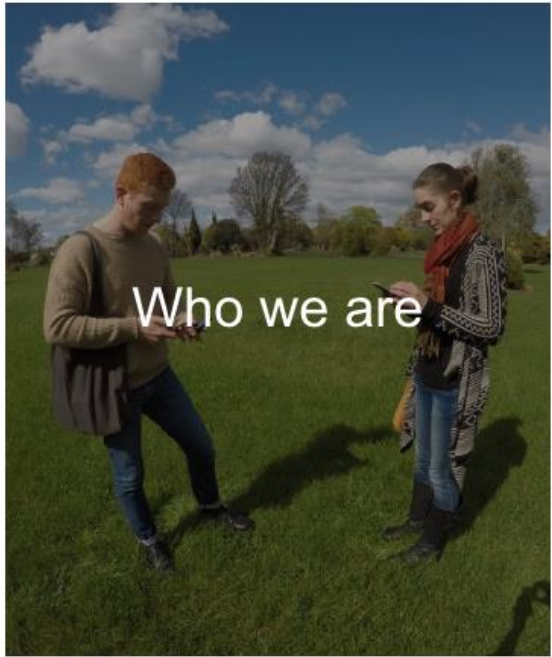


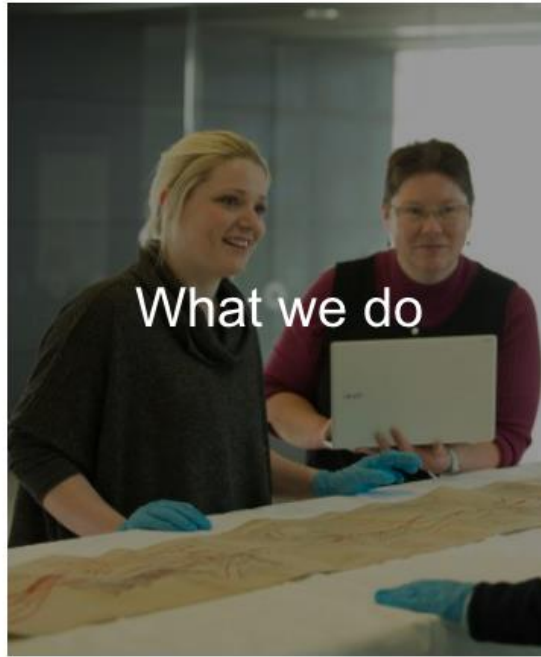


# Introductions

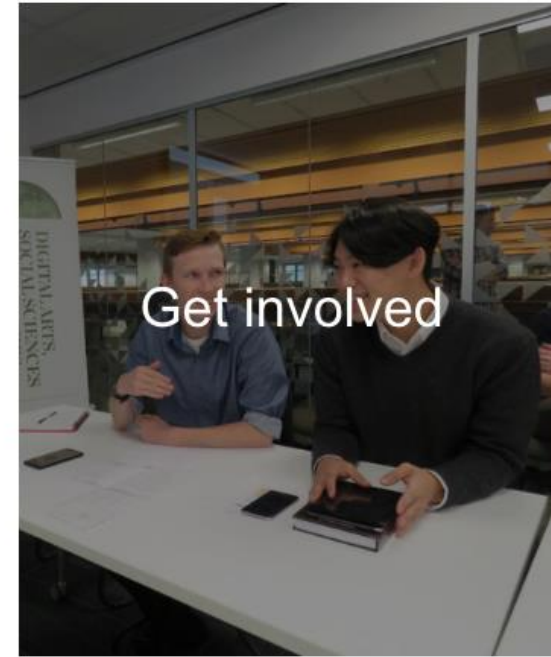
DIGI405: Texts, discourses and data



Who we are



What we do



Get involved

Arts Digital Lab  
University of Canterbury  
Private Bag 4800  
Christchurch 8140  
New Zealand  
email: [kaspar.middendorf@canterbury.ac.nz](mailto:kaspar.middendorf@canterbury.ac.nz)



Arts Digital Lab  
Te Pokapū Aronui ā-Matihiko

Arts Digital Lab: Logie 306

Lab manager: Kaspar Middendorf: [kaspar.middendorf@canterbury.ac.nz](mailto:kaspar.middendorf@canterbury.ac.nz)

Contact for DH research seminars / events.



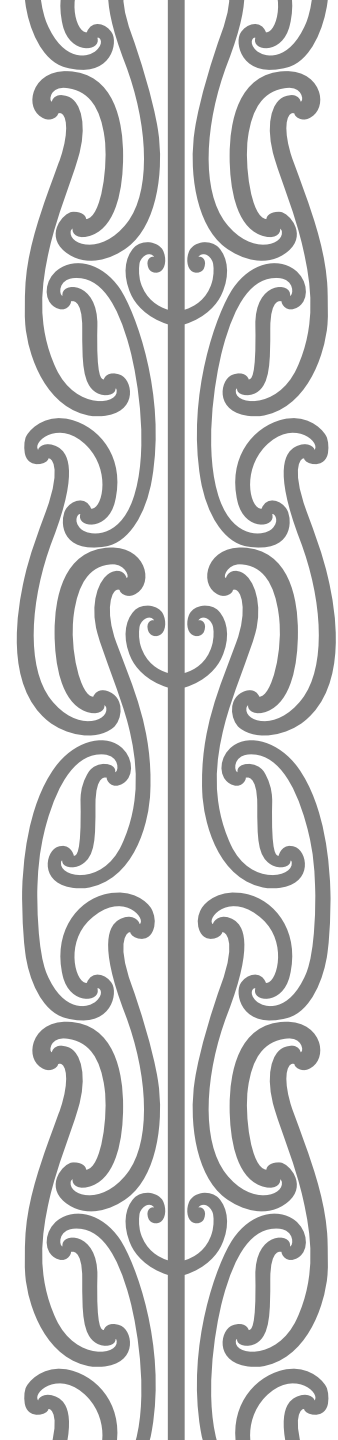
# Outline

- Introduction to the course
- Language (as data): interesting! pervasive! important! challenging!
- Defining key concepts
- Example of text as data to introduce some of the big challenges



# Introduction to the course

- Introducing the teaching staff
- AKO LEARN site
- Core Concepts
- Overview of topics and structure
- Lab Classes
- Assessments
- Week-to-week participation
- This week – Introduce yourself!
- Active Learning
- What about LLMs?
- Class rep
- Questions?!



WIKIPEDIA  
The Free Encyclopedia

[Main page](#)  
[Contents](#)  
[Current events](#)  
[Random article](#)  
[About Wikipedia](#)  
[Contact us](#)  
[Donate](#)

[Contribute](#)

[Help](#)  
[Learn to edit](#)  
[Community portal](#)  
[Recent changes](#)  
[Upload file](#)

[Tools](#)

[What links here](#)  
[Related changes](#)  
[Special pages](#)  
[Permanent link](#)  
[Page information](#)

Not logged in [Talk](#) [Contributions](#) [Create account](#) [Log in](#)

[Article](#)

[Talk](#)

[Read](#)

[Edit](#)

[View history](#)



# Master's degree

From Wikipedia, the free encyclopedia

A **master's degree**<sup>[note 1]</sup> (from [Latin](#) *magister*) is an [academic degree](#) awarded by [universities](#) or [colleges](#) upon completion of a course of study demonstrating mastery or a high-order overview of a specific [field of study](#) or area of [professional practice](#).<sup>[1]</sup> A master's degree normally requires previous study at the [bachelor's](#) level, either as a separate degree or as part of an integrated course. **Within the area studied, master's graduates are expected to possess advanced knowledge of a specialized body of theoretical and applied topics; high order skills in analysis, critical evaluation, or professional application; and the ability to solve complex problems and think rigorously and independently.**

## Contents [hide]

- 1 Historical development
  - 1.1 Medieval era to 18th century
  - 1.2 Nineteenth century
  - 1.3 Twentieth century
  - 1.4 Twenty-first century
- 2 Titles
  - 2.1 Types
- 3 Structure
  - 3.1 Duration

# Active Learning

- Attendance (should weigh very heavily).
- Participate in class discussion on a regular basis.
- Write and accumulate study notes on the readings.
- Volunteer (in advance) to take responsibility (as an individual or as a group) to initiate discussion in class (discussion questions must be prepared in advance in consultation with the lecturer).
- Come to class with written questions on scheduled readings / activities.
- Write a brief review of an [Academic Skills Centre workshop](#) or resource and how it might apply to this course, and share it with the class (in consultation with the lecturer).
- Meet with a study group on a regular basis (once a week, once every two weeks; each student needs to document meeting).
- Visit your lecturer and/or tutors during office hours to discuss issues raised in class and in the readings; bring one or two classmates with you.
- Contribute possible exam questions.
- Watch a documentary film or read a text that explores how race, gender or sexuality is connected to topics studied in this course, then share your reaction and/or critical insights with your classmates and in a written commentary to the lecturer (one paragraph).
- Offer novel learning resources or ideas to the class.
- Offer reflections on your own personal experiences as they relate to topics studied in the course.
- Notify the class of events on- or off-campus that address the issues discussed in class.
- Attend a relevant community event and share your reaction with your classmates and in a written commentary to the lecturer (one paragraph).
- Volunteer at a local organization; document your experience.
- Any other creative way you can actively participate in class.



# #GENAI WHAT ABOUT LLMs?

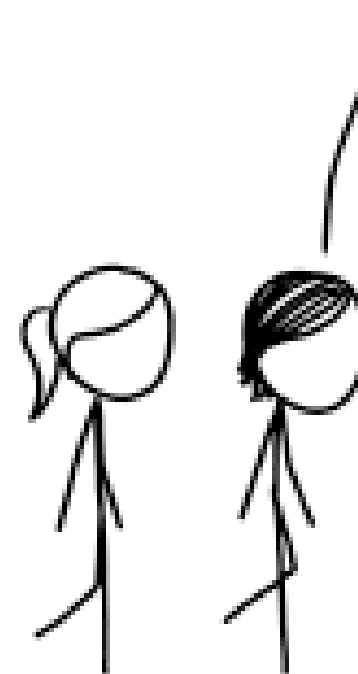
We are all prompt engineers now!



# Language (as data)

- Language as data, the key claim
- Interesting
- Pervasive
- Important
- Challenging

EVERY CHOICE OF PHRASING AND  
SPELLING AND TONE AND TIMING  
CARRIES COUNTLESS SIGNALS AND  
CONTEXTS AND SUBTEXTS AND MORE,  
AND EVERY LISTENER INTERPRETS  
THOSE SIGNALS IN THEIR OWN WAY.  
LANGUAGE ISN'T A FORMAL SYSTEM.  
LANGUAGE IS GLORIOUS CHAOS.



<https://xkcd.com/1576/>

CC BY-NC 2.5





# Some definitions

- Texts
- Discourses
- Text as Data
  - Corpus linguistics, corpus, corpora
  - Text mining
- Humanities
- Digital Humanities



## Some definitions

### **Corpus Linguistics:**

“the study of language based on examples of  
‘real life’ language use”

(McEnery & Wilson, Corpus Linguistics, 1996)



## Some definitions

### **Text mining (or text analytics/analysis):**

“the process of extracting interesting information and knowledge from unstructured text. ...[a] field in the intersection of the related areas information retrieval, machine learning, statistics, computational linguistics and especially data mining”

(Hotho et al., “A Brief Survey of Text Mining”, 2005)



# **What is discourse?**



## What is discourse?

Communication through language, speech or writing, from individual utterances to huge collections of texts.  
OR **‘real life’ language use** (remember our definition of corpus linguistics)

We can apply this to think about what we can study:

- Focus on discourse (language use) associated with institutions/settings/roles. e.g. parliamentary discourse, academic discourse
- Focus on discourse (language use) about \_\_\_\_\_  
e.g. discourse about killer robots / meritocracy / “the economy” / COVID-19 / AI



## Discussion

1. Think about your introduction and the context you were introducing yourself. What might be expected/unexpected in an introduction in this context?
2. US President Donald Trump is well known for talking about “the fake news media”. What might be the effect of using this label?





## **Back to introductions ...**

Think about real life examples of specific kind of language use in an analytical way ...

... to be continued in labs!



# Example of text as data: Google Ngrams

Google Ngram Viewer example: “[the economy](#)”

Ngrams ...

Unigrams (1-grams): “Texts”, “Discourses”, “and”, “Data”

Bigrams (2-grams): “Texts, Discourses”, “Discourses and”, “and Data”

Trigrams (3-grams): “Texts, Discourses and”, “Discourses and Data”



