11th International Conference on Computational Social Science IC^{2}S^2
July 21-24, 2025, Norrköping, Sweden

Masculine Defaults via Gendered Discourse in Podcasts and Large Language Models

Keywords: Masculine Defaults, LLM, Discourse, Disfluency, Podcasts

Extended Abstract

Masculine defaults are a type of gender bias "in which characteristics and behaviors associated with the male gender role are valued, rewarded, or regarded as standard, normal, neutral, or necessary aspects of a given cultural context" [1], and hence result in the *other-ing* of women [2]. Considerable prior research has examined gender differences in social media (e.g., [3]–[6]) and in LLMs (e.g., [7]–[10]). But how do *masculine defaults* manifest on social media? And how do they impact emerging systems like large language models (LLMs) that are trained in part over social media?

While masculine defaults are highly connected to gender differences,[1] there is a research gap in identifying and analyzing masculine defaults that arise through *gender differences in discourse*. Specifically, we focus on patterns of discourse in spoken communication, including fillers (e.g., *uh*, *um*), discourse markers (e.g., *well*, *you know*, *I mean*), false starts (e.g., *It was, anyways, I went to Target yesterday*) and more [15], [16]. Such discourse words are non-content related words that serve important social purposes with respect to gender, such as to *"hold the floor"* in conversation [16], [17]. Previous work notes gender differences in how men and women use specific types of *discourse words* – for example, men use more filled pauses and repeats [17], [18] than women. However, these studies lack an automated method for large-scale discourse word discovery and gender analysis, primarily relying on the Switchboard corpus [19] – a corpus which is not representative of the range of natural speech patterns, as the phone calls were recorded in the manufactured, awkward situation of randomly-pairing two callers and assigning them a topic to discuss.

Hence, we propose in this paper a twofold framework for (i) the large-scale discovery and analysis of gendered discourse words in spoken content via our **Gendered Discourse Correlation Framework (GDCF, shown in Figure 1)**; and (ii) the measurement of the gender bias associated with these gendered discourse words in LLMs via our **Discourse Word-Embedding Association Test (D-WEAT, shown in Figure 2)**. Concretely, we focus our study on podcasts, a popular and growing form of social media [20], [21]. We analyze 15,117 podcast episodes from the Spotify Podcast Dataset [20], to discover the *rewards* associated with *masculine discourse words* in terms of (i) correlated domains with substantial economic rewards, and (ii) more stable LLM representations. The presence of rewards for these *masculine discourse words* means that they indeed constitute *masculine defaults* [1]. We structure our work in terms of research questions, as detailed in the following sections.

Research Question 0: How are women and men's discourse different? We first introduce our *Gendered Discourse Correlation Framework (GDCF)* as shown in Figure 1, a framework

[1] We consider the binary definitions of sex (female/male) and gender (women/men, feminine/masculine) in our work due to (i) continuity with previous work in the gender debiasing task in the NLP community [8], [10], and (ii) modeling constraints – i.e., *inaSpeechSegmenter* [11] for gender approximation via audio signal. This definition, however, is not representative of the sex and gender spectrums – and transgender, intersex, intersectional identities, and other identities are also not represented in this binary definition [12]–[14]. This is an important direction for future work.

for discovering gendered discourse words, with features which are centered around spoken content – specifically, an audio-based GENDER SEGMENTER [11], a TOPIC MODELER via LDA [22] and BERTopic [23], and a specialized CONVERSATIONAL PARSER [24]. We analyze correlations between *gender* and *discourse words* to automatically form gendered discourse word lists, as shown in Tables 1 and 2. Additionally, GDCF is a flexible framework which can be extended to other forms of audio speech data – such as short videos that are prevalent on TikTok, Instagram, and YouTube, long videos on YouTube, streamers on Twitch, and more.

Research Question 1: Are discourse-based masculine defaults present in domain-specific contexts? We then study the prevalence of these gendered discourse words in domain-specific contexts, as shown in Table 3. We find that masculine discourse words are positively correlated with the business domain, the technology/politics domain, and the video games domain. Participation in these domains grants economic *rewards* [1], hence there are indeed discourse-based masculine defaults present.

Research Question 2: Are discourse-based masculine defaults present in LLM embeddings? Finally, we study the representation of these gendered discourse words as shown in Figure 2, using a state-of-the-art LLM embeddings model from OpenAI, `text-embedding-3-large`. We find that the masculine discourse words have a more stable and robust representation than the feminine discourse words, as shown in Figures 3 and 4, resulting in better system performance on downstream tasks for men. Hence, men are *rewarded* [1] for their discourse patterns with better system performance by one of the state-of-the-art language models – and therefore this difference in the embedding representations for women and men constitutes a masculine default [1] and a *representational harm* [25].

We consider a few key types of implications pertaining to our study of gendered discourse. **(1) Theoretical Implications**: First, the use of gendered discourse words can be considered a type of *gender performativity* [26]–[30], wherein the discourse words are part of a *gender schema* [28], [31]. Hence, we identify specific words which are part of the current *hegemonic masculine* strategy [32], [33] – and in the domain of technology, discourse words which are part of the *technomasculine* strategy [34]–[36]. We contribute GDCF (Figure 1) for the discovery and analysis of gendered discourse words. Second, we contribute D-WEAT as an intrinsic metric which can be used to debias LLMs, broadening the debiasing task in natural language processing. **(2) Policy Implications**: Policymakers – in government or platforms such as Spotify – could implement measures by which to mitigate bias in LLMs with respect to gender. Specifically, policymakers could regulate the use of D-WEAT to impose an unbiased representation of discourse words with respect to gender. D-WEAT could be run regularly, and a threshold could be set to determine what an "acceptable" level of bias is in a given LLM. Broadly, D-WEAT can join *a set of debiasing methods, tools, and datasets* [7]–[10], [37]–[41] which can be employed to regulate bias in LLMs. **(3) Ethical Implications**: A potential ethical concern is that tools used to remove bias can also be used to exacerbate bias. GDCF and D-WEAT could potentially be used to discover discourse words in audio-text corpora, and then *increase* the gender bias of the LLM embeddings. This abuse of the framework would be a *representational harm* [25]. However, a more important point is that it is hard to undo bias issues without knowing how that bias manifests; here, we provide a framework to identify and quantify this subtle gender bias so that it can be undone in powerful LLMs.

Note: This work has been accepted to the 2025 International AAAI Conference on Web and Social Media (ICWSM 2025).

References

[1] S. Cheryan and H. R. Markus, "Masculine defaults: Identifying and mitigating hidden cultural biases.," *Psychological Review*, vol. 127, no. 6, pp. 1022–1052, 2020, ISSN: 0033-295X. DOI: 10.1037/rev0000209.

[2] S. d. Beauvoir, *The Second Sex*. 1949.

[3] A. Wang, A. Pappu, and H. Cramer, "Representation of music creators on wikipedia, differences in gender and genre," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 15, 2021, pp. 764–775.

[4] G. Kalhor, H. Gardner, I. Weber, and R. Kashyap, "Gender gaps in online social connectivity, promotion and relocation reports on linkedin," in *Proceedings of the International AAAI Conference on Web and Social Media*, 2023.

[5] I. Johnson, F. Lemmerich, D. Sáez-Trumper, R. West, M. Strohmaier, and L. Zia, "Global gender differences in wikipedia readership," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 15, 2021, pp. 254–265.

[6] Y. Wang and E.-Á. Horvát, "Gender differences in the global music industry: Evidence from musicbrainz and the echo nest," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 13, 2019, pp. 517–526.

[7] X. Dong, Z. Zhu, Z. Wang, M. Teleki, and J. Caverlee, "Co^2PT: Mitigating bias in pretrained language models through counterfactual contrastive prompt tuning," in *Findings of the Association for Computational Linguistics: EMNLP 2023*, H. Bouamor, J. Pino, and K. Bali, Eds., Association for Computational Linguistics, Dec. 2023, pp. 5859–5871. DOI: 10.18653/v1/2023.findings-emnlp.390. [Online]. Available: https://aclanthology.org/2023.findings-emnlp.390.

[8] A. Caliskan, J. J. Bryson, and A. Narayanan, "Semantics derived automatically from language corpora contain human-like biases," *Science*, vol. 356, no. 6334, pp. 183–186, 2017, ISSN: 0036-8075. DOI: 10.1126/science.aal4230.

[9] C. May, A. Wang, S. Bordia, S. R. Bowman, and R. Rudinger, "On measuring social biases in sentence encoders," *arXiv*, 2019. DOI: 10.48550/arxiv.1903.10561.

[10] T. Bolukbasi, K.-W. Chang, J. Zou, V. Saligrama, and A. Kalai, "Man is to computer programmer as woman is to homemaker? debiasing word embeddings," *arXiv*, 2016. DOI: 10.48550/arxiv.1607.06520.

[11] D. Doukhan, J. Carrive, F. Vallet, *et al.*, "An open-source speaker gender detection framework for monitoring gender equality," in *ICASSP*, IEEE, 2018.

[12] B. Ghai, M. N. Hoque, and K. Mueller, "Wordbias: An interactive visual tool for discovering intersectional biases encoded in word embeddings," in *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021, pp. 1–7.

[13] A. Ovalle, P. Goyal, J. Dhamala, Z. Jaggers, K.-W. Chang, A. Galstyan, R. Zemel, and R. Gupta, """i'm fully who i am": Towards centering transgender and non-binary voices to measure biases in open language generation," *2023 ACM Conference on Fairness, Accountability, and Transparency*, pp. 1246–1266, 2023. DOI: 10.1145/3593013.3594078.

[14] K. Seaborn, S. Chandra, and T. Fabre, "Transcending the "male code": Implicit masculine biases in nlp contexts," *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–19, 2023. DOI: 10.1145/3544548.3581017.

[15] Merriam-Webster, *Discourse*, 2024. [Online]. Available: https://www.merriam-webster.com/dictionary/discourse.

[16] E. E. Shriberg, "Preliminaries to a theory of speech disfluencies," Ph.D. dissertation, 1994.

[17] E. Shriberg, "Disfluencies in switchboard," in *International Conference on Spoken Language Processing*, 1996.

[18] H. Bortfeld, S. D. Leon, J. E. Bloom, *et al.*, "Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender," *Language and Speech*, vol. 44, no. 2, pp. 123–147, 2001.

[19] M. Mitchell, B. Santorini, M. Marcinkiewicz, *et al.*, "Treebank-3 ldc99t42 web download," *Linguistic Data Consortium*, vol. 3, p. 2, 1999.

[20] A. Clifton, S. Reddy, Y. Yu, *et al.*, "100,000 podcasts: A spoken english document corpus," in *COLING*, 2020, pp. 5903–5917.

[21] C. S. A. The Pew Research Center. "Audio and Podcasting Fact Sheet," The Pew Research Center. (2023), [Online]. Available: https://www.pewresearch.org/journalism/fact-sheet/audio-and-podcasting/.

[22] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, no. Jan, pp. 993–1022, 2003.

[23] M. Grootendorst, "Bertopic: Neural topic modeling with a class-based tf-idf procedure," *arXiv preprint arXiv:2203.05794*, 2022.

[24] P. Jamshid Lou and M. Johnson, "Improving disfluency detection by self-training a self-attentive model," in *ACL*, 2020.

[25] S. L. Blodgett, S. Barocas, H. D. III, and H. Wallach, "Language (technology) is power: A critical survey of "bias" in nlp," *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 5454–5476, 2020. DOI: 10.18653/v1/2020.acl-main.485.

[26] J. Butler, "Performative acts and gender constitution an essay in phenomenology and feminist theory," *Theatre Journal*, vol. 40, no. 4, p. 519, 1988.

[27] J. Butler, "Performativity, precarity and sexual politics.," *AIBR. Revista de Antropología Iberoamericana*, vol. 4, no. 3, 2009.

[28] C. West and D. Zimmerman, "Doing gender," *Gender and Society*, vol. 1, pp. 125–151, 1987.

[29] R. K. Unger, "Toward a redefinition of sex and gender," *American Psychologist*, vol. 34, no. 11, pp. 1085–1094, 1979, ISSN: 0003-066X. DOI: 10.1037/0003-066x.34.11.1085.

[30] C. L. Muehlenhard and Z. D. Peterson, "Distinguishing between sex and gender: History, current conceptualizations, and implications," *Sex Roles*, vol. 64, no. 11–12, pp. 791–803, 2011, ISSN: 0360-0025. DOI: 10.1007/s11199-011-9932-5.

[31] S. L. Bem, "Androgyny and gender schema theory: A conceptual and empirical integration.," *Nebraska Symposium on Motivation. Nebraska Symposium on Motivation*, vol. 32, pp. 179–226, 1984, ISSN: 0146-7875.

[32] R. Connell, *Masculinities*. Allen & Unwin, 1995.

[33] R. Connell, *Gender and power: society, the person, and sexual politics*. Stanford University Press, 1987.

[34] M. Cooper, "Being the "go-to guy": Fatherhood, masculinity, and the organization of work in silicon valley," *Qualitative Sociology*, vol. 23, no. 4, pp. 379–405, 2000, ISSN: 0162-0436. DOI: `10.1023/a:1005522707921`.

[35] E. A. Lockhart, "Nerd/geek masculinity: Technocracy, rationality, and gender in nerd culture's countermasculine hegemony," Ph.D. dissertation, 2015.

[36] E. Bulut, *The Illusion of Dream Jobs in the Video Game Industry*. Ithaca, NY: Cornell University Press, 2020, ISBN: 9781501746543. DOI: `doi:10.1515/9781501746543`. [Online]. Available: `https://doi.org/10.1515/9781501746543`.

[37] N. Nangia, C. Vania, R. Bhalerao, and S. R. Bowman, "Crows-pairs: A challenge dataset for measuring social biases in masked language models," *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1953–1967, 2020. DOI: `10.18653/v1/2020.emnlp-main.154`.

[38] M. Nadeem, A. Bethke, and S. Reddy, "Stereoset: Measuring stereotypical bias in pretrained language models," *arXiv*, 2020. DOI: `10.48550/arxiv.2004.09456`.

[39] Y. Guo, Y. Yang, and A. Abbasi, "Auto-debias: Debiasing masked language models with automated biased prompts," *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1012–1023, 2022. DOI: `10.18653/v1/2022.acl-long.72`.

[40] J. He, M. Xia, C. Fellbaum, and D. Chen, "Mabel: Attenuating gender bias using textual entailment data," *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pp. 9681–9702, 2022. DOI: `10.18653/v1/2022.emnlp-main.657`.

[41] M. Cheng, E. Durmus, and D. Jurafsky, "Marked personas: Using natural language prompts to measure stereotypes in language models," *arXiv*, 2023. DOI: `10.48550/arxiv.2305.18189`.

Figures

`images/BigDiagram.pdf`

Figure 1: GDCF (Gendered Discourse Correlation Framework) Diagram: Testing for correlations with an example of a significant correlation and an insignificant correlation – all $(\vec{f_i}, \vec{f_j})$ pairs are labeled *significant* or *insignificant*. $|\vec{f_i}| = 15,117$ podcast episodes. $z = \binom{124}{2} = 7,626$ correlation tests for the 124 total feature vectors.

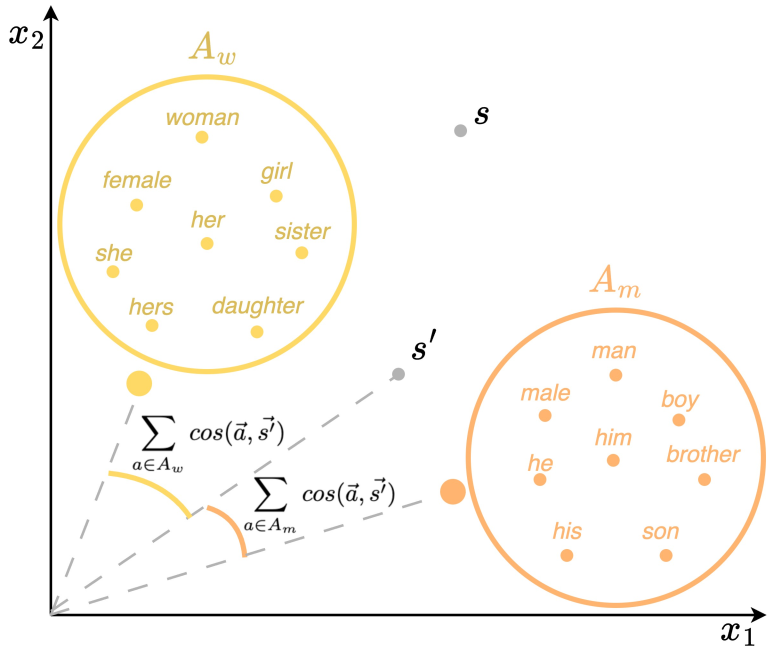

Figure 2: D-WEAT: Plot of the segment vectors $\vec{s}$ and $\vec{s'}$, and the word vectors, $\vec{w} \in A_w$, and $\vec{w} \in A_m$, projected into a two-dimensional space for illustrative purposes. The cosine similarity for s' and A_w, and s' and A_m is depicted; the cosine similarity for s and A_w, and s and A_m is calculated in the same way.

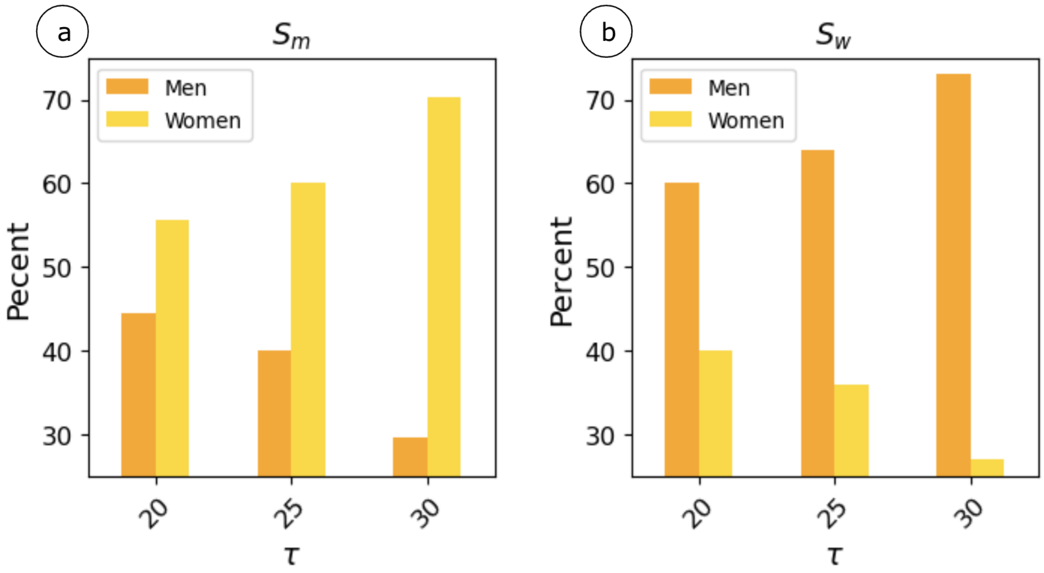

Figure 3: ⓐ Impact of τ on the average percentage of S_m segments which move closer to the *women* concept (A_w) versus the *men* (A_m) concept. ⓑ Impact of τ on the average percentage of S_w segments which move closer to the *women* concept (A_w) versus the *men* (A_m) concept.

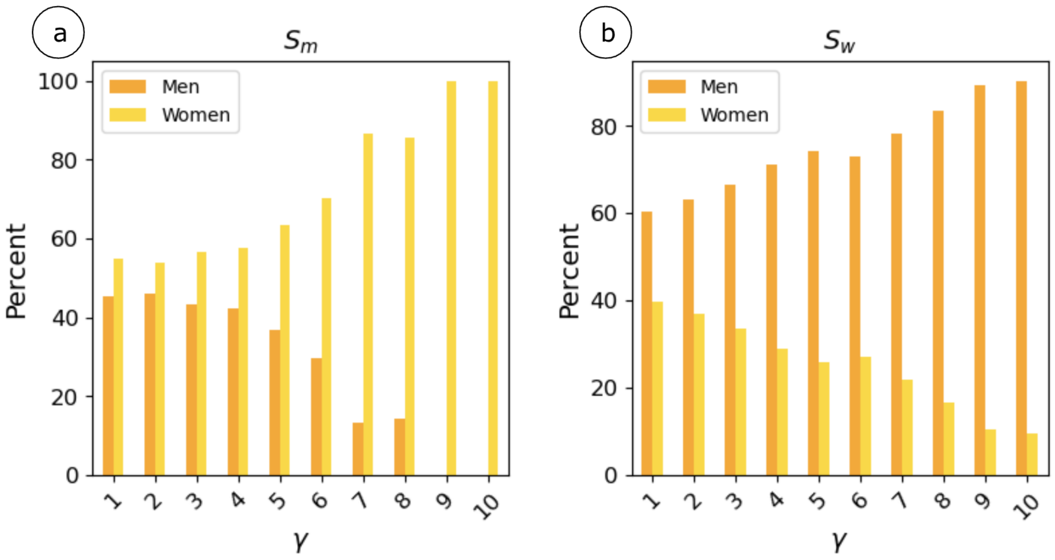

Figure 4: ⓐ Impact of γ on the average percentage of S_m segments which move closer to the women concept (A_w) versus the men (A_m) concept. ⓑ Impact of γ on the average percentage of S_w segments which move closer to the women concept (A_w) versus the men (A_m) concept.

Tables

Table 1: **LDA with Non-Contextual Embeddings (Bag-Of-Words):** The complete set of significant correlations between gender features and topic features – *both content topics and discourse topics*. Based on r, the Topic N Gender forms the **gendered discourse word lists** via Topics 54 and 60 (the masculine word lists) and Topic 62 (the feminine word list).

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			

Table 2: **BERTopic with Contextual Embeddings (BERT, ChatGPT, Llama):** The complete set of significant correlations between gender features and topic features for *discourse topics only* (content topics are omitted).

Topic N	Gender	r	Topic N Word List	Topic N Categories	Topic N Gender
Topic 0	Women	-0.08	like, yeah, know, oh, right, podcast, got, going, think, really	Discourse	Men
	Men	0.10			
Topic 2	Women	0.08	life, know, things, really, people, feel, like, want, love, going	Discourse	Women
	Men	-0.08			
Topic 5	Women	0.08	like, know, think, yeah, episode, really, going, anchor, kind, right	Discourse	Women
	Men	–			

Table 3: LDA with Non-Contextual Embeddings (Bag-Of-Words): Significant correlations between content topic features and **gendered discourse word lists** (discourse topic features 54, 60, 62, see Table 1) for content topic features which *do not* have direct, significant correlations with gender features, but may broadly be more used by one gender.

Topic N	**Topic M**	*r*	**Topic N Word List**	**Topic N Categories**	**Topic M Word List**	**Topic M Categories**
Topic 11	Topic 54	0.11	data, new, technology, public, bill, theory, science, system, security, article	Content - Technology/ Political	get, like, know, right, people, going, podcast, make, want, one	Discourse (Men)
	Topic 62	-0.20			like, know, really, going, people, want, think, get, things, life	Discourse (Women)
Topic 12	Topic 54	0.24	business, money, company, market, buy, right, million, companies, pay, sell	Content - Business	get, like, know, right, people, going, podcast, make, want, one	Discourse (Men)
Topic 79	Topic 60	0.18	game, games, play, playing, like, played, nintendo, video, fun, switch	Content - Video Games	going, know, think, get, got, one, really, good, well, yeah	Discourse (Men)
	Topic 62	-0.13			like, know, really, going, people, want, think, get, things, life	Discourse (Women)