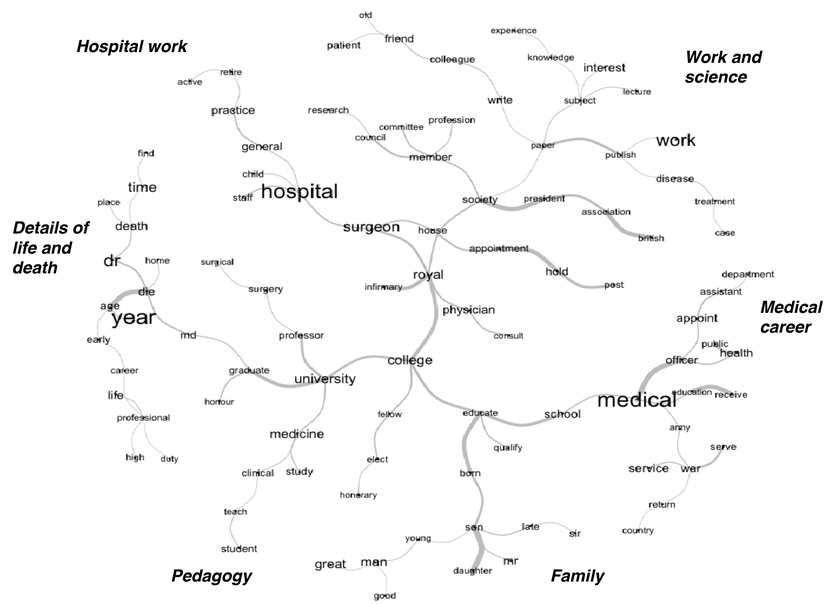


Academic obituaries spiked during World War 1

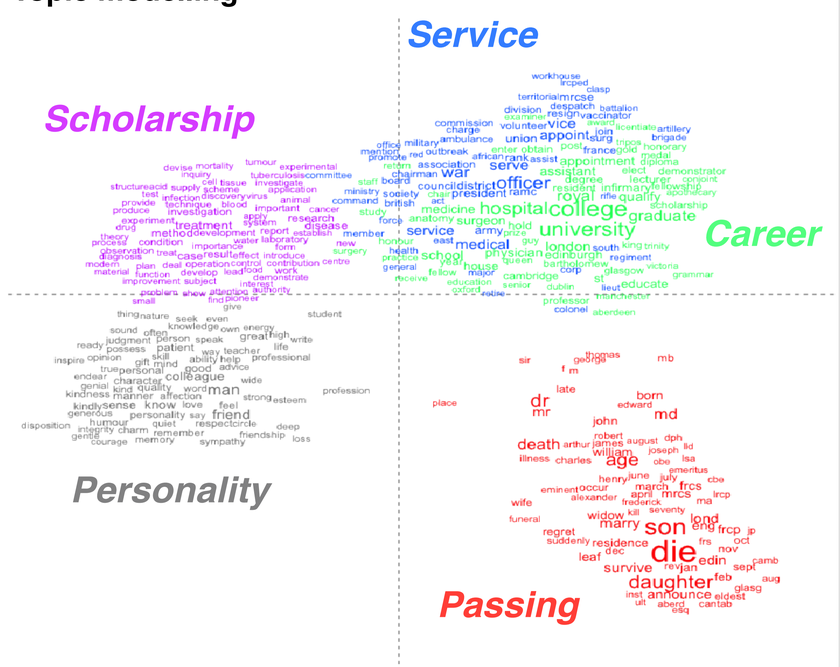
"The death is announced of Captain D. B. MacLean, killed in action at the front while attached to the East Lancashire Regiment …
After being on sick leave till March Ist of this year he was sent to a military hospital, but while on duty there he succumbed to pneumonia."

What is said in obituaries?

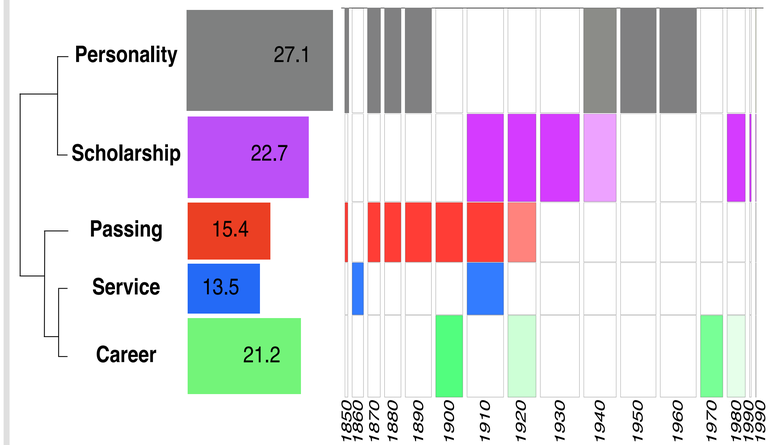
Word co-occurence



Topic modelling



How have obituaries changed over time?



The **relative representation** of topics has **changed over time.**

Details of family and **Passing** were over-represented until 1920.

Service was most represented during World War 1.

After WW 1, **Scholarship** and **Personality** became more common

Recently, **Scholarship** and **Career** are having a resurgence.

In Conclusion, Scientific obituaries reflect the importance of world events, professional career, and personality, but the genre evolves

Future work aims to construct a larger corpus of obituaries from popular and scientific venues

How has the academic obituary evolved, and what differences exist by venue, and the subjects's gender, discipline, and nationality?

dakota.s.murray@gmail.com

@dakotasmurray