

Your Name: liping xiong

Your Andrew ID: lipingx

## Homework 2

### 1 question1

#### 1.1 result of full data

Sorry, I didn't have enough time to get the result. I should ran it earlier.

#### 1.2 result of war data

world war	2.198738926212309	2.2120851100041663	-0.013346183791857086
civil war	1.2337463211609394	1.2354436247196132	-0.00169730355867384
war ii	0.966533763447829	0.9863173760121391	-0.01978361256431005
cold war	0.21118920557343865	0.2170153977469689	-0.005826192173530245
korean war	0.13636106877023146	0.13592983879904116	4.312299711903137E-4
war department	0.12842815434421706	0.12476231770575279	0.0036658366384642693
revolutionary war	0.09236431715727837	0.09426394625129954	-0.0018996290940211
war years	0.0821451031651612	0.08107177292850926	0.001073330236651949
nuclear war	0.07548791085995638	0.07818782992151509	-0.0026999190615587085
war effort	0.06923800226628793	0.07046272807727699	-0.0012247258109890572
american war	0.06162352997028235	0.061275772472878264	3.4775749740408805E-4
mexican war	0.058082223039053034	0.05783780183431693	2.4442120473610457E-4
war broke	0.054461366079685616	0.05319443107247734	0.0012669350072082829
war office	0.054273835697977	0.05067335706848027	0.0036004786294967306
declared war	0.05372983916506078	0.05215902811983628	0.0015708110452244959
war period	0.05367526470611614	0.052321360510699355	0.0013539041954167862
great war	0.05135207966499543	0.05069461164333759	6.574680216578338E-4
crimean war	0.048846598997409924	0.047418807091086396	0.0014277919063235263
japanese war	0.04493677130535046	0.04374122411740842	0.0011955471879420434
war ended	0.03799697097501001	0.038433465090523436	-4.364941155134244E-4

### 2 question2

1970-1979	1980-1989	1990-1999
world war	world war	world war
civil war	war ii	war ii
war ii	civil war	civil war
cold war	nuclear war	cold war
vietnam war	vietnam war	gulf war
revolutionary war	cold war	vietnam war
korean war	korean war	korean war
war department	revolutionary war	revolutionary war
war years	war effort	nuclear war
war effort	war department	war effort
american war	war years	war department
mexican war	american war	war years
war period	mexican war	american war
war broke	great war	great war
nuclear war	war ended	mexican war
declared war	war period	war ended
great war	war broke	war era
japanese war	war college	war crimes
crimean war	declared war	war broke
war ended	war office	declared war

1970-1979				1980-1989				1990-1999			
world war	2.32120197	2.328461722	-0.0072598	world war	2.63825237	2.61818039	0.02007198	world war	2.41451063	2.41460048	-8.99E-05
civil war	1.28450336	1.283632454	8.71E-04	war ii	1.43191983	1.4135512	0.01836863	war ii	1.39211471	1.37888564	0.013229074
war ii	1.11090254	1.120864707	-0.0099622	civil war	1.26723346	1.27074308	-0.0035096	civil war	1.33886637	1.33592559	0.002940774
cold war	0.1845341	0.192382982	-0.0078489	nuclear war	0.22275598	0.20567288	0.0170831	cold war	0.51757411	0.48496747	0.032606637
vietnam war	0.14772859	0.147590136	1.38E-04	vietnam war	0.19759666	0.19250673	0.00508993	gulf war	0.25647492	0.19490671	0.061588211
revolutionary war	0.13640935	0.134103844	0.00230551	cold war	0.18546766	0.19348258	-0.0080149	vietnam war	0.22400902	0.21488827	0.009120755
korean war	0.12774369	0.128255666	-5.12E-04	korean war	0.12854266	0.12897889	-4.36E-04	korean war	0.13366499	0.13330472	3.60E-04
war department	0.10696628	0.106081299	8.85E-04	revolutionary war	0.12391123	0.12357736	3.34E-04	revolutionary war	0.11973093	0.12001145	-2.81E-04
war years	0.08660208	0.085198375	0.0014037	war effort	0.09183971	0.09104598	7.94E-04	nuclear war	0.093022	0.09464449	-0.001622483
war effort	0.08300459	0.082974029	3.06E-05	war department	0.08907633	0.09028636	-0.00121	war effort	0.08879427	0.08837496	4.19E-04
american war	0.06912251	0.067917409	0.0012051	war years	0.07907691	0.07892542	1.51E-04	war department	0.08161691	0.08377491	-0.002157996
mexican war	0.06010518	0.059469921	6.35E-04	american war	0.05572547	0.05640163	-6.76E-04	war years	0.06230296	0.06422577	-0.00192281
war period	0.05688307	0.055280546	0.00160253	mexican war	0.05372887	0.05399255	-2.64E-04	american war	0.05643707	0.05708912	-6.52E-04
war broke	0.05118504	0.050543225	6.42E-04	great war	0.04575695	0.04646867	-7.12E-04	great war	0.05340968	0.05329801	1.12E-04
nuclear war	0.04965327	0.053705623	-0.0040524	war ended	0.04269376	0.04281516	-1.21E-04	mexican war	0.0516189	0.05211698	-4.98E-04
declared war	0.04948072	0.048743064	7.38E-04	war period	0.04164922	0.04216373	-5.15E-04	war ended	0.04978704	0.04903534	7.52E-04
great war	0.04869323	0.048765036	-7.18E-05	war broke	0.04044246	0.04099767	-5.55E-04	war era	0.04878412	0.04412279	0.004661328
japanese war	0.04616834	0.044916931	0.00125141	war college	0.0383104	0.03706115	0.00124925	war crimes	0.04201406	0.04051102	0.001503039
crimean war	0.04322096	0.042551722	6.69E-04	declared war	0.03768296	0.0383429	-6.60E-04	war broke	0.03708672	0.03797098	-8.84E-04
war ended	0.04063845	0.040901137	-2.63E-04	war office	0.03477579	0.0344858	2.90E-04	declared war	0.03463785	0.0356403	-0.001002451

We see some trend given by the occurrence of the event. Such as “Gulf war”(1990-1991) got high Informativeness score which indicates it's a novel event. The background corpus is from a time where this event did not happen, this phrase have a very high informativeness score.

However, since the Informativeness score is orders of magnitude smaller than the phraseness score, this trend is not been clearly shown. A score normalization of both scores would make that more clear.

### 3 question3

world war	2.32120197	2.328461722	-0.00725975
civil war	1.28450336	1.283632454	8.71E-04
war ii	1.11090254	1.120864707	-0.009962163
cold war	0.1845341	0.192382982	-0.007848884
vietnam war	0.14772859	0.147590136	1.38E-04
revolutionary war	0.13640935	0.134103844	0.002305507
korean war	0.12774369	0.128255666	-5.12E-04
war department	0.10696628	0.106081299	8.85E-04
war years	0.08660208	0.085198375	0.001403705
war effort	0.08300459	0.082974029	3.06E-05
american war	0.06912251	0.067917409	0.001205105
mexican war	0.06010518	0.059469921	6.35E-04
war period	0.05688307	0.055280546	0.001602529
war broke	0.05118504	0.050543225	6.42E-04
nuclear war	0.04965327	0.053705623	-0.004052356
declared war	0.04948072	0.048743064	7.38E-04
great war	0.04869323	0.048765036	-7.18E-05
japanese war	0.04616834	0.044916931	0.001251411
crimean war	0.04322096	0.042551722	6.69E-04
war ended	0.04063845	0.040901137	-2.63E-04

In the top20 phrases from the war dataset, there are several phrases that are high ranked and should not be, such as “war effort”, “war department”. They are very high ranked because they have a high phraseness score. Two words in these phrases co-occur very often, but they are not very meaning as a phrase(low informativeness score).

This is because phraseness score is much larger than informativeness score that phraseness score dominates the final score. So it's important to scale both scores before adding them up.

### 4 question4

a)

$X=\text{toast} \wedge Y=\text{breakfast}$  3

$X=\text{likes} \wedge Y=\text{breakfast}$  2

$X=\text{likes} \wedge Y=\text{dinner}$  2

$X=\text{steak} \wedge Y=\text{dinner}$  2

b)

Toast  $C[w \wedge Y=\text{breakfast}] = 3$

Likes  $C[w \wedge Y=\text{breakfast}] = 2$ ,  $C[w \wedge Y=\text{dinner}] = 2$

Steak  $C[w \wedge Y=\text{dinner}] = 2$

c)

Jane // ~ctr to id1  
ordered // ~ctr to id1  
eggs // ~ctr to id1  
and // ~ctr to id1  
toast // ~ctr to id1

d)

id1 ~ctr for Jane is  $C[w^{\wedge} Y = \text{breakfast}] = 1$   
id1 ~ctr for and is  $C[w^{\wedge} Y = \text{breakfast}] = 1, C[w^{\wedge} Y = \text{dinner}] = 1$   
id1 ~ctr for toast is  $C[w^{\wedge} Y = \text{breakfast}] = 3$

e)

id1 Jane ordered eggs and toast  
~ctr for Jane is  $C[w^{\wedge} Y = \text{breakfast}] = 1$   
~ctr for and is  $C[w^{\wedge} Y = \text{breakfast}] = 1, C[w^{\wedge} Y = \text{dinner}] = 1$   
  
~ctr for toast is  $C[w^{\wedge} Y = \text{breakfast}] = 3$

## Part 2

- a.  $K * V$
- b.  $K * V$
- c.  $K * V$
- d.  $V$
- e.  $N$
- f.  $V + N$
- g.  $N$

## 5 Question 5

I asked liangke gui and qiang zhu some questions about the meaning of questions. But not about the solutions.