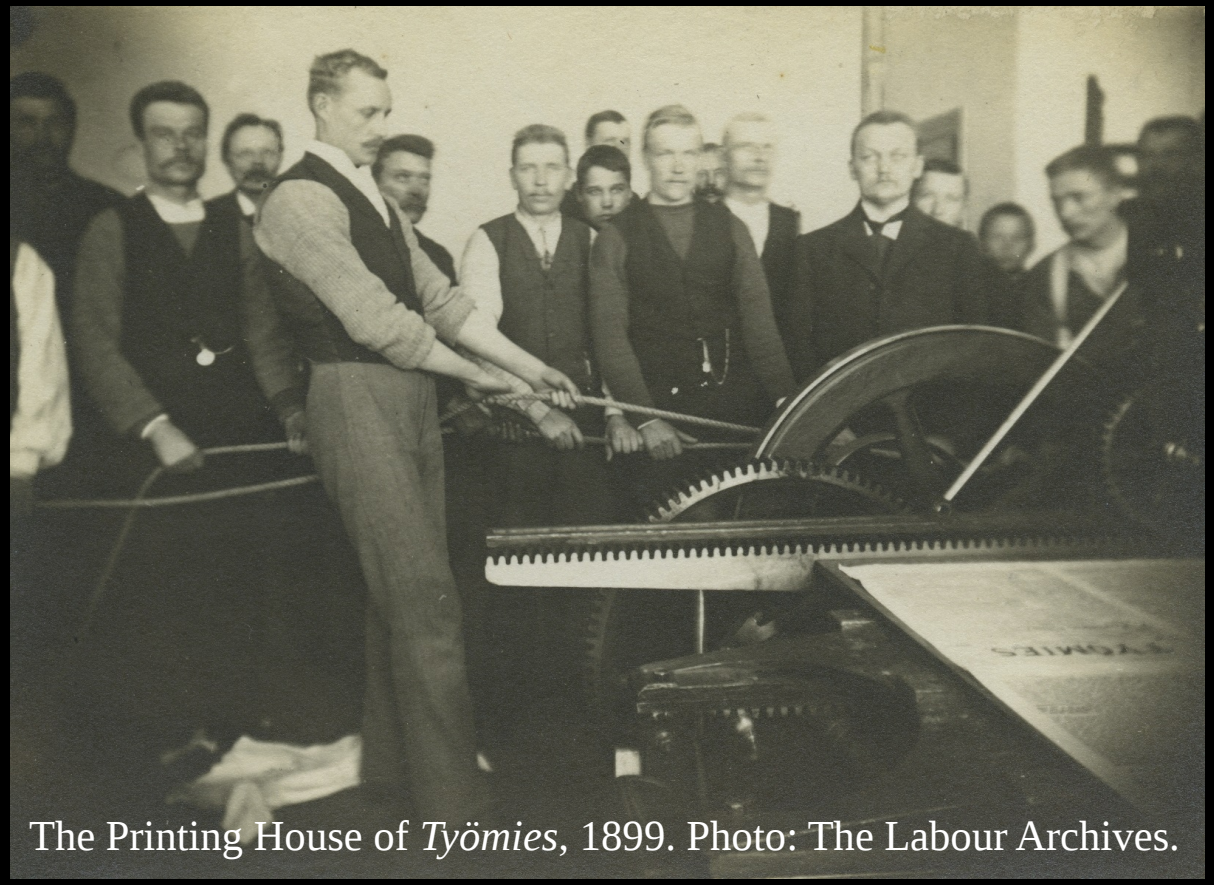
Sculpting Time:
Temporality in the Political Language of Finnish Socialism, 1900–1917



The Printing House of *Työmies*, 1899. Photo: The Labour Archives.

Risto Turunen
risto.turunen@uta.fi
School of Social Sciences
University of Tampere

QUESTIONS

I) What are the key differences between socialist temporality and non-socialist temporalities in the early twentieth century?

II) What kinds of meaning do socialist attach to the past, present, and future?

III) Do the actual revolutionary events in Finland – such as the General Strike of 1905 – change the socialist perception of time in any way?

DATA

Digitized Finnish Newspapers, 1900–1917

- Downloaded txt.files from the National Library.
- Lemmatized with the LAS command-line tool.
- Political affiliation NOT in the metadata → determined using earlier research on Finnish newspapers (Tommila 1987).
- Today's presentation concentrates on the newspapers in Helsinki: socialist *Työmies*, liberal-nationalist *Helsingin Sanomat* (*Päivälehti* until 3.7.1904), conservative-nationalist *Uusi Suometar*, and Christian *Kotimaa*.

METHOD I

Keyness method ≈
which words are used more frequently than expected by chance
in socialist texts compared to non-socialist texts?

Subjective classification ≈
which of these socialist key words are connected to temporality?

Some temporal words that appear more frequently than expected in socialist *Työmies* in 1906-1910, compared to three non-socialist newspapers in Helsinki

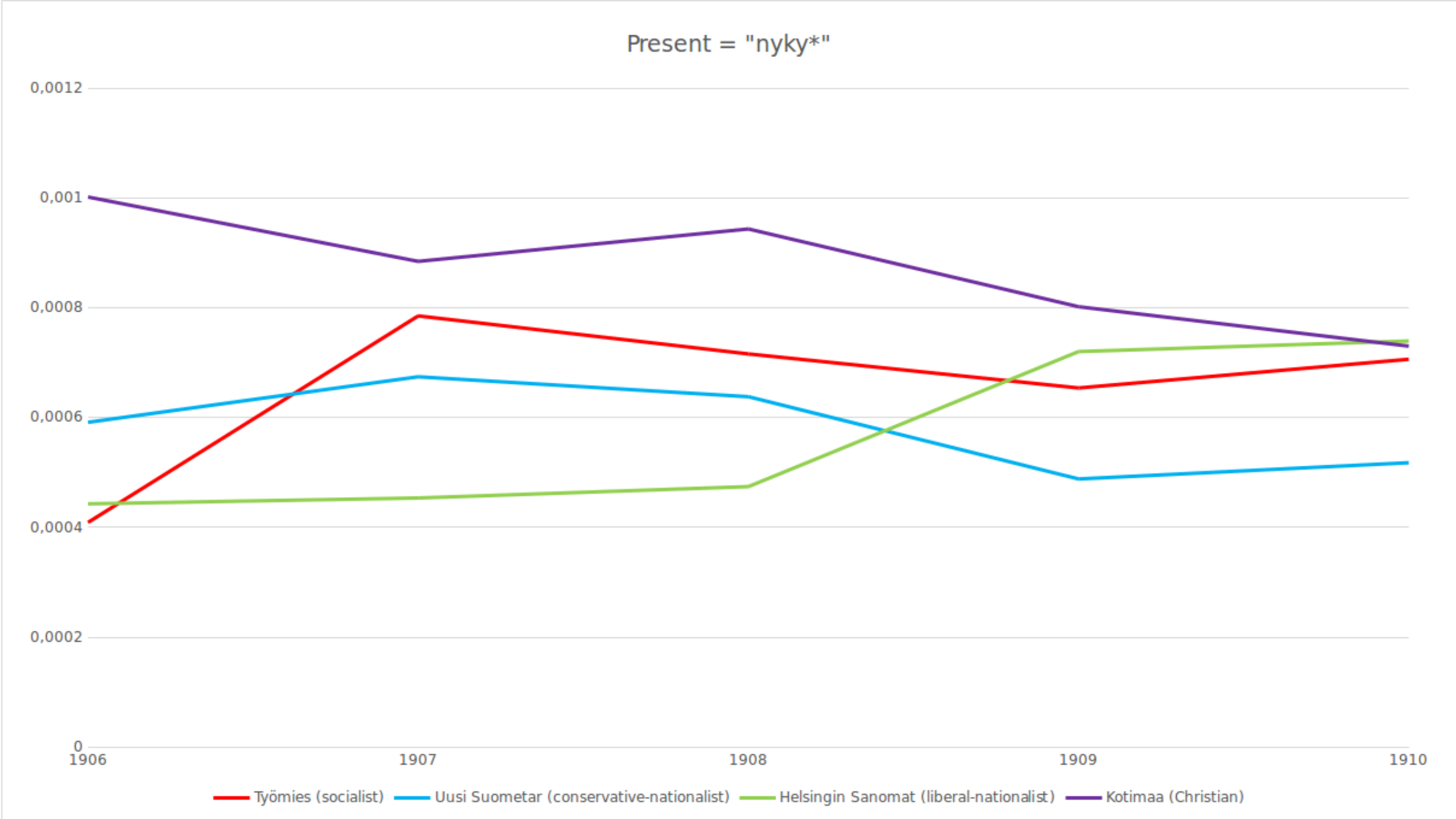
WORD	TRANSLATION	KEYNESS VALUE	ABSOLUTE HITS
nykyinen	present (adjective)	324,54	8013
nykyään	nowadays	152,94	5168
nykynen	present (adjective)	98,38	49
nykyisin	nowadays	42,47	1531
nykyisellään	nowadays	16,52	64

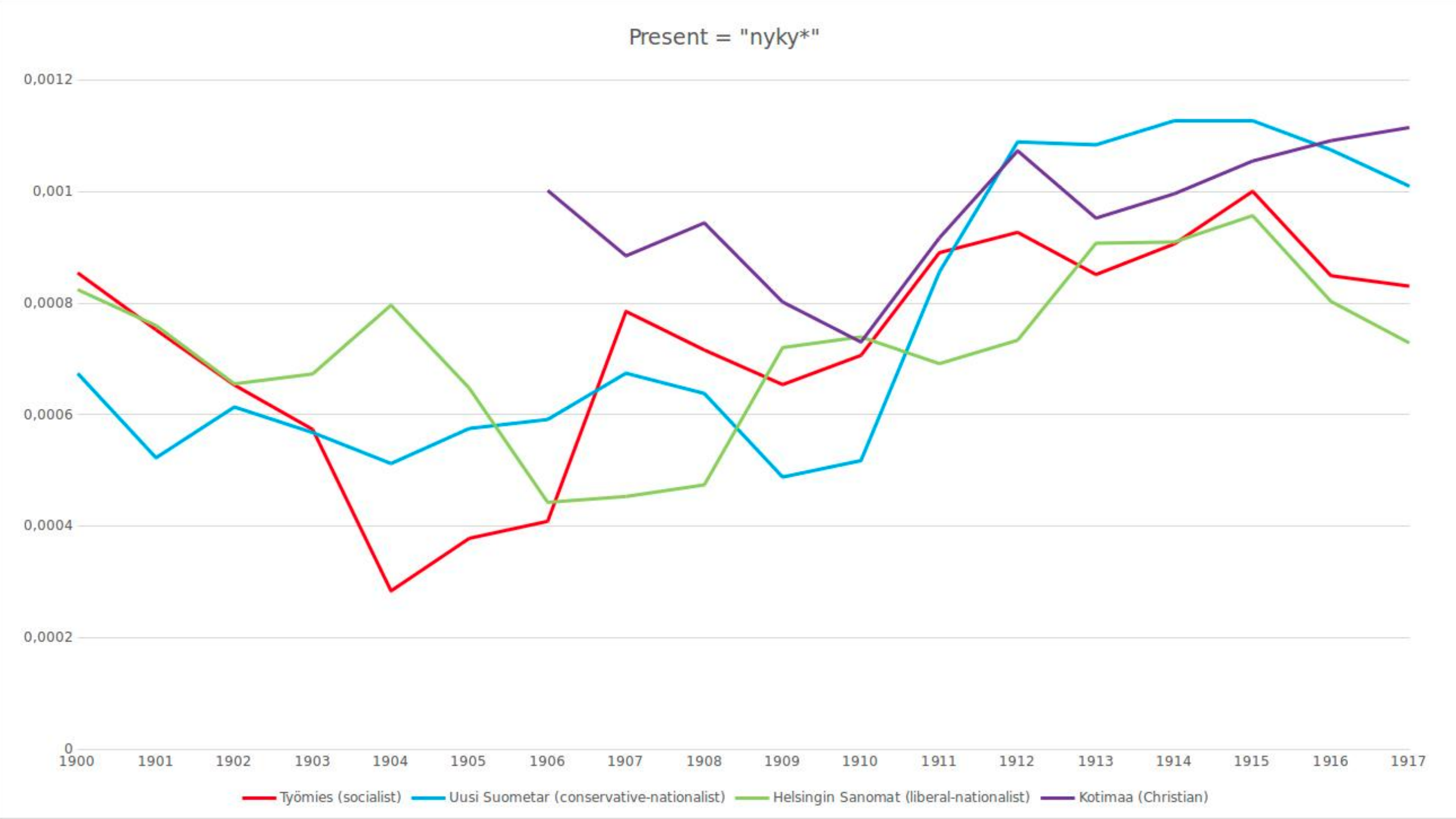
Vaasa: *Vapaa Sana* (socialist) versus *Vaasa* (conservative-nationalist), 1906-1910

WORD	TRANSLATION	KEYNESS VALUE	ABSOLUTE HITS
nykyinen	present (adjective)	25,83	1993
nykynen	present (adjective)	20,93	17
nykyjään	nowadays	19,28	30

Kuopio: *Savon Työmies* (socialist) versus *Otava* (liberal-nationalist), 1906-1910

WORD	TRANSLATION	KEYNESS VALUE	ABSOLUTE HITS
nykyinen	present (adjective)	211,32	1899
nykyisellään	nowadays	27,67	27
nykyisin	nowadays	19,66	391


Present = "nyky*"
0,0012
0,001
0,0008
0,0006
0,0004
0,0002
0
1906
1907
1908
1909
1910
Työmies (socialist)
Uusi Suometar (conservative-nationalist)
Helsingin Sanomat (liberal-nationalist)
Kotimaa (Christian)


Present = "nyky*"
0,0012
0,001
0,0008
0,0006
0,0004
0,0002
0
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
Työmies (socialist)
Uusi Suometar (conservative-nationalist)
Helsingin Sanomat (liberal-nationalist)
Kotimaa (Christian)

METHOD II

Collocation ≈
which words are used more frequently than expected by chance in close proximity to the search word
(five words to the left or right from the present in this case)?

Subjective interpretation ≈
what do the collocates tell about the meaning of the present in the political language of Finnish socialism?

Features of the Socialist Present, Based on Its Collocations in *Työmies*

- The present is *negative*: misery, miserable, to suffer, to threaten, grievance, shortage...
- ...and it is getting even worse: venereal diseases, landless people, unemployment, housing misery, dissatisfaction etc. are *increasing nowadays*.
- The present is a *system* (”järjestelmä”, ”yhteiskuntajärjestelmä”, ”talousjärjestelmä”) that needs to be *explained*: meaning (”merkitys”), to mean (”merkitys”), question (”kysymys”), cause (”syy”), to cause (”aiheuttaa”), ”relation” (suhde).
- The present is also something that can be changed: to fix (”korjata”), to remove (”poistaa”), to improve (”parantaa”), to change (”muuttaa”).

METHOD III

”Key collocation” ≈
all the words appearing in close proximity to the present *after* the General Strike are collected and combined into one mini-corpus, which was then compared with all the words appearing in close proximity to the present *before* the General Strike

Subjective interpretation ≈
what do these key collocates tell about the semantic differences between socialist present before and after the General Strike?

TOP10 key collocates of the present, comparing *Työmies* 7.11.1905-31.12.1907 to *Työmies* 1.1.1904-30.10.1905.

WORD	TRANSLATION	KEYNESS VALUE	ABSOLUTE HITS
oma	OCR-crap?	74,81	686
lakko	strike	62,24	225
moi	OCR-crap?	56,70	207
hallitus	government	42,67	324
nykyinen	present	40,95	3012
maa	nation	37,29	754
duuma	Duma	36,90	85
senaatti	senate	32,03	177
tilaton	landless	29,85	85
keittiö	kitchen	29,12	177

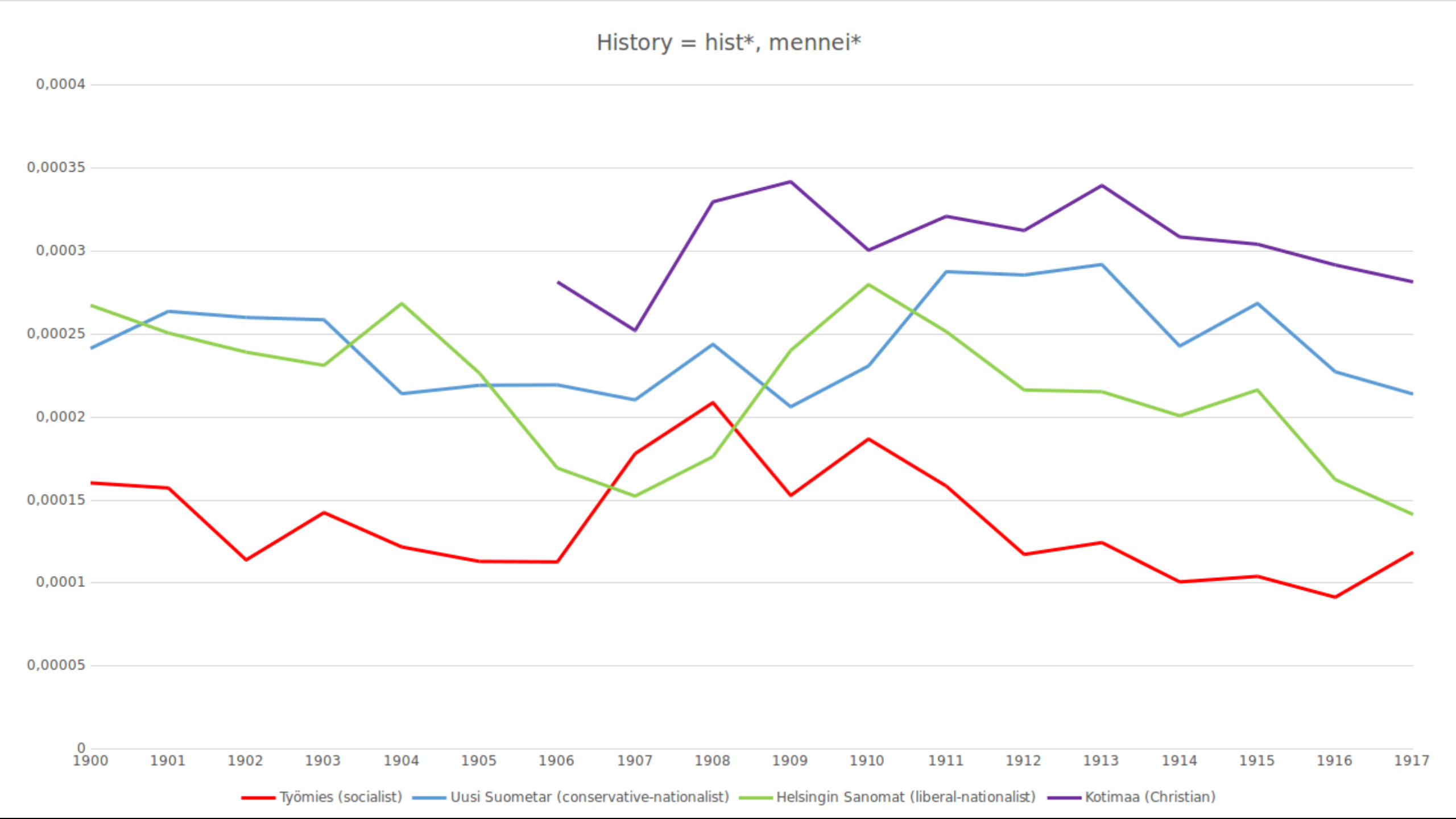
CONCLUSIONS

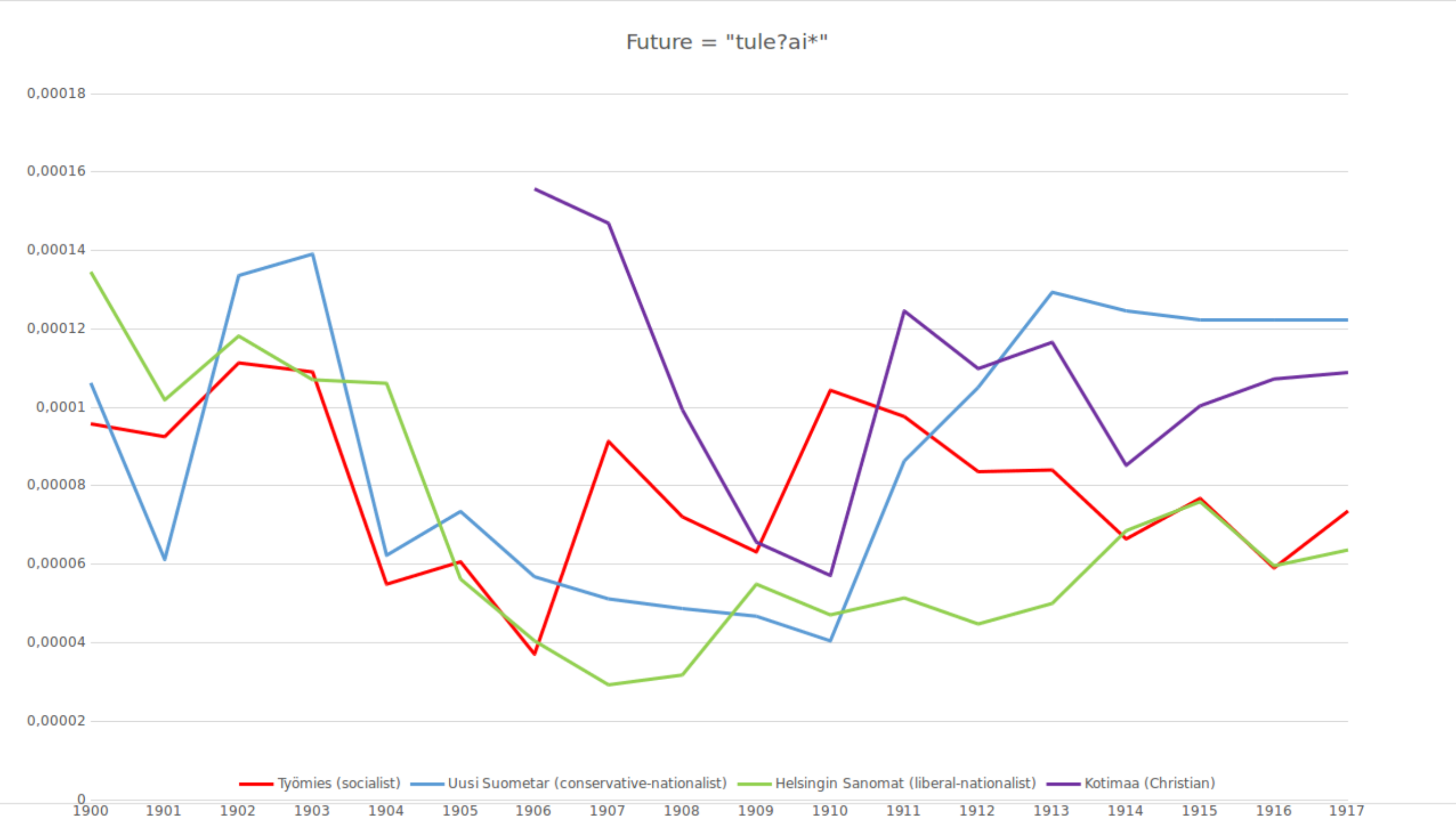
?
?
?
?
?
?
?

Is It Possible to Quantify The Temporality of the Political Language of Finnish Socialism?

- The quantitative amount of the present in the press seems to vary strongly in time.
- The qualitative structure of the socialist present seems not to be affected (as strongly as I expected) by the major political upheavals.
- Primitive methods, ”sculpting time with a sledge hammer”...then again, what are the alternative methods to analyze temporalities?

EXTRAS


History = hist*, mennei*
0,0004
0,00035
0,0003
0,00025
0,0002
0,00015
0,0001
0,00005
0
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
Työmies (socialist)
Uusi Suometar (conservative-nationalist)
Helsingin Sanomat (liberal-nationalist)
Kotimaa (Christian)


Future = "tule?ai*"
0,00018
0,00016
0,00014
0,00012
0,0001
0,00008
0,00006
0,00004
0,00002
0
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
Työmies (socialist)
Uusi Suometar (conservative-nationalist)
Helsingin Sanomat (liberal-nationalist)
Kotimaa (Christian)

How to calculate log likelihood

Log likelihood is calculated by constructing a contingency table as follows:

	Corpus 1	Corpus 2	Total
Frequency of word	a	b	a+b
Frequency of other words	c-a	d-b	c+d-a-b
Total	c	d	c+d

Note that the value 'c' corresponds to the number of words in corpus one, and 'd' corresponds to the number of words in corpus two (N values). The values 'a' and 'b' are called the observed values (O), whereas we need to calculate the expected values (E) according to the following formula:

$$E_i = \frac{N_i \sum_i O_i}{\sum_i N_i}$$

In our case N1 = c, and N2 = d. So, for this word, E1 = c*(a+b) / (c+d) and E2 = d*(a+b) / (c+d). The calculation for the expected values takes account of the size of the two corpora, so we do not need to normalize the figures before applying the formula. We can then calculate the log-likelihood value according to this formula:

$$-2\ln \lambda = 2\sum_i O_i \ln\left(\frac{O_i}{E_i}\right)$$

This equates to calculating log-likelihood G2 as follows: G2 = 2*((a*ln (a/E1)) + (b*ln (b/E2)))

Note 1: (thanks to Stefan Th. Gries) The form of the log-likelihood calculation that I use comes from the Read and Cressie research cited in Rayson and Garside (2000) rather than the form derived in Dunning (1993).

Note 2: (thanks to Chris Brew) To form the log-likelihood, we calculate the sum over terms of the form x*ln(x/E). For strictly positive x it is easy to compute these terms, while if x is zero ln(x/E) will be negative infinity. However the limit of x*ln(x) as x goes to zero is still zero, so when summing we can just ignore cells where x = 0. Calculating ln(0) returns an error in, for example, MSExcel and the C-maths library.

The higher the G2 value, the more significant is the difference between two frequency scores. For these tables, a G2 of 3.8 or higher is significant at the level of $p < 0.05$ and a G2 of 6.6 or higher is significant at $p < 0.01$.

- 95th percentile; 5% level; $p < 0.05$; critical value = 3.84
- 99th percentile; 1% level; $p < 0.01$; critical value = 6.63
- 99.9th percentile; 0.1% level; $p < 0.001$; critical value = 10.83
- 99.99th percentile; 0.01% level; $p < 0.0001$; critical value = 15.13

Check It:
https://github.com/rt80119/dhn2018