Masculine Defaults via Gendered Discourse in Podcasts and Large Language Models

ICWSM '25

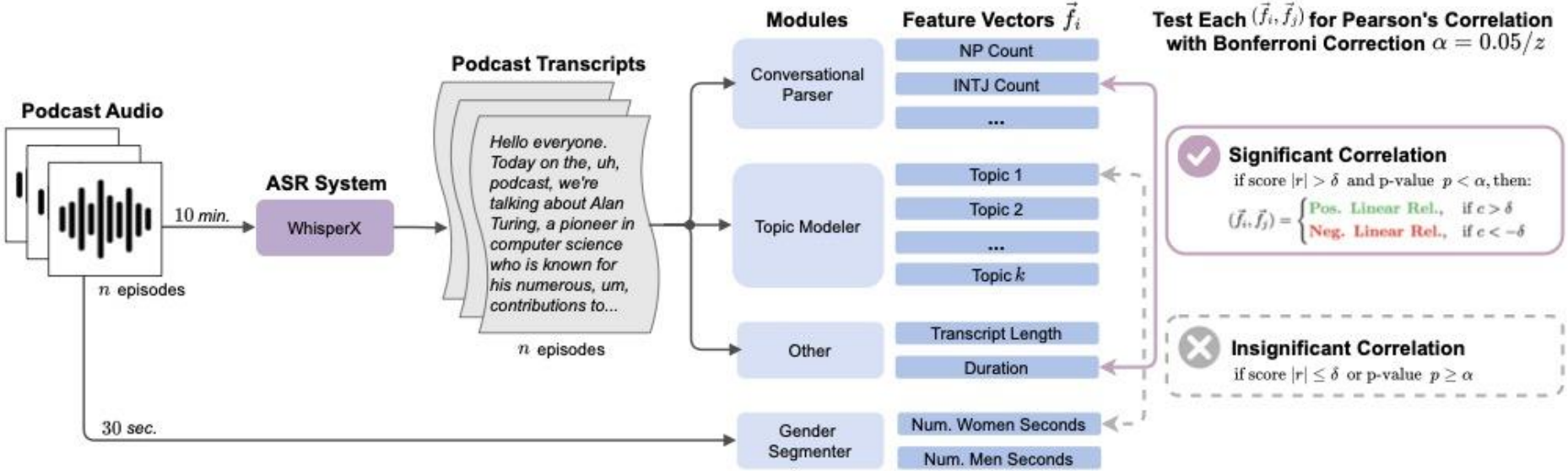
MARIA TELEKI

XIANGJUE DONG

HAORAN LIU

DR. CAVERLEE

Gendered Discourse Correlation Framework (GDCF)



We obtain ***audio and text-based features*** for the ***Spotify Podcasts***, and test for **significant correlations between these features**.

Gendered Discourse Correlation Framework (GDCF)

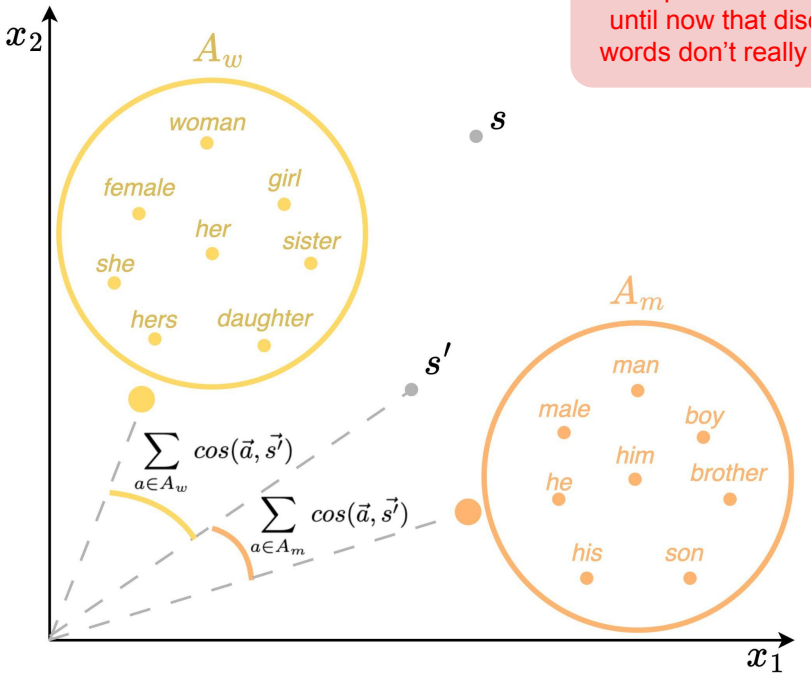
Topic N	Gender	r	Topic N Word List	Topic N Categories	Topic N Gender
Topic 3	Women	0.15	women, woman, men, baby, pregnant, girls, men, doctor, health, birth	Content - Pregnancy	Women
	Men	-0.14			
Topic 10	Women	0.10	energy, body, feel, mind, space, yoga, love, beautiful, feeling, meditation	Content - Yoga	Women
	Men	-0.12			
Topic 49	Women	-0.21	game, know, think, team, going, mean, play, year, one, good	Content - Sports	Men
	Men	0.17			
Topic 71	Women	0.14	christmas, sex, girl, hair, love, get, date, girls, let, wear	Content - Dating	Women
	Men	-0.14			
Topic 54	Women	–	get, like, know, right, people, going, podcast, make, want, one	Discourse	Men
	Men	0.12			
Topic 60	Women	-0.27	going, know, think, get, got, one, really, good, well, yeah	Discourse	Men
	Men	0.20			
Topic 62	Women	0.33	like, know, really, going, people, want, think, get, things, life	Discourse	Women
	Men	-0.28			

$s =$ *And I was* ***going****, hey, it's cold outside...*

Discourse Word-Embedding Association Test (D-WEAT)

Topic 60	Women	-0.27	going, know, think, get, got, one, really, good, well, yeah	Discourse	Men
	Men	0.20			
Topic 62	Women	0.33	like, know, really, going, people, want, think, get, things, life	Discourse	Women
	Men	-0.28			

People have assumed until now that discourse words don't really matter!



We set up an experiment to measure: What happens if we swap the discourse words? Does the sentence "move closer" to the other gender?

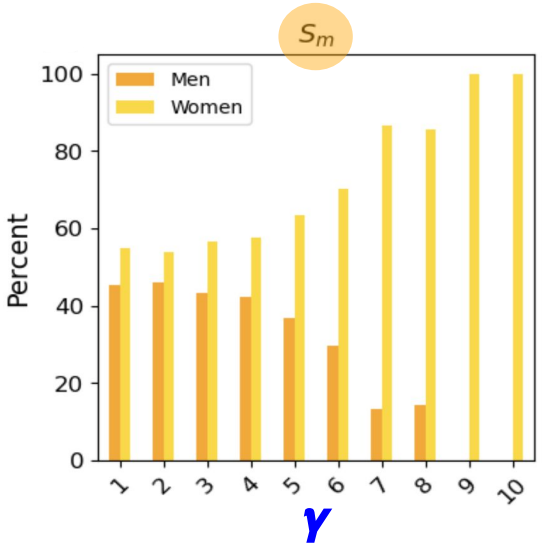
$s =$ *And I was* ***going****, hey, it's cold outside...*

$s' =$ *And I was* ***like****, hey, it's cold outside...*

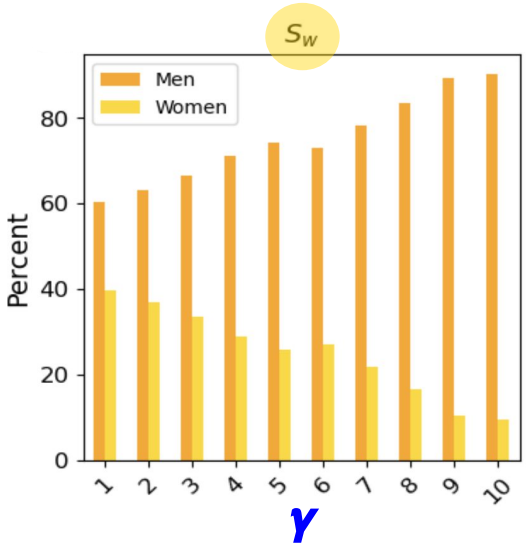
S_m: masculine → feminine discourse word replacement

S_w: feminine → masculine discourse word replacement

Impact of γ



We see that the embedding moves towards the **feminine concept** in the embedding space.



We see that the embedding moves towards the **masculine concept** in the embedding space.

We also see that the overall ***gap is bigger for the S_w sentences than the S_m sentences*** – meaning **men have a more robust discourse embedding representation than women.**

What is γ?

s = *And I was* ***going****, hey, it's cold outside...*
s' = *And I was* ***like****, hey, it's cold outside...*

In this example, γ=1, because we do 1 discourse word replacement.

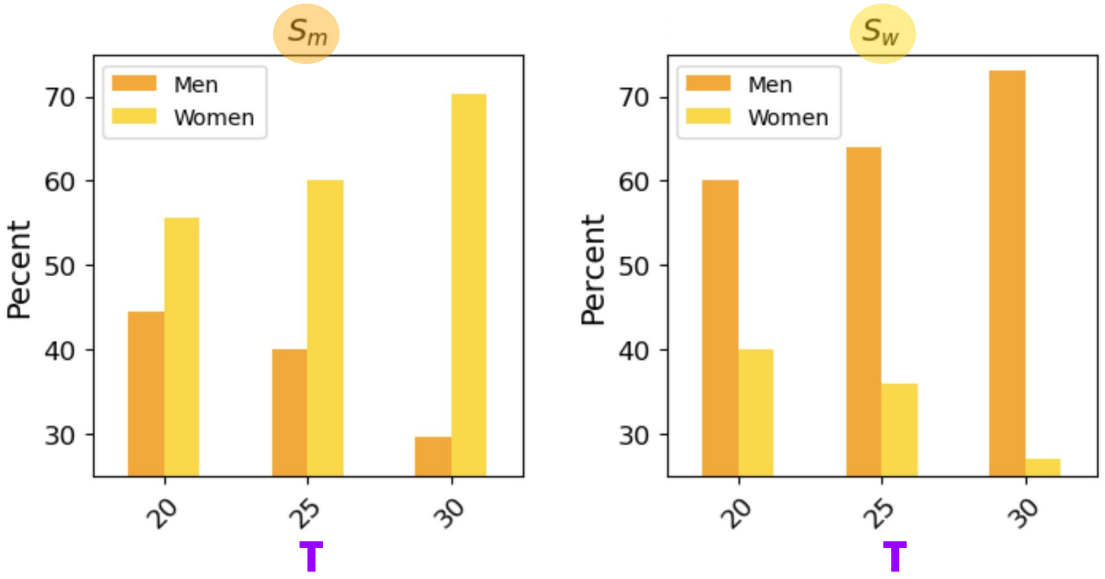
What are S_m and S_w?

s = *And I was* ***going****, hey, it's cold outside...*
s' = *And I was* ***like****, hey, it's cold outside...*

S_m: masculine → feminine discourse word replacement

S_w: feminine → masculine discourse word replacement

Impact of T

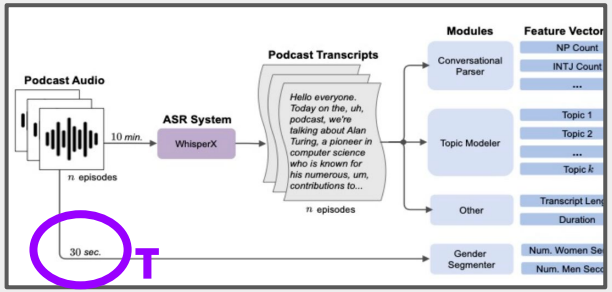


We see that the embedding moves towards the **feminine concept** in the embedding space.

We see that the embedding moves towards the **masculine concept** in the embedding space.

We also see that the overall ***gap is bigger for the S_w sentences than the S_m sentences*** – meaning **men have a more robust discourse embedding representation than women.**

What is T?



This variable is T, the # of seconds we take from the podcast audio for our gender features in the GDCF pipeline.

What are S_m and S_w?

s = *And I was* ***going****, hey, it's cold outside...*

s' = *And I was* ***like****, hey, it's cold outside...*

S_m: masculine → feminine discourse word replacement

S_w: feminine → masculine discourse word replacement

Thanks!

mariateleki@tamu.edu