Text search informed by word frequency and topic modeling

By Seth D Temple (sdtemple@uw.edu)
STAT 599 consulting project
December 15, 2021

W

Background

Setting

- **C**ollaborative **O**nline **I**nternational **L**earning (COIL)
  - Instructors, students in America and Iraq
  - (A)synchronous
  - [Learning outcomes](#)
    - Perspective Taking
    - Cultural Self-Awareness
    - Intercultural Communication

Data

- Pre- and post-course surveys
  - International Interests (II)
  - Intercultural Awareness and Citizenship (IAIC)
  - English Language Use and Learning (ELUL)
- *Four discussion boards*
  - Learning in a pandemic
  - Gender issues
  - English as a lingua franca
  - Food and music

Questions

1. Is there evidence in the survey responses that the students achieved COIL and course-specific learning outcomes?

2. Can we find instances of intercultural learning in texts on the discussion boards?

3. How did the American and Iraqi student experiences in the COIL course differ and/or align? Provide examples in the texts.

4. How did English use in the course facilitate intercultural learning?

Strategy

Surveys

- Paired responses
- Summary statistics on the Likert-scale items (5-point scale)
- Tabular presentation

Discussions

- Study *word frequencies* and associations
- Use NLP toolkit to identify interesting words
- Search for *anchor words* in the text and include flanking sentences for context
- Organize sentences, data viz. to *juxtapose* US and IQ responses

Survey statistics

Table 2 Summary of paired survey responses for English Language Use and Learning (ELUL)

| Prompt	Count of Likert responses				Percent of Likert response >= 4		
	Decreased	Same	Increased	Incr. - Decr.	Pre-survey	Post-survey	Post - Pre
ELUL.2A	3	17	6	3	0.77	0.89	0.12
ELUL.2B	4	12	8	4	0.67	0.73	0.06
ELUL.2C	3	14	9	6	0.85	0.92	0.08
ELUL.2D	1	18	7	6	0.81	0.96	0.15
ELUL.2E	2	16	8	6	0.89	0.92	0.04
ELUL.2F	6	13	7	1	0.81	0.77	-0.04
ELUL.2G	3	16	6	3	0.85	0.96	0.11
ELUL.2H	6	14	6	0	0.81	0.77	-0.04
ELUL.2I	4	13	8	4	0.46	0.48	0.02
ELUL.3A	2	9	9	7	0.56	0.91	0.34
ELUL.3B	1	13	5	4	0.63	0.86	0.23
ELUL.3C	2	10	10	8	0.73	0.82	0.09

Discussion Prompt

Consider issues that present as gender related inequities or challenges facing individuals pursuing their education and/or career/professional development.

Word clouds

US students
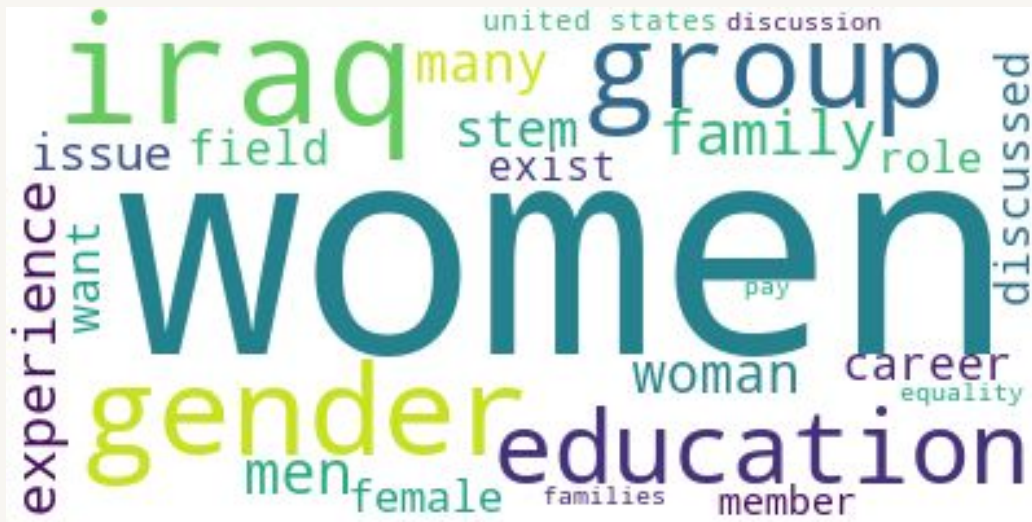
stem, pay, field, career

IQ students

freedom, right, society

n-grams

- Words that appear adjacent to each other w/ high frequency
- Interpretation:
  - Compound words
  - Associated words

print() of Python dictionary

{'gender_inequality': 15, 'gender_issues': 14,

 'united_states': 9, 'gender_equality': 8,

 'religious_freedom': 7, 'gender_inequalities': 7,

 'group_discussed': 5, 'medical_care': 5,

 'coil_partners': 5, 'talked_lot': 5,

 'higher_education': 4, 'similarities_differences': 4,

 'causes_gender': 4, 'group_members': 4}

Topic modeling

- Non-negative matrix factor. (NMF)
- Latent Dir. allocation (LDA)

- Small corpus
- Topics corr. to discussions

Out[32]:

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6
0	english	women	learning	power	girls	lingua
1	language	gender	pandemic	circle	education	franca
2	people	men	covid	masculinity	agreed	message
3	lingua	legal	online	ideally	drop	translation
4	learn	iraq	students	ground	pressured	con
5	languages	inequalities	education	seemingly	perusing	intended
6	franca	woman	impact	break	girl	things
7	learning	abuse	study	unbreakable	partner	effectively
8	used	family	personal	address	students	lost
9	know	inequality	first	linkage	marriage	benefit

9

Sentences

Anchors: stem, field, career

- US students discussed gender issues in scientific careers.
- IQ students did not use the words field, stem, or career as much.

e.g. US student

" in our society i feel that opportunities exist for woman to pursue what they wish however it may be harder for those going in to stem **field**s leadership positions or roles generally considered for men because of different expectations and judging women harsher than men in these **field**s "

Sentences

Anchor: freedom

- IQ students shared their experiences and frustrations with gender-related restrictions in society.
- US students reflected on their privileges and hearing about gender issues in IQ.

e.g. IQ student

" i like to be given the **freedom** to complete my education process and be an important person in my society but there are obstacles and restrictions that bind and suffocate me to play my role perfectly "

11

Python package

Motivation

- Easy-to-use to promote collaboration with experts
- Reproduce for next COIL course

Software Design

- Combine pre-processing and NLP models
- Document functions for future use
- Installation instructions
- Tutorial analyses
- GitHub repository

Next steps

- Summarize methods and analyses in final report

- Annotate word-anchored sentences
- Conduct analyses for next COIL course
- Assist in publication
  - Data visualization
  - Interpretation of NLP methods

References

1. Bird, Steven, Edward Loper and Ewan Klein (2009), Natural Language Processing with Python. O'Reilly Media Inc.
2. Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. Journal of Machine Learning Research, 3, 993-1022.
3. Jonathan K Pritchard, Matthew Stephens, Peter Donnelly, Inference of Population Structure Using Multilocus Genotype Data, Genetics, Volume 155, Issue 2, 1 June 2000, Pages 945–959, https://doi.org/10.1093/genetics/155.2.945
4. Rajaraman, A., & Ullman, J. (2011). Data Mining. In Mining of Massive Datasets (pp. 1-17). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139058452.002
5. Rehurek, R., & Sojka, P. (2010). Software framework for topic modelling with large corpora. In In Proceedings of the LREC 2010 workshop on new challenges for NLP frameworks.
6. Scikit-learn: Machine Learning in Python, Pedregosa et al., Journal of Machine Learning, 12, pp. 2825-2830, 2011.

Appendix

Survey codes

ABBREV	QUESTION
ELUL.2A	It will allow me to get good grades at university.
ELUL.2B	It will allow me to pass exams.
ELUL.2C	I will need it for further study.
ELUL.2D	I will need it for my future career.
ELUL.2E	It will allow me to meet and communicate with more and varied people from many different cultures.
ELUL.2F	It will allow me to travel to many different countries and to learn about different cultures.
ELUL.2G	It will allow me to meet and communicate with native speakers of English.
ELUL.2H	It will allow me to have a fun and enjoyable experience.
ELUL.2I	Other people will respect me more if I have knowledge of the English language.
ELUL.3	Are you planning to study in an English-speaking country in the future?
ELUL.3A	How well do you think your English classes have prepared you for studying abroad?
ELUL.3B	Do you feel well prepared for intercultural communication when studying abroad?
ELUL.3C	Do you hope to develop an identity or feeling of intercultural/global citizenship when you are abroad?