

Wanita in Parliaments: The attitude of Malaysian MPs towards women

COMPTExT conference 2022

Soon Yung (Soung) Low, MSc Applied Social Data Science, LSE
Research interest: Social inequality, stereotypes (races and gender)

Introduction

An exploratory analysis of how women are perceived in the parliament
Data: Hansards of parliamentary debates in Malaysia from 1990 to 2020*

1. The linguistic usage of gender labels (women and men)
 - In terms of frequency and semantic meaning
 - Kulkarni et al. (2014)
2. Stereotypes against women as opposed to men
 - In terms of personal traits and occupations
 - Garg et al. (2018) and Jones et al. (2020)

**This period covers 7 parliaments, from the 8th until the 14th.*

Research question

Attitude conceptualized by **attention** and **association** (stereotypes).

1. How frequent are women being mentioned in the parliamentary debates?
2. What are being associated with women? Has it changed over time?
3. Which personal traits / occupations are more closely associated with women in comparison to men?

RQ1: Frequency of gender labels

- Measure raw frequency and normalized frequency¹

Gender	Labels ²
Female	perempuan (woman), wanita (woman), isteri (wife), pemudi (girl), gadis (girl), datin (an honorary title for Malaysian women)
Male	lelaki (man), male, laki (man), suami (husband), pemuda (boy), datuk (an honorary title for Malaysian men)

¹Normalized frequency = Raw frequency / total word counts

²Primarily sourced from the excerpts of Hansards in David and Ngeow (2007)

RQ2: Semantic shift of gender labels

- Train word embeddings* at the yearly level to capture semantic meanings
- Measure semantic shifts by computing the difference in cosine similarity between their vectorized representations in 1991 and the rest of years
- Explore the semantic relationship of the gender label with other words by identifying the closest words with the label in the embedding space.

**word2vec (window=5 +, skip-gram + negative sampling)*

RQ3: Gender stereotypes

- personal trait stereotypes and occupational stereotypes

Personal traits	Occupations
<p>Mainly consist of adjectives:</p> <ul style="list-style-type: none">• David and Yeok (2007)• Garg et al. (2018)• POS tagging using ALBERT (Lan et al. 2020)	<ul style="list-style-type: none">• Garg et al. (2018)• Occupations identified in the Hansards

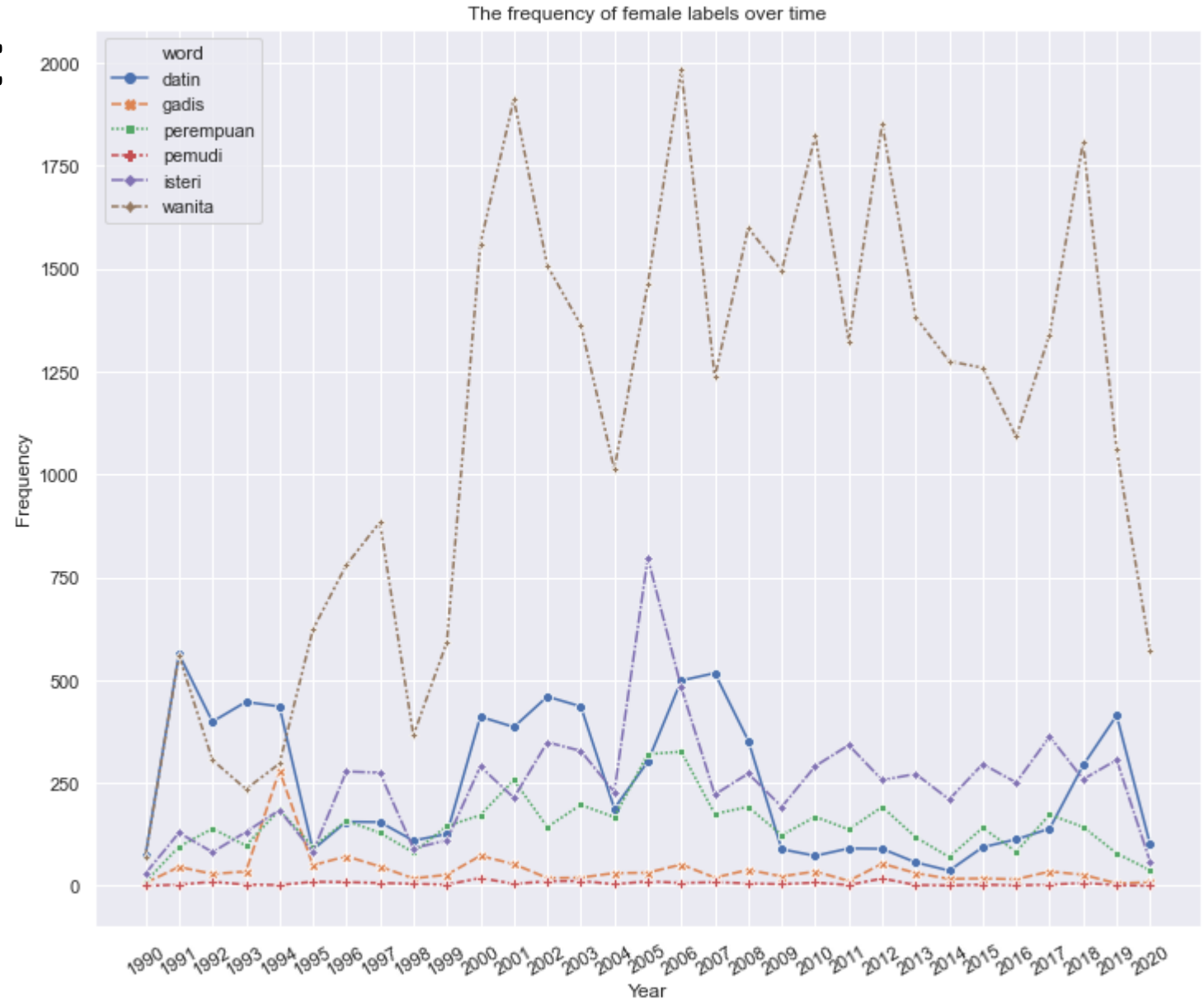
RQ3: Gender stereotypes

1. Calculate the **gender group vector** which equals the average vector of the gender labels
2. Calculate the **cosine distance** between the gender group vector and occupations/adjectives.
3. Compute the **stereotype score** by taking the difference in distances for female and male groups.

$$\textit{stereotype distance} = \textit{Dist}_{f,adj} - \textit{Dist}_{m,adj}$$

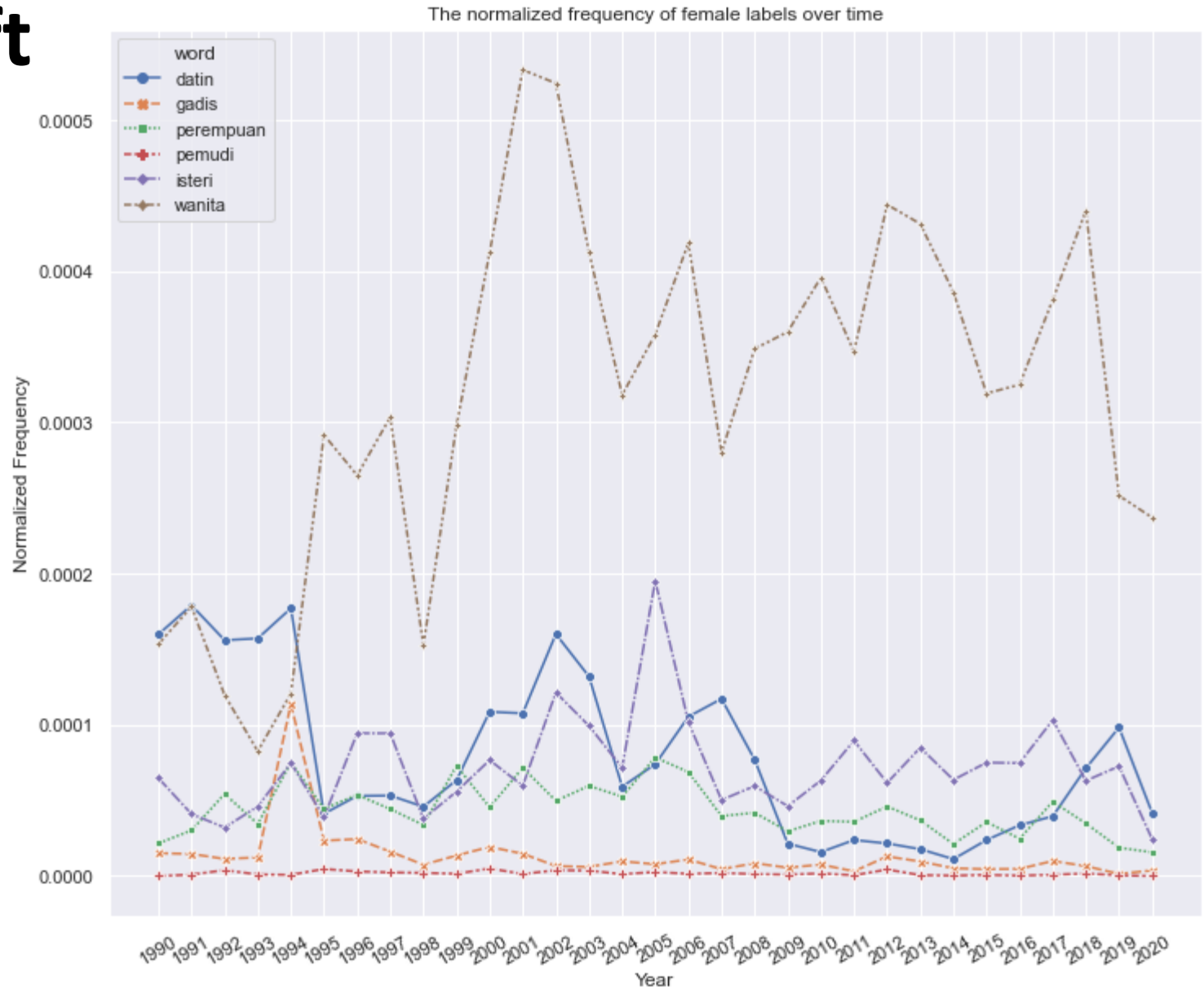
Frequency-based shift

- Mentions of women have remained at a higher level in the last 20 years
- wanita (woman) being the most frequently used female label
- Rahim (2005): the word wanita is often used in a formal context such as parliamentary debates



Frequency-based shift

- Highest normalized frequency of “wanita” in 2001
- the establishment of the Ministry of Women, Family and Community Development (KPWKM) in 2001



Keywords-in-context (KWIC)

- Some examples of the context where “wanita” appeared in the Hansards of 2001:

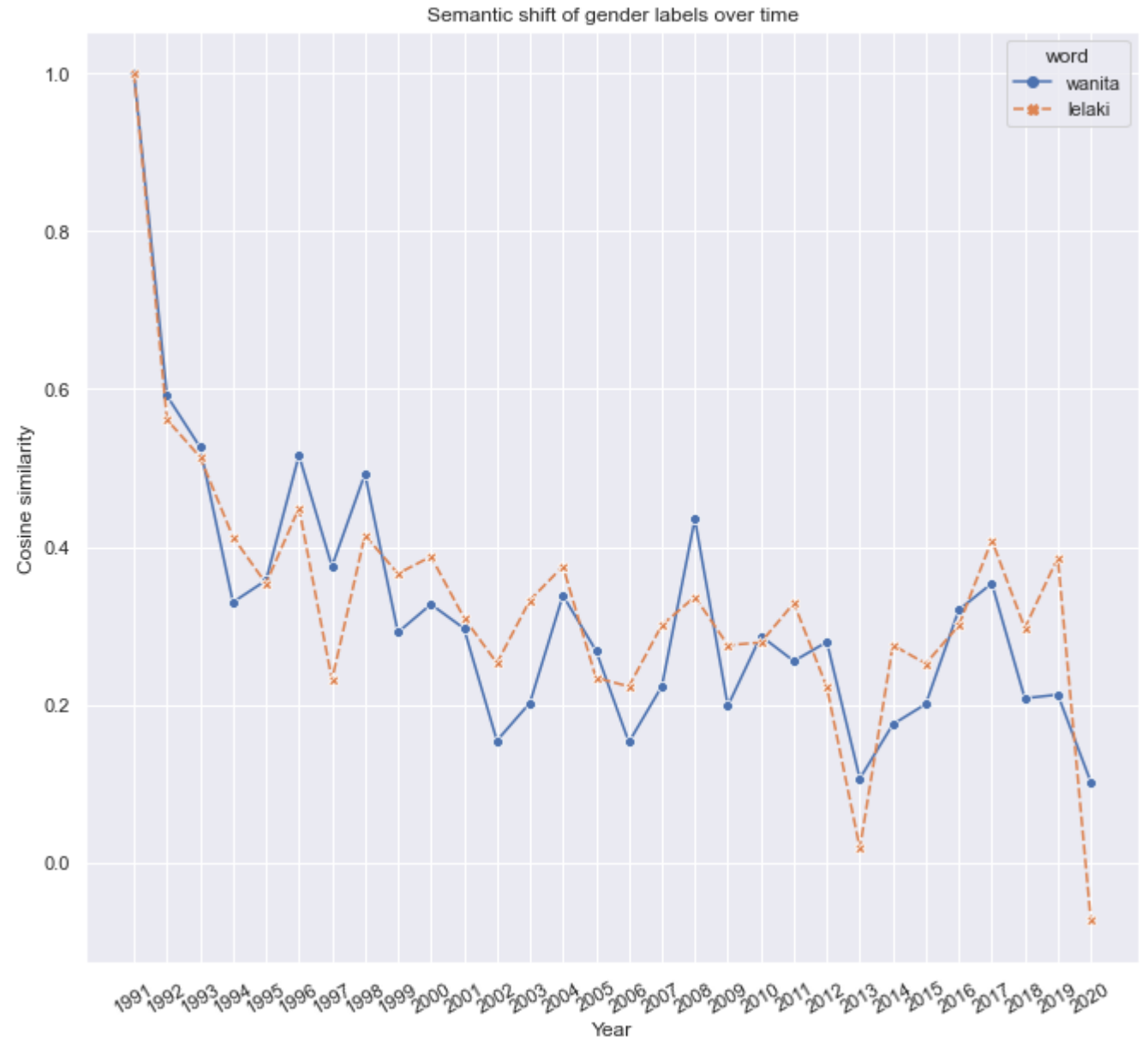
*“Datuk Effendi Norwawi (Kuala Rajang) Menteri Pembangunan **Wanita** dan Keluarga...”
(2 April 2001)*

*“perancangan Kementerian Pembangunan **Wanita** dan Keluarga kita untuk mempromosikan ...”
(3 May 2001)*

*“yang Menteri Pembangunan **Wanita** dan Keluarga telah tiga kali menyebut peruntukan”
(1 August 2001)*

Semantic shift

- Changes in self-distance using the representation in 1991 as the baseline
- the perception of genders in the parliament has remained roughly stable over the last 30 years
- a close semantic relationship between “wanita” and “lelaki” as evidenced by highly correlated trends
- Aziz (2019): high collocation of wanita-lelaki/woman-man using MI score



Closest words in the embedding space

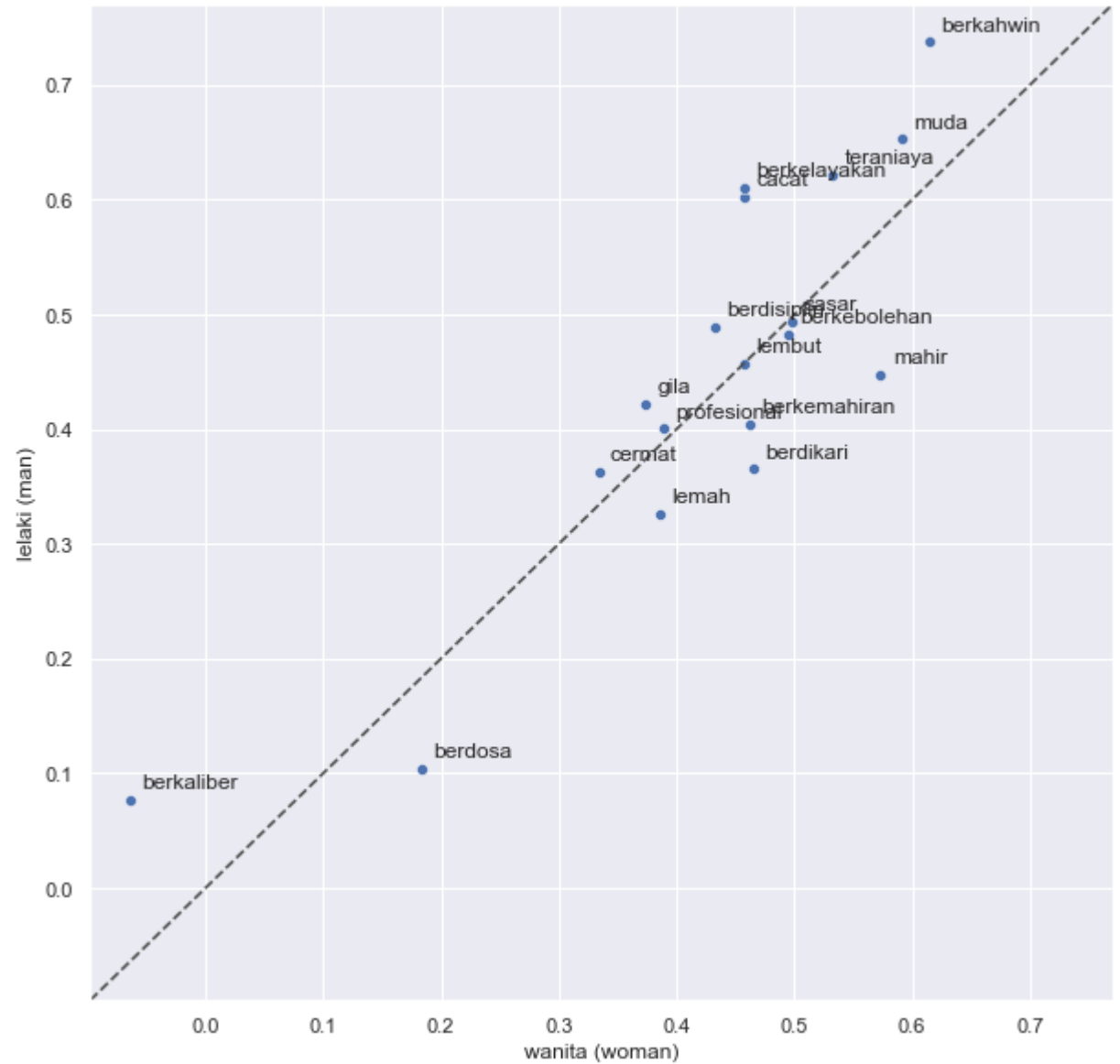
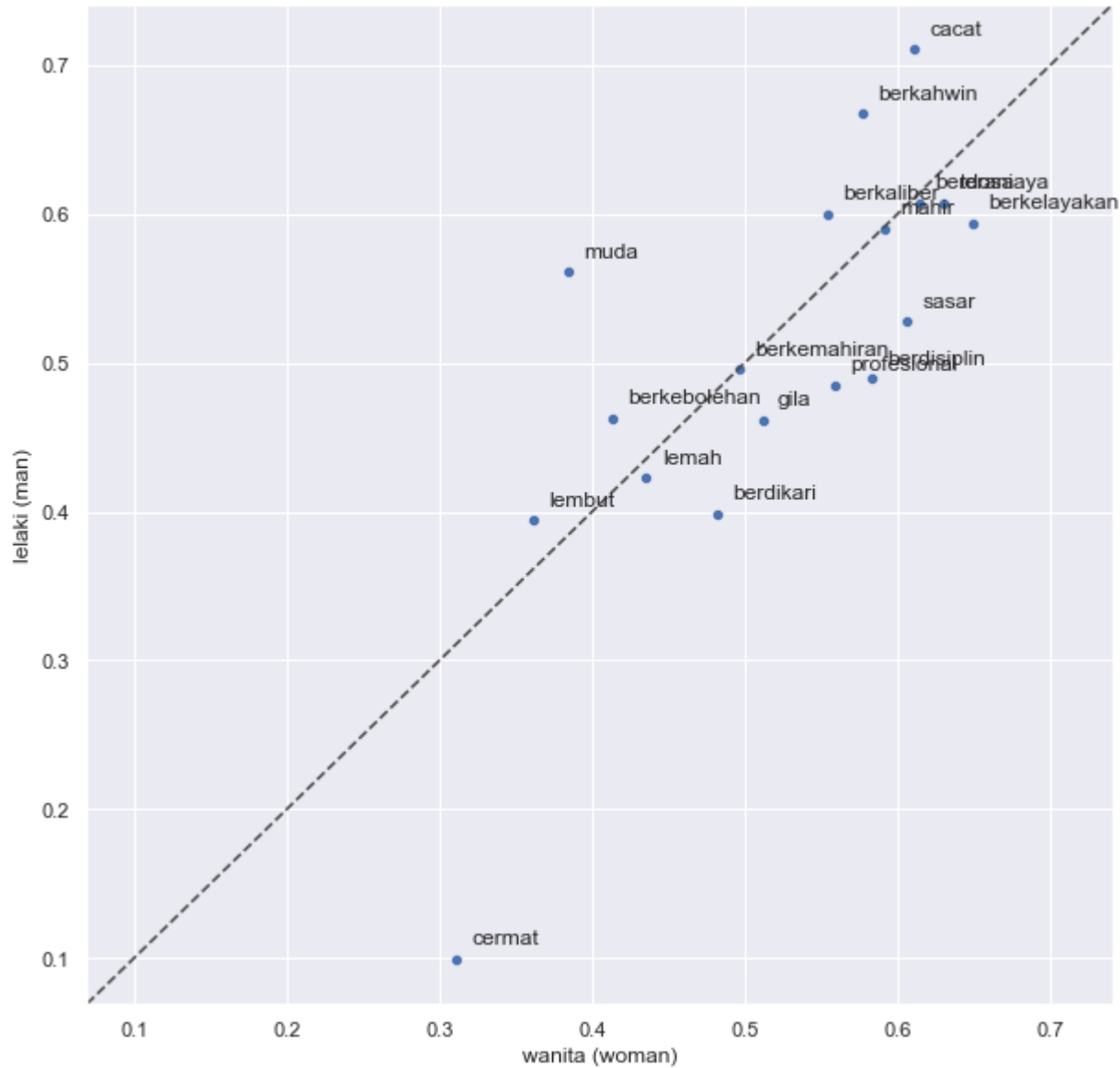
Year	Top six associated words with wanita					
1991	lelaki (man)	uniform (uniform)	terjebak (stuck)	bekerja (working)	penggajian (payroll)	rumahtangga (household)
1995	komuniti (community)	gadis (girl)	remaja (adolescents)	mengharumkan (make better)	saling (mutual)	penaja (sponsor)
2000	kaum (race/group)	lelaki (man)	teraniaya (humiliated)	sayang (dear)	mengandung (pregnant)	didera (abused)
2005	keluarga (family)	masyarakat (society)	berpecah (broken)	jelebu (a place name)	pembangunan (development)	orangnya (the person)
2010	keluarga (family)	masyarakat (society)	pembangunan (development)	lelaki (man)	lumut (a place name)	direstui (blessed)
2015	keluarga (family)	masyarakat (society)	pembangunan (development)	selangau (a place name)	tuala (towel)	datin (woman's title)
2020	masyarakat (society)	keluarga (family)	menjawabnya (answer)	ketereh (a place name)	sasar (target/crazy)	lelaki (man)

Stereotypes in personal traits and occupations

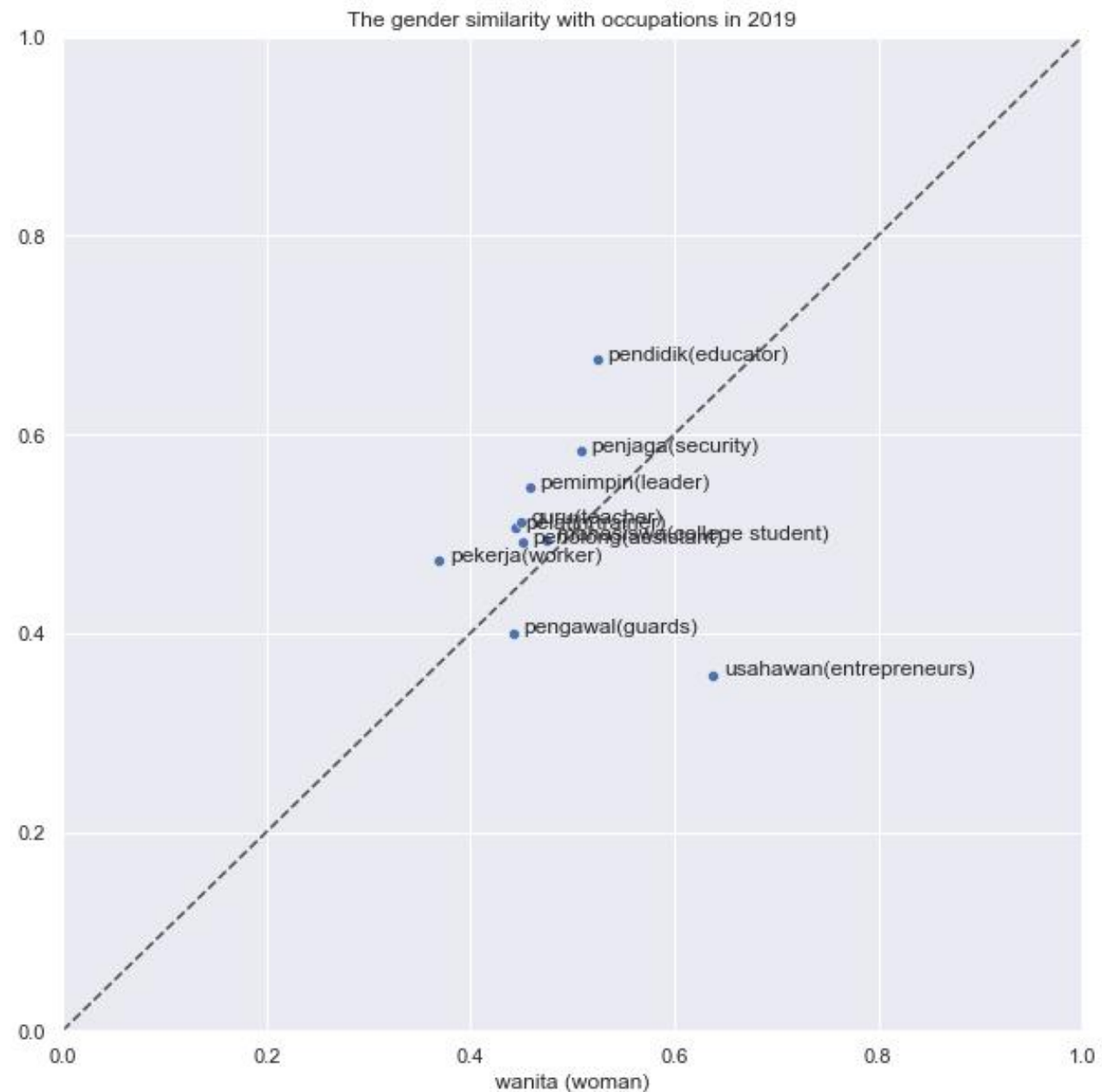
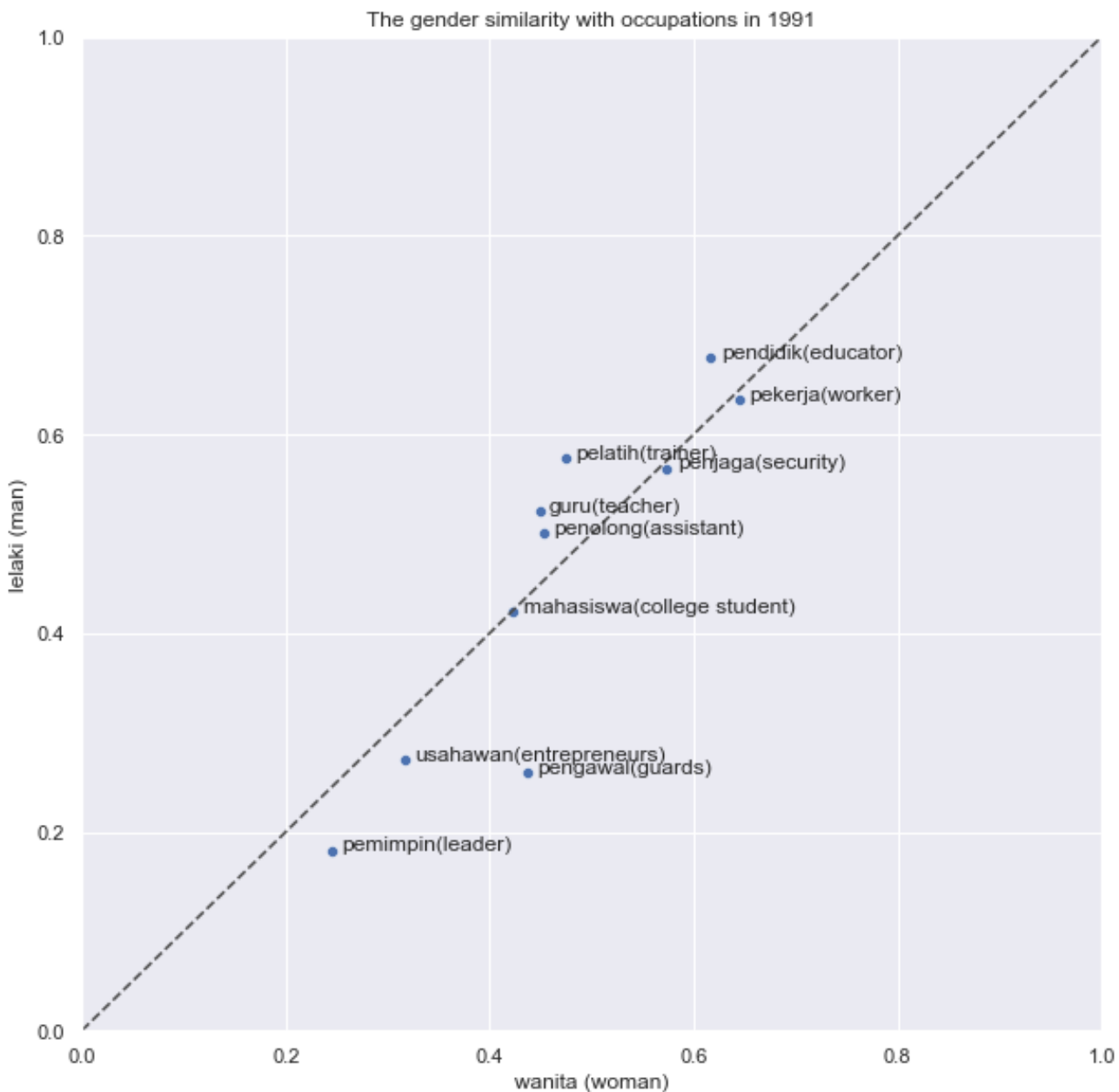
Words	
Adjectives*	lemah (weak), cacat (handicap), lembut (soft), sasar (crazy), teraniaya (humiliated), muda(young), berkelayakan (qualified), berkemahiran (skilled), berdosa (sinful), professional (professional), berkebolehan (capable), berkahwin (married), berkaliber (caliber), cermat (careful), berdikari (independent), mahir (skilled), berdisiplin (disciplined), gila (crazy)
Occupations	penjaga (security), penolong (assistant), pendidik (educator), usahawan (entrepreneurs), pelatih (trainer), mahasiswa (college student), pengawal (guards), pekerja (worker), guru (teacher), pemimpin (leader)

** Manually selected from the resulting list of pos-tagging is 4,260 (including duplicates)*

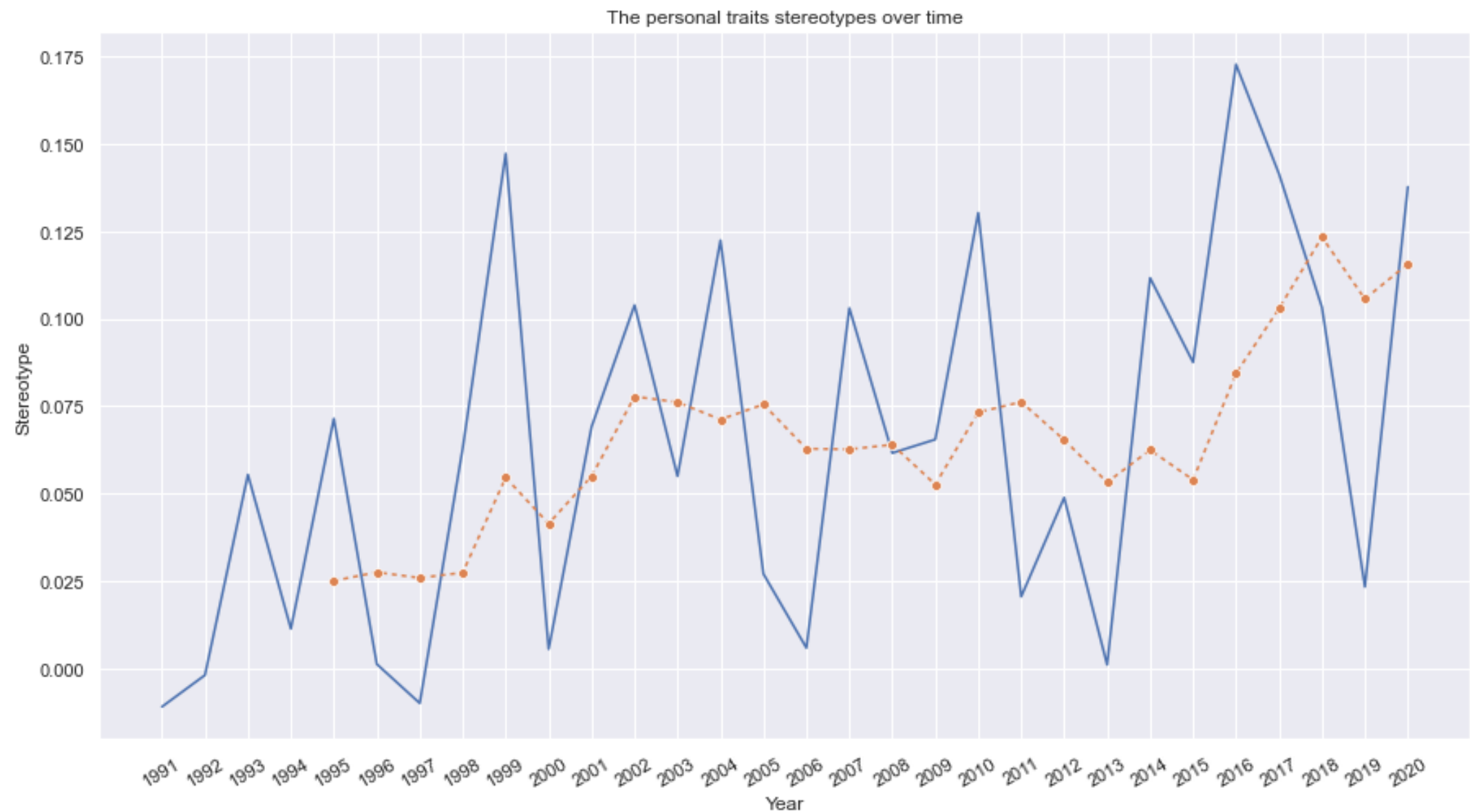
Personal trait similarity



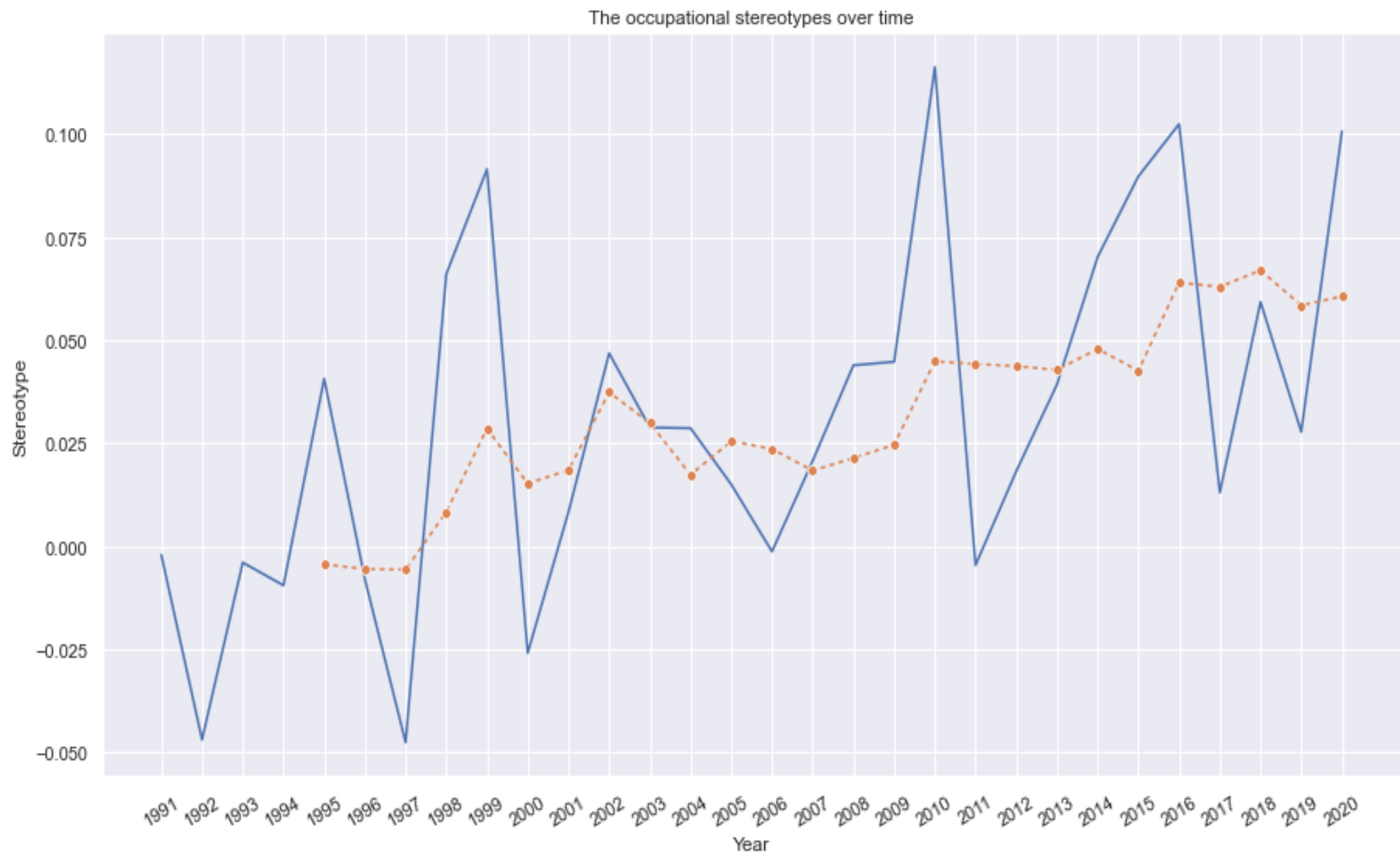
Occupation similarity



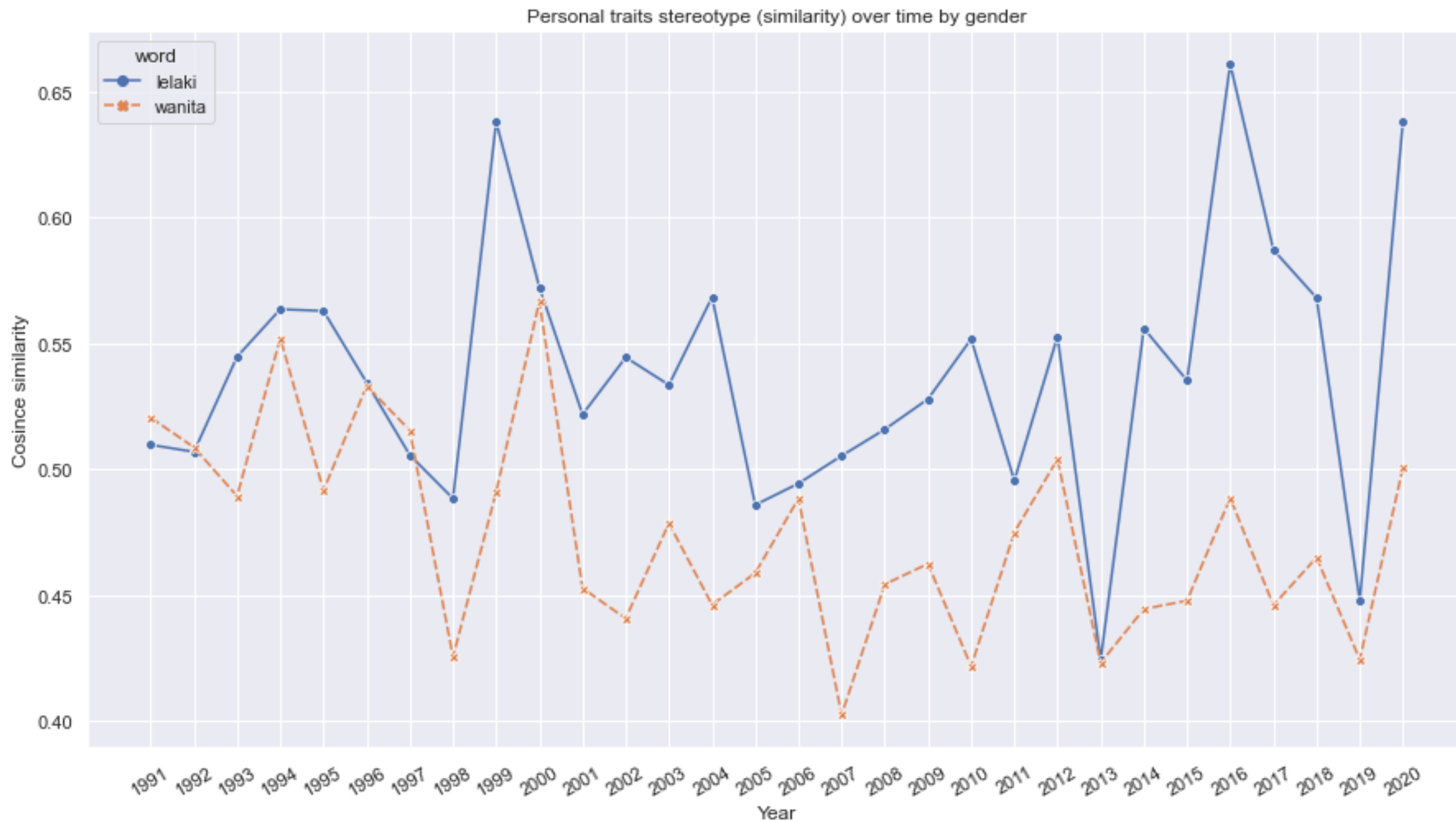
Stereotypes over the years



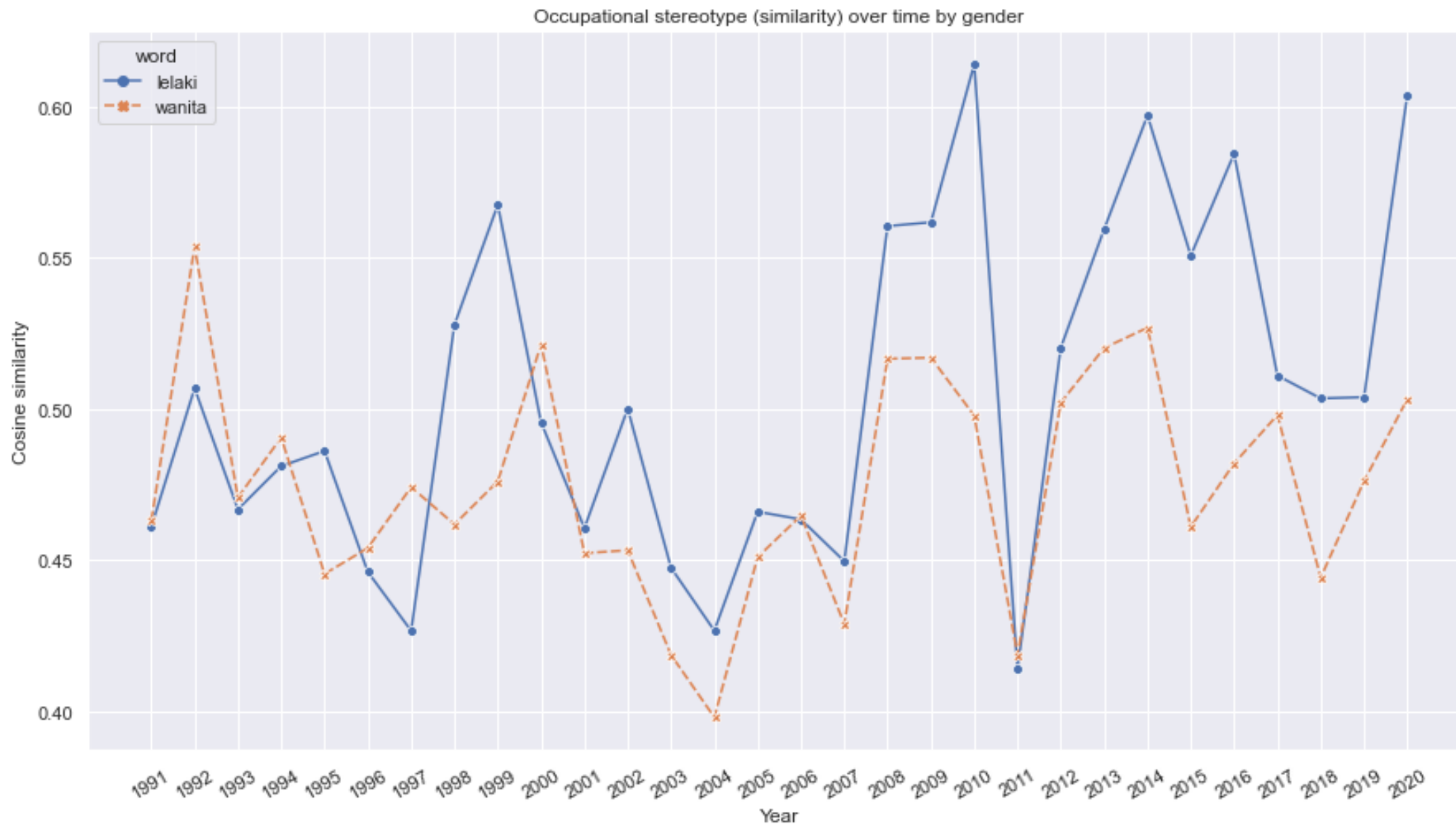
Stereotypes over the years



Validation: stereotype by gender

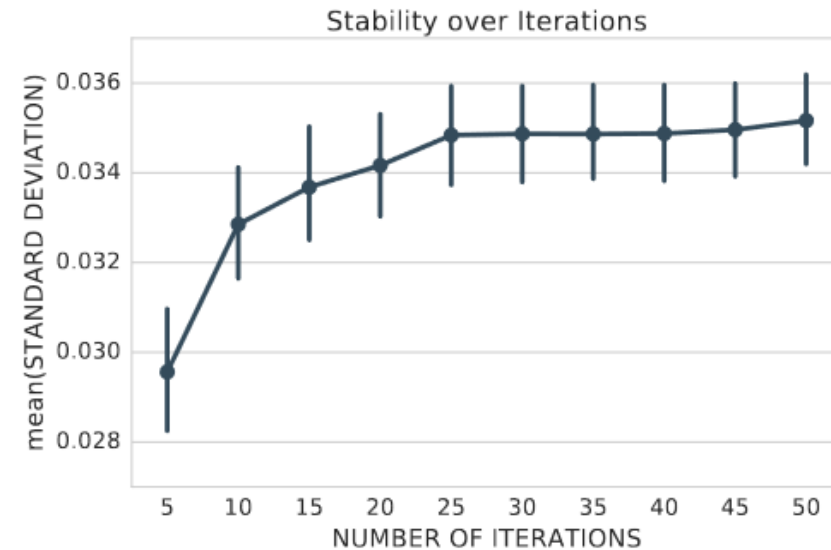
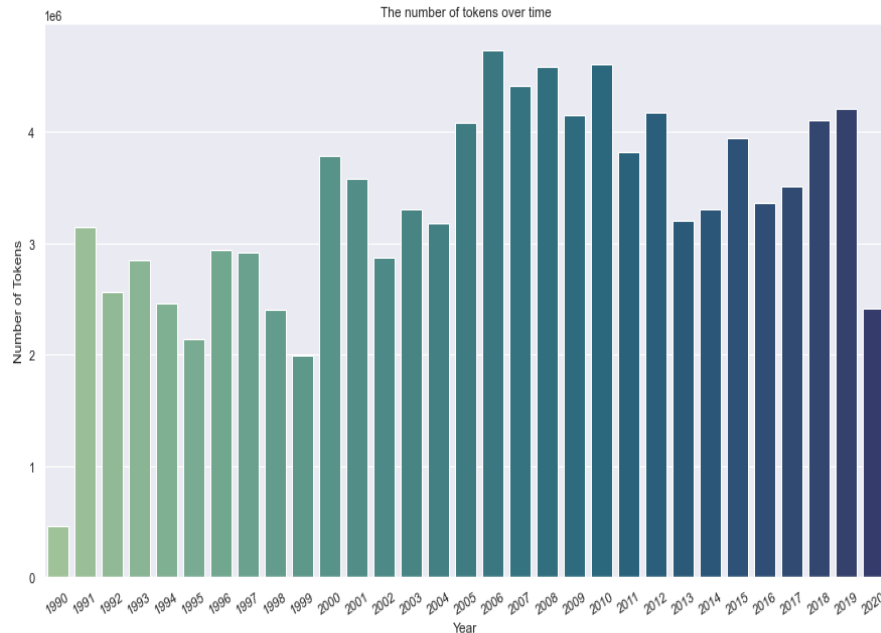


Validation: stereotype by gender



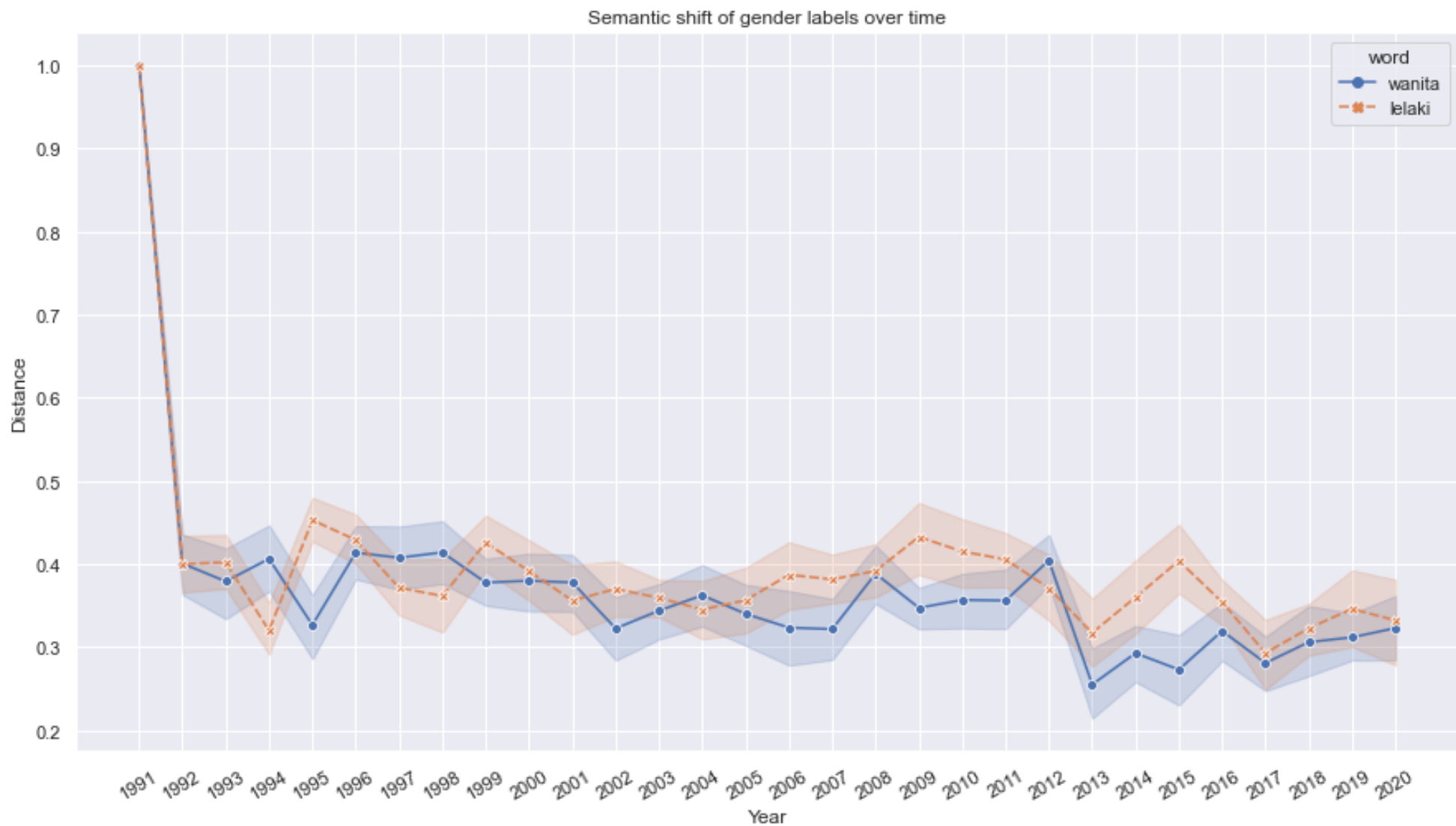
Limitations of word embeddings

- Comparability of word embeddings trained on different years of data
 - e.g., different document lengths, randomness in the algorithm
- For distance calculation, even changing the order of documents causes some variations in the mean cosine similarities and standard deviations (Antoniak and Mimno, 2018)
- A more robust approach is to average over multiple bootstrap samples, ideally 25 or above

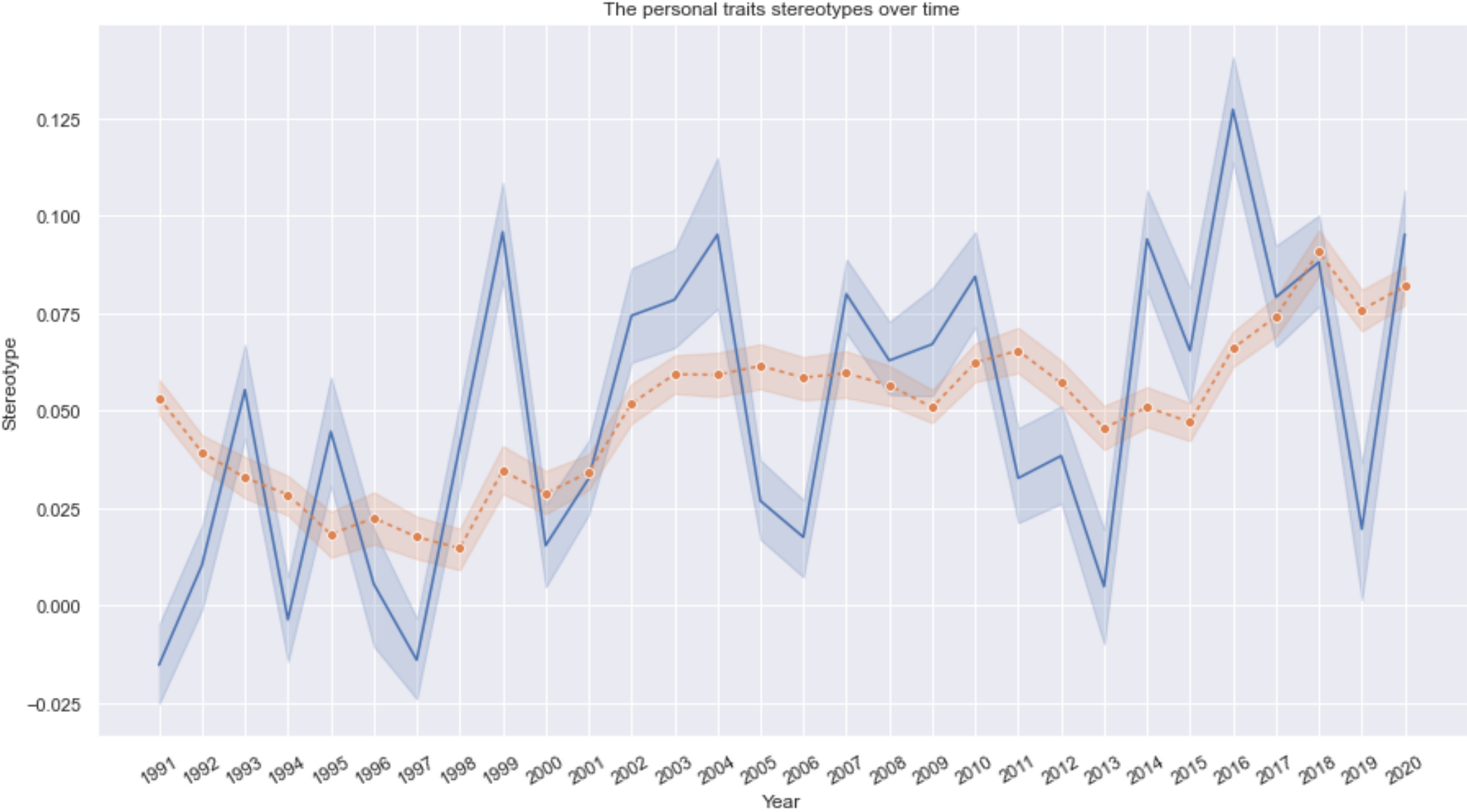


Antoniak and Mimno (2018)

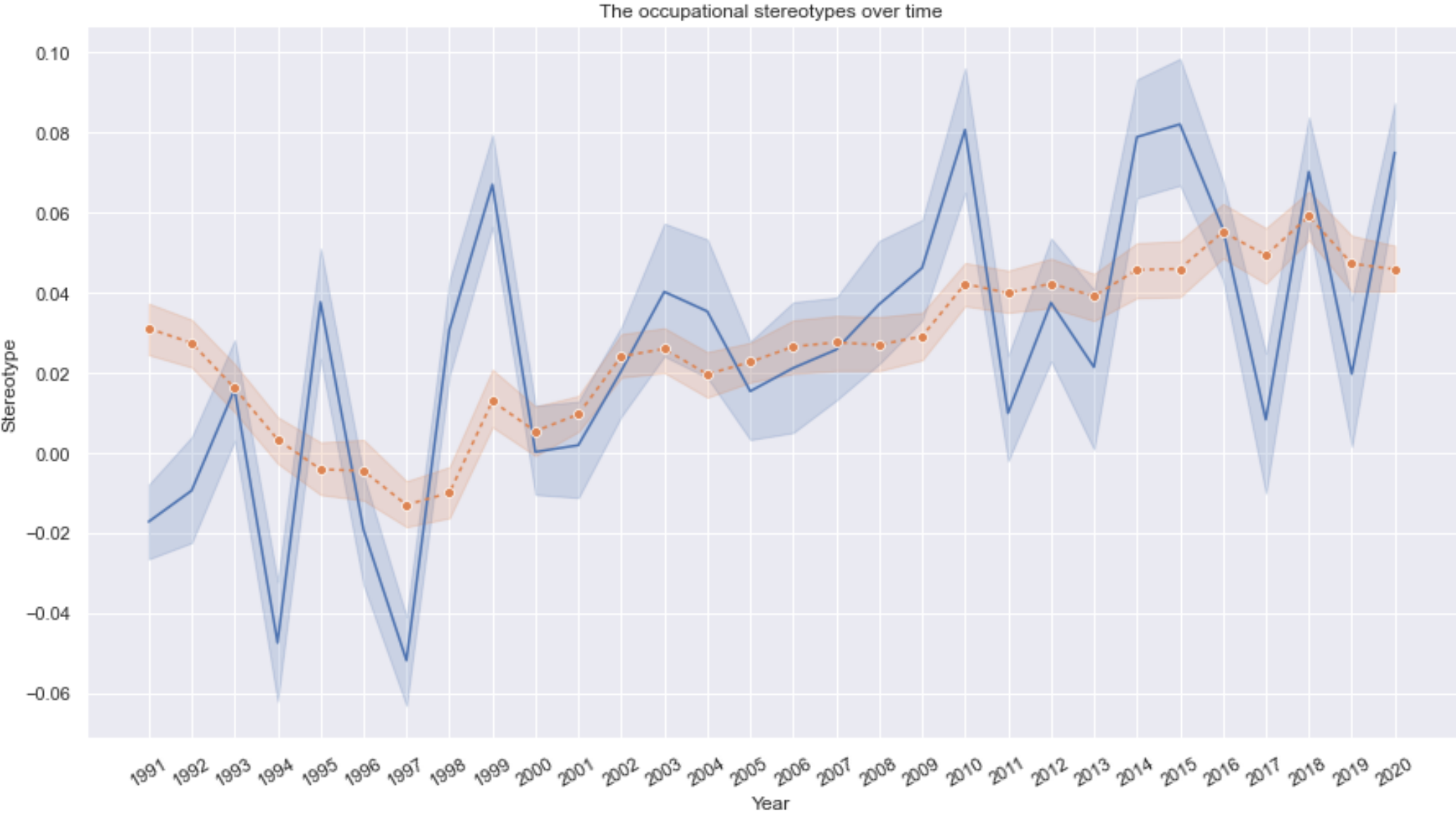
Validation: Semantic shift



Stereotypes with 25 bootstrap samples



Stereotypes with 25 bootstrap samples



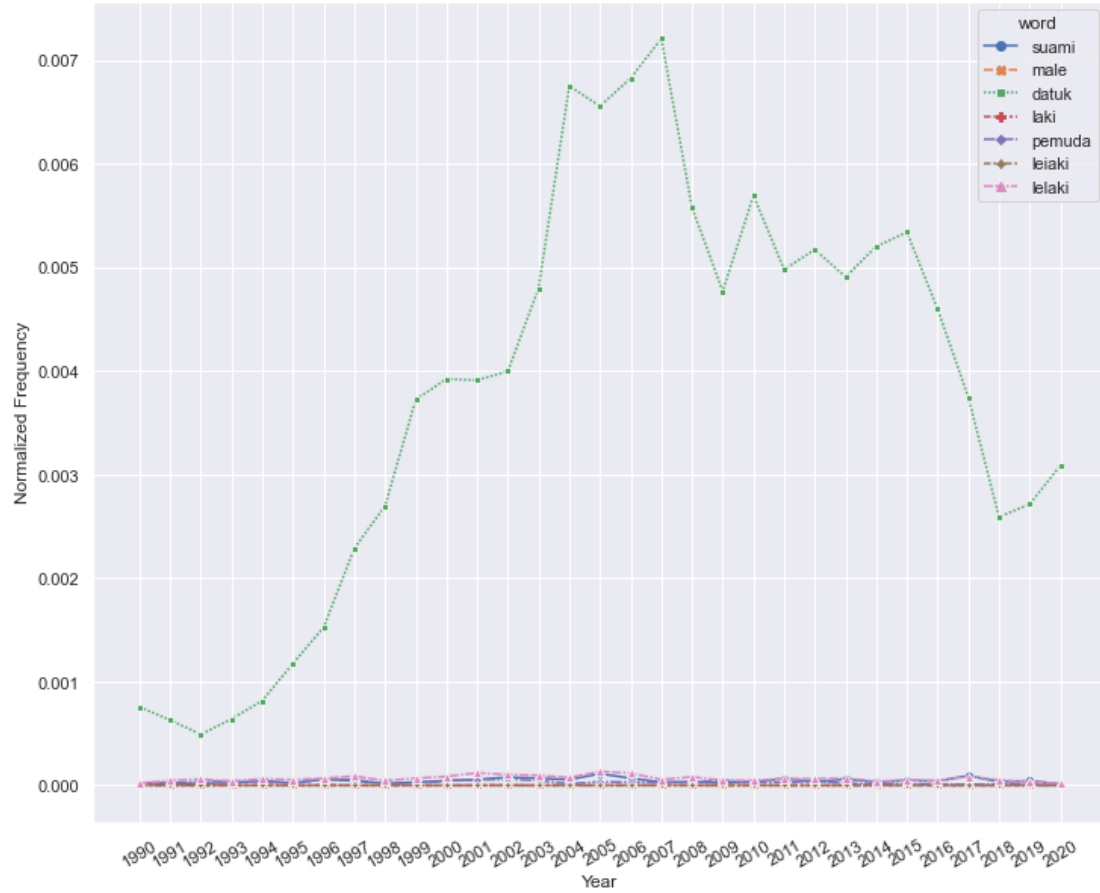
Conclusion

- More attention has been given to women after the set up of Ministry of Women in 2001 without any noticeable change in the perception (semantic meaning)
- An improvement in the stereotypes against women in terms of personal traits and occupations
- Increasing distances indicate these stereotypes are getting less connected to women
- Many potential avenues
 - Interaction between women's stereotypes and the number/proportion of female MPs

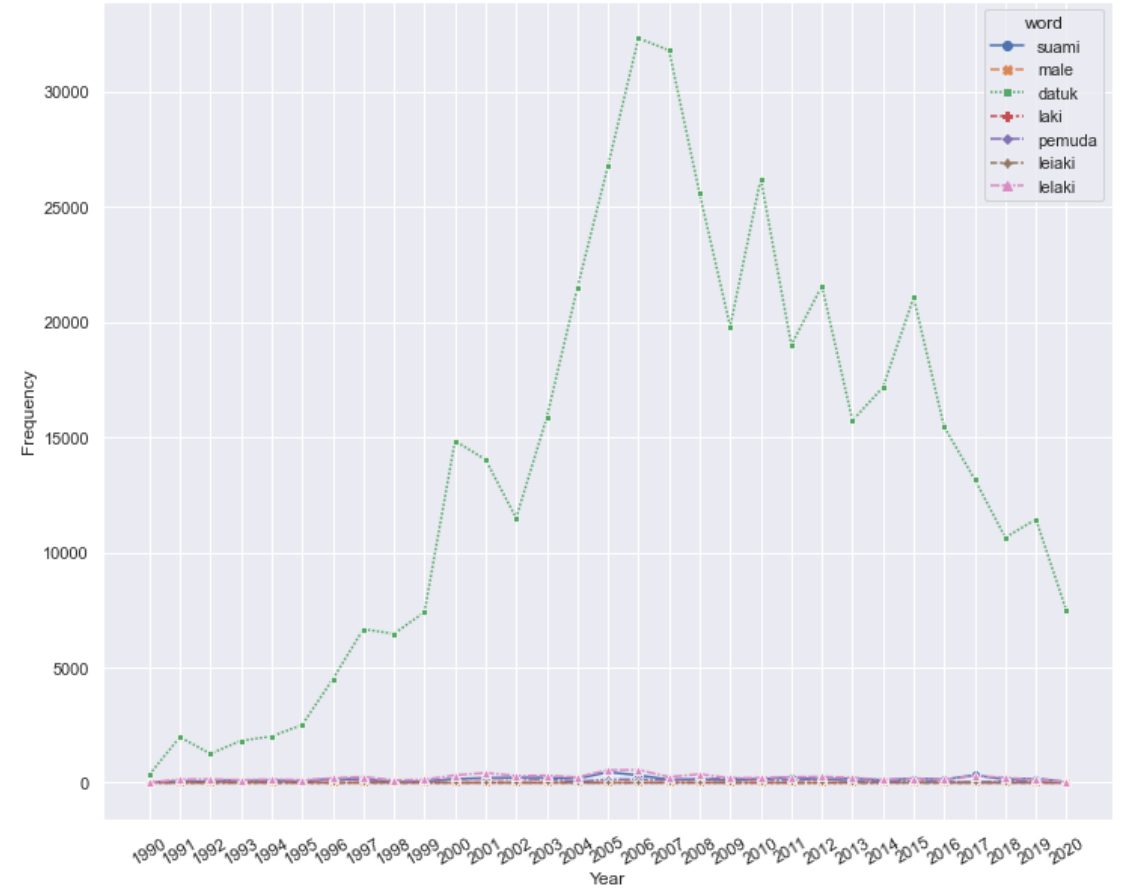
Data description

Parliament	Period	Number of Hansards	Number of tokens
8 th Parliament	1990 – 1995	315	11,471,654
9 th Parliament	1995 – 1999	354	12,222,640
10 th Parliament	1999 – 2004	292	13,702,330
11 th Parliament	2004 – 2008	323	16,454,424
12 th Parliament	2008 – 2013	366	21,288,644
13 th Parliament	2013 – 2018	301	18,305,502
14 th Parliament	2018 – Present	173	9,759,033

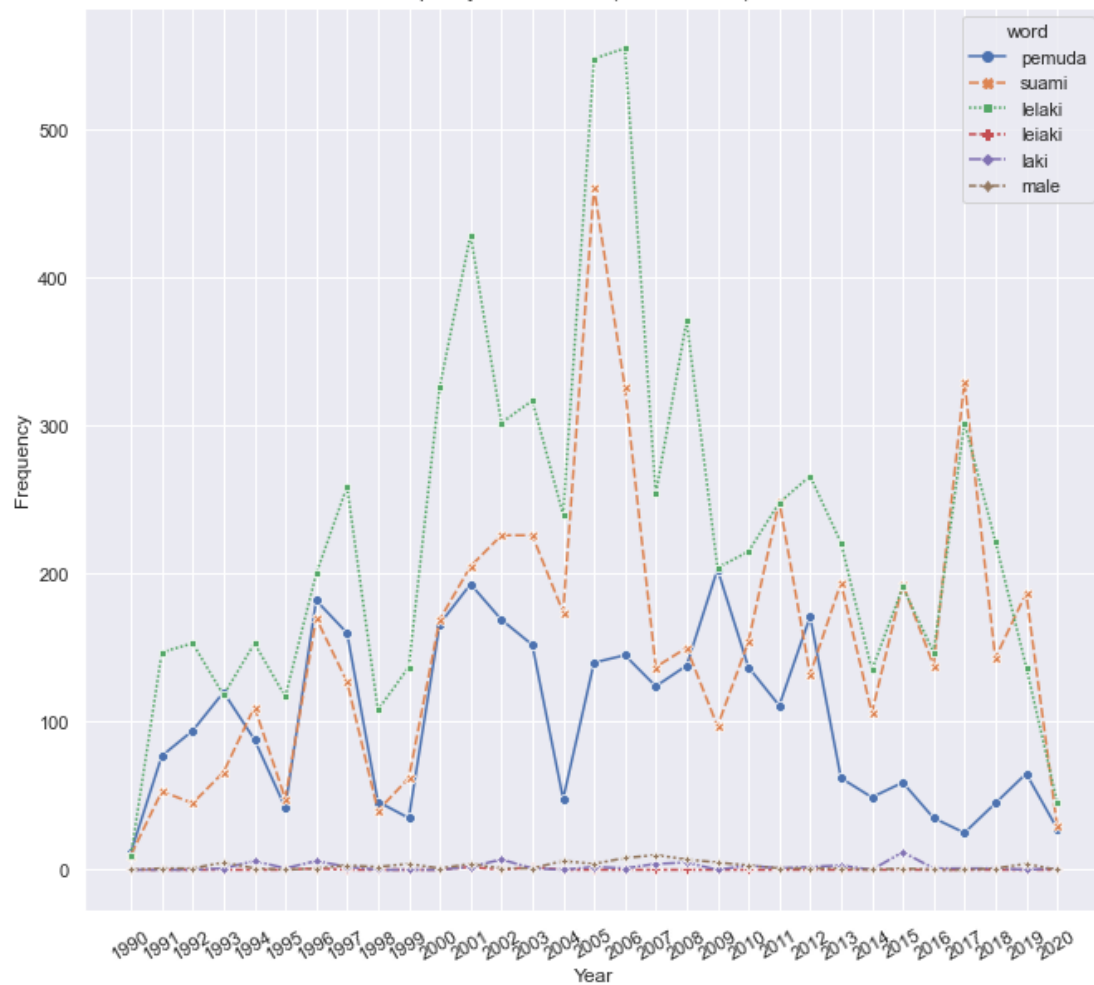
The normalized frequency of male labels over time



The frequency of male labels over time



The frequency of male labels (without 'datuk') over time



The normalized frequency of male labels (without 'datuk') over time

