Macroscoping the Sun of Socialism

Distant Readings of Temporality in Finnish Labour Newspapers, 1895–1917

Risto Turunen

Introduction

> But in spite of all indifference, the sun of socialism has cast its first rays there. Even there, great and clear thoughts on the injustices of the present system are silently smouldering.[1]

The optimistic quote above was written in 1903, by a labour journalist outlining the preconditions of socialism in the eastern periphery of the Grand Duchy of Finland. Characteristic of the socialist discourse of the time, he used the phrase 'the sun of socialism'. It was one of the most important symbols of the Finnish labour movement in the early 20th century, figuring not only in newspapers, but also in poetry and red banners. Without doubt, there was something in the red sun of socialism that captured the contemporary proletarian imagination.

Many studies in social and cultural history have proven that symbols acting as 'simplified objectifications of ideologies' play a crucial role in the making

of political movements.[2] The sun is the starting point for this chapter, for we believe that this symbol carries rich temporal information from a century ago. Thus, it can be used as a symbolic key to unlock socialist perceptions of the imagined past, present and future. The breakthrough of Finnish socialism has been analysed from a variety of perspectives,[3] but the focus has not been on 'temporality', that is, the way human beings experience time. There are some occasional comments on the socialist temporality in the previous research, mainly concentrating on the Marxian interpretation of history or on the future expectations in the form of socialist utopianism and eschatology.[4] The third dimension of time, the present, has largely escaped scholarly attention. For example, the sun of socialism has been seen in the context of the future, as a symbol for a better tomorrow and freedom.[5] The future-oriented meaning certainly existed, but we can add more interpretative depth to the investigation of the sun by also including the present in our analysis.

According to Reinhart Koselleck's thesis on temporality, the emergence of modernity, especially the unexpected rupture of the French Revolution of 1789, diminished the value of experience in forecasting the future.[6] While Koselleck's argument concerned the German-speaking world, we argue that the General Strike of 1905 had a similar effect in the Finnish context, expanding the gap between the experiences (of the past) and the expectations (towards the future) and, simultaneously, creating a new understanding of the present. The General Strike from 30 October to 6 November in 1905 was not only a direct result but rather an active extension of the 1905 Russian Revolution to the Grand Duchy Finland.[7] For the first time in Finnish history, workers momentarily seized a great part of political power, and this brief moment, a mere one week of imagined proletarian rule, meant that neither the old rules of politics nor old temporalities applied to the new situation. The General Strike led to a set of parliamentary reforms and to universal suffrage in 1906, and finally in 1907, just four years after the quote at the beginning of this chapter, Finland had the largest socialist party with parliamentary representation in Europe.[8]

This chapter has a threefold goal. First, regarding historical content, it constitutes a case study that tries to decipher the intriguing symbol of the rising sun and, thus, to broaden our understanding of the socialist temporality in Finland. The focus lies on the relation between the sun and the present, or more precisely, on how the sun illuminates the proletarian perception of their reality at the turn of the century. Second, methodologically speaking, we introduce 'macroscopic' approaches that allow historians to see something in the sources that is unavailable to the naked eye.[9] In practice, this means quantifying comparable word frequencies, collocates and key collocates. Third, we describe what it means to write digital history, by sketching a simple theoretical model, which sheds a new light on the intellectual journey the scholar undertakes on her way from original sources to historical wisdom.

Relative Word Frequencies: Counting the Heartbeats of Finnish Politics?

We begin our journey to the core of the socialist sun with an already well-established practice in digital history, that is, counting relative word frequencies over time. First, we download the dataset from the National Library of Finland: the raw text files of the biggest socialist (*Työmies*, '*The Working Man*'), conservative-nationalist (*Uusi Suometar*, '*New Finland*') and liberal-nationalist (*Helsingin Sanomat*, '*Helsinki News*', and before 1904 *Päivälehti*, '*The Daily Paper*') newspapers from 1900 to 1917.[10] Then, we find the words referring to the present in each year by using the search string 'nyky*', which covers the most common Finnish words denoting the present moment: 'nykyinen' ('present' as an adjective), 'nykyisyys' ('present' as a noun) and 'nykyisin' / 'nykyään' (adverbs for the present moment).[11]

Figure 17.1 shows a trend. The socialist newspaper *Työmies* has the highest frequency of 'the present' in 1900, but by the year 1904 the references to
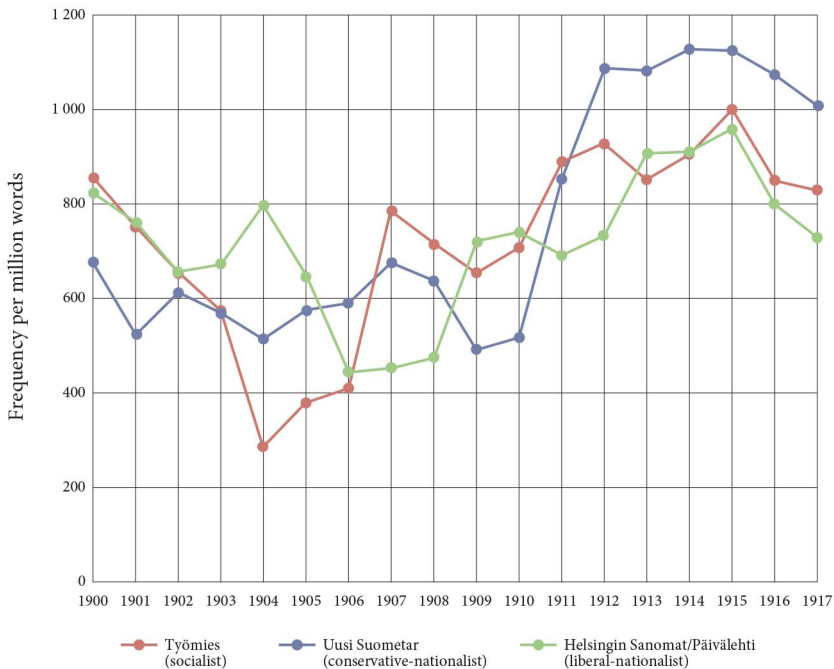

Figure 17.1: The relative amount of the present ('nyky*') in three newspapers, 1900–1917. Source: Raw text files of *Työmies*, *Uusi Suometar*, and *Helsingin Sanomat* / *Päivälehti*, distributed by the National Library of Finland, https://digi.kansalliskirjasto.fi/opendata.

the present have dropped far below the two bourgeois newspapers. What does this trend, this piece of information, mean? Based on previous research, censorship might lead us to the correct explanation. According to Antti Kujala, the censorship of the labour press in Finland became considerably tighter in 1903 when the Finnish Social Democratic Party officially adopted Marxism. Bans and warnings succeeded in silencing the most radical socialist discourse, for during the next year political commentary disappeared 'almost completely' from *Työmies*.[12] However, the General Strike of 1905 broke the silence on political matters as preventive censorship ended temporarily.

The renowned Finnish socialist poet Kössi Kaatra summarised the dramatic new temporality summoned by the strike in his poem *Suurina päivänä* ('*During the Great Days*') from 1906:

> It is great to be alive,
> when in a single day, in a night
> now we create more new things than in the work of many centuries.[13]

Reading Kaatra's words and focusing especially on the temporal marker 'now', it is not a surprise that we see a sharp rise in the socialist present, especially during 1907, which happens to be the year when Finland held the first parliamentary elections. Could the new political situation (electoral speculation, campaigns and aftermath) explain the peak of 1907? Based on both close and distant reading of *Työmies*, this seems to be the case. The words such as 'strike', 'government', 'nation, land', 'Duma' and 'senate' increase greatly in close proximity to the present after the General Strike.[14] Thus, the rise of the present means, in fact, the rise of the *political* present.

We could explain this finding in the light of Benedict Anderson's theory of 'imagined communities' which argues that between 1500–1800 technological innovations and the advance of print-capitalism profoundly changed our experience of time and space.[15] In the case of Finnish working people, these changes probably took place much later, beginning approximately from the mid-19th century onwards.[16] When looking at the date on the front page of the daily socialist newspaper, a Finnish worker could see with her own eyes that time was moving linearly forwards day after day. In addition, she could imagine that *meanwhile* there were thousands of other workers like just like her reading the very same edition, although she had never actually met them.[17] Using Anderson's theory to explain the dissemination of socialism instead of nationalism, as it has usually been applied, guides our analysis towards the close connection between temporalities and print media, or in our case, between the socialist interpretation of the present and the Finnish labour press. Because temporalities are always constructed, they can be manipulated. The leading socialist newspaper reacted to the changing political conditions after the General Strike by accelerating the flow of time, by repeating an imperative temporal message: the time to act is *now*.

One should not forget that there might be other alternative explanations for the peak of 1907. For example, there could have been more adverts in the socialist newspaper in 1907 than before, as the adverts of the early 20th century often referred to the present in order to sell their products better. The lack of information on what constitutes the peaks and valleys of word frequencies is not a trivial problem, but rather characteristic of word frequency charts in digital humanities. Far too often, they neglect variation inside a given corpus. Figure 17.1 is slightly better than the usual combination of a relative word frequency (y-axis) and time (x-axis) in the sense that it contains the extra dimension of political affiliation. However, the figure would be even better if it showed the frequency of 'the present' in different newspaper genres (editorials, foreign section, adverts, poems, letters to the editor, etc.) for each newspaper under investigation. The distribution of genres would show us in which journalistic context the present is discussed in each major political language of the time.

Despite the weaknesses, simple word frequency charts can reveal useful, low-level information to historians. In this case, it revealed above all that the amount of the socialist present varies strongly over time. The valley of 1904 is probably due to censorship, whereas the peak of 1907 is explained by the heated political situation. However, the general trend is that all the newspapers increase their references to the present with the passage of time. Does the trend reflect the increasing heartbeat of Finnish politics, or the rise of present-intensive advertising in all newspapers, or perhaps something completely different? We do not want to get entangled in that question in the context of this chapter, but we do want to highlight the importance of keeping an open mind when attaching meanings to the figures. As doctors know, if the heart beats faster than normal, the possible causes are many and varied.

Collocation: Mining the Semantic Structure of the Socialist Present

Historians inspired by conceptual history, discourse studies or the Cambridge school of intellectual history have for long been interested in the linguistic contexts in which their historical objects of interest (concepts, discourses, intellects) figure.[18] Nowadays, it is possible to quantify such linguistic contexts, given that the textual sources are in a machine-readable form. One approach to operationalise 'the linguistic context' is to define the context as all the words appearing in a window of x words to the left or right of the studied word. Since we are dealing with a highly inflected language, Finnish, it is important to lemmatise all the words in the text files, that is, to replace all word variations with their base form, before the actual analysis in order to get more reliable results.[19] In our case study, we could quantify all the words that exist in proximity of five words from the words referring to the present in the three biggest socialist newspapers. Why five words? There is no magic formula for defining the

Table 17.1: The most frequent words in a window of five words from the 'present' (search strings 'nykyi*' and 'nykyä*') in three socialist newspapers, 1895–1917.

RANK	WORD	TRANSLATION	FREQUENCY
1	olla	to be	60,941
2	ja	and	25,893
3	se	it, that	19,693
4	että	that	14,741
5	joka	which	14,673
6	ei	no	12,174
7	tämä	this	8,618
8	ne	they	6,348
9	kun	when	4,779
10	saada	to get	4,721

Source: Lemmatised raw text files of *Työmies*, *Kansan Lehti* and *Länsisuomen Työmies / Sosialisti*, 1895–1917, distributed by the National Library of Finland, https://digi.kansalliskirjasto.fi/opendata.

perfect window size. Historians must decide, usually through trial and error, which is the most appropriate selection for their own research questions.

As we can see in Table 17.1, the problem with this approach is that the most frequent words connected to the present are common words which do not reveal anything relevant from the historian's perspective. Fortunately, corpus linguists have developed a statistically more sophisticated method in collocation analysis that produces more meaningful raw information for historians to contemplate. Collocates are words that appear *more frequently than statistically expected* in close proximity to the search word.[20]

When looking at Table 17.2, one can immediately see that it contains useful information inviting a further human analysis. After close reading of the concordances, the list of examples of 'the present' as they occur in the socialist newspaper texts, we found three categories of technical errors: some collocates had, not surprisingly, OCR errors; some were lemmatised into a wrong base form; and some suffered from an incorrect word segmentation. We also increased the minimal frequency of collocation to a relatively high cut-off point of 200 instances, in order to filter out advertisements that plague all quantitative analyses of the Finnish newspaper corpus.[21] Then, after cleaning up Table 17.2 for errors and function words, we created a simple visualisation that is hopefully easier to understand for most historians.[22] Apart from absolute frequencies and Finnish originals, Figure 17.2 contains the same information as that in Table 17.2, but in a more accessible and user-friendly form. Figure 17.2 shows what could be poetically defined as 'the architecture of the concept'.[23] It is based on the principle that the human brain can intuitively understand: the closer the word is to the centre, the more strongly it is connected to the socialist present.

Table 17.2: The collocates of the present ('nykyä', 'nykyi*') in three socialist newspapers, 1895–1917, min. frequency 200.

RANK	WORD	TRANSLATION	FREQUENCY	MI-VALUE
1	yhteiskuntajär-jestelmä	societal system	354	10.29
2	tilanne	situation	1,460	9.91
3	järjestelmä	system	1,506	8.58
4	mallita	to prevail (OCR error)	448	8.44
5	vallita	to prevail	1,464	8.14
6	oleskella	to stay	404	8.11
7	yhteiskunta	society	2,026	8.11
8	tätä	this	282	7.56
9	olosuhteet	conditions	488	7.35
10	kapitalistinen	capitalist	394	7.32
11	kallis	expensive	587	6.95
12	pula	shortage	293	6.87
13	kurja	miserable	322	6.86
14	politiikka	politics	299	6.84
15	valtiollinen	state-	500	6.79
16	työttömyys	unemployment	452	6.76
17	muoto	form	801	6.68
18	asema	position, condition	3,547	6.64
19	olo	condition	1,731	6.63
20	kanta	view	897	6.56
21	kunnallinen	municipal	487	6.55
22	sota	war	1,415	6.55
23	taloudellinen	economic	658	6.53
24	vaikea	hard, difficult	295	6.52
25	käytäntö	practice	594	6.51
26	kehitys	progress, development	555	6.46
27	kurjuus	misery	280	6.42
28	säilyttää	to preserve, to save	323	6.39
29	epäkohta	grievance, shortcoming	307	6.36
30	mahdoton	impossible	485	6.35

Source: Lemmatised raw text files of *Työmies*, *Kansan Lehti* and *Länsisuomen Työmies* / *Sosialisti*, 1895–1917, distributed by the National Library of Finland, https://digi.kansalliskirjasto.fi/opendata.
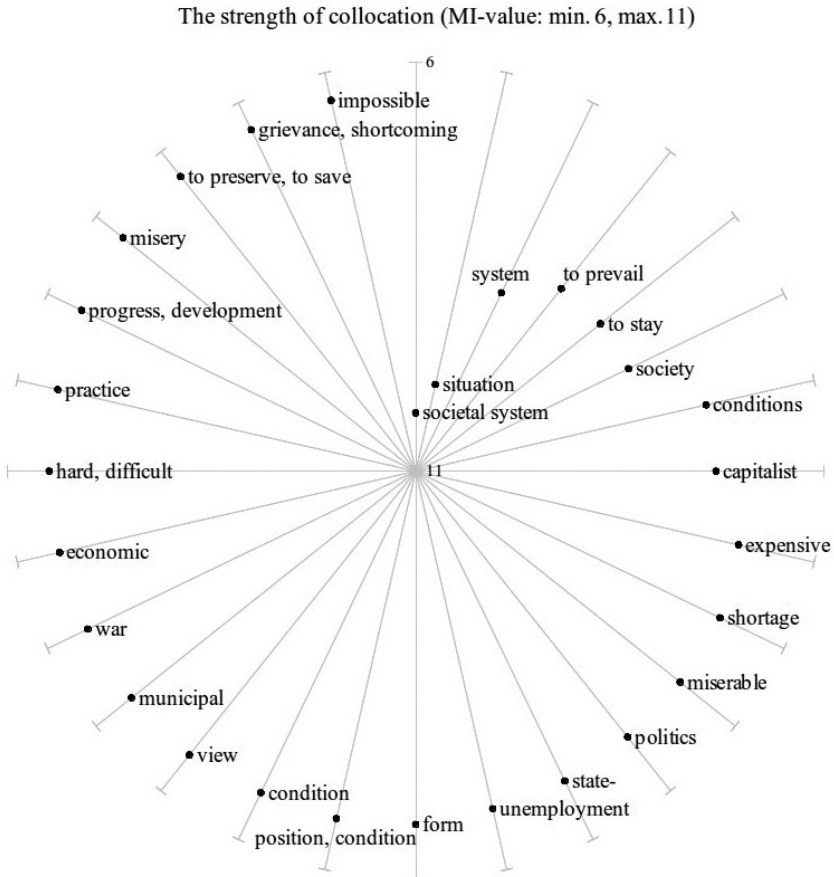
The strength of collocation (MI-value: min. 6, max. 11)

- impossible
- grievance, shortcoming
- to preserve, to save
- misery
- progress, development
- practice
- hard, difficult
- economic
- war
- municipal
- view
- condition
- position, condition

- form
- unemployment
- state-
- politics
- miserable
- shortage
- expensive
- capitalist
- conditions
- society
- to stay
- to prevail
- system
- situation
- societal system

6

11

Figure 17.2: The collocates of the present ('nykyi*', 'nykyä*') in three social-ist newspapers, 1895–1917. Source: Raw text files of *Työmies*, *Kansan Lehti* and *Länsisuomen Työmies / Sosialisti*, 1895–1917, distributed by the National Library of Finland, https://digi.kansalliskirjasto.fi/opendata.

It is relatively easy to find patterns in the figure for a historian with prior knowledge on the topic. First, the socialist present seems to attract phenomena that are considered to be negative, if not universally, at least as perceived by most people. In addition to the abstract concept of misery, the readers of the socialist newspapers were frequently introduced with the more concrete evils of 'shortage', 'unemployment' and 'war'. Negativity was enforced with the adjectives 'miserable' and 'hard, difficult', especially when talking about current 'conditions'. A close reading also reveals that current 'politics', referring to both tsarist repression and domestic bourgeois oppression, belong to the same nega-tive semantic field.[24] The critique of 'politics' is understandable since the social-ists, despite polling the most votes in every election, were not represented in the

government until in the year 1917.[25] Of course, in the socialist understanding, their interpretation of the present was not negative in an exaggerated sense, but rather a realistic portrayal of the inherent problems in capitalism.

This leads us to the second feature worth highlighting in Figure 17.2: the misery of the socialist present was not accidental, but *systematic*. The words 'societal system', 'system' and 'society' in the figure hint at this socialist pattern of conceptualising the present. The everyday had a structure, and those great 'capitalist', 'state', 'economic' and 'municipal' forces shaping the present were not mystical nor divine, but explainable through a rational analysis. Let us take a quote from each of the three main socialist newspapers to illustrate the logic:

> The eyes of many workers have opened even here [in Eura] to see the misery into which the present system has brought our society.[26]

> The curse of the present system is precisely that the more satisfied the capitalists can be with their lives, the more miserable the life of the workers has become under the constraining conditions.[27]

> Everyday experience shows us that it is not possible to achieve sufficient improvements for the condition of the majority of the people on the grounds of the present bourgeois system …[28]

Thus, socialists not only disapproved of the present with negative words, but they also tried to explain the causal mechanism behind it, by arguing that the present system was the root of all evil.

Although the present was in essence systematically bad, there was hope, or as the grand old man of Finnish labour history Hannu Soikkanen has argued in his seminal study *Arrival of Socialism in Finland* (1961), 'the present and future conditions were contrasted as starkly as possible'.[29] According to Soikkanen, this contrast was one of the main features of socialism that made it psychologically so attractive for the working people. The third pattern in Figure 17.2, visible to an experienced eye, refers to this connection between the present and the future, that is, to the words that imply changes and movement. 'The present situation' indicates a way of thinking that is not limited by eternal conditions created by God in the beginning of time. In addition, the phrase 'the prevailing conditions' is used regularly in the political language of Finnish socialism, in order to undermine the foundations of the present status quo:

> Modern socialism is thus a product of the prevailing economic and social conditions in the present.[30]

> Class struggle is rooted in the unsolvable conflicts between employers and employees. A collective agreement cannot remove the conflict, and neither can it abolish the hegemony of the capital over work in the prevailing conditions of the present.[31]

> Capital is accumulating in the hands of fewer and fewer people while the propertied class is not growing. Marx taught that the natural result is that the conditions themselves must change, there will be a fall of the prevailing system.[32]

Finally, the concept of 'progress' is fundamental to understanding the socialist temporality, for it tied together the past, present and future. If the present capitalist society was indeed a temporary product of a historical process, then society would surely change in the future too, or like one socialist journalist foretold: 'By the force of historical progress, the present system of oppression shall be once wiped from the stage, and its wretched henchmen shall get the reward they deserve.'[33]

We have seen that the collocation method, combined with different visualisation techniques (tables, figures), can produce a massive amount of low-level information on the semantic content of a concept in historical texts, in this case on the concept of the present in three leading socialist newspapers. It would have taken several years, perhaps a decade, to manually close read the more than 180 million words printed in these newspapers. However, a computational distant reading helped us to discover that the socialist present was (1) negative, (2) systematic and (3) changeable.

In the end, it always depends on the skills of a historian whether or not elementary quantitative information is successfully transformed into historical knowledge. Here, instead of limiting our critical thinking only to the meanings of preliminary 'results', we should also inquire into the presuppositions embedded in each quantitative method. For example, from a historian's point of view, the collocation method is lacking comparative contexts for it operates only within the political language of socialism. How do we know if these found features of the socialist present are unique, or if they belong to a more general discourse of the time?

Key Collocation: Placing the Socialist Present into the Contemporary Context

The collocation method shows the strength of mutual relation between two words. Another useful method historians interested in language could borrow from corpus linguistics is the keyness method, which can be used to show differences between two discourses. Keyness detects the words which appear more frequently than expected by pure chance in the text collection A ('target corpus') compared to the text collection B ('reference corpus').[34] Next, we combine the main ideas behind the two methods under the concept of key collocation, which aims to reveal semantic differences in the use of a certain historical concept.

First, as previously, we collect all the words appearing in a window of five words of the present (using search strings 'nykyä*' and 'nykyi*') in the socialist newspapers, and combine these words into one unified corpus. Then, we

do exactly the same for the bourgeois newspapers, in this case for the biggest liberal-nationalist newspaper *Helsingin Sanomat / Päivälehti* and the biggest conservative-nationalist newspaper *Uusi Suometar*. Now we have two corpora, and we can utilise the keyness method in order to see which words appear more frequently in the socialist discourse on the present compared with the bourgeois discourse on the present.

Table 17.3 places our findings on the socialist present in the previous section into the wider context of the early 20th-century newspaper discourse. The only clearly negative word in the top 20 key collocates is 'unemployment'.

Table 17.3: The key collocates of the socialist present ('nykyä*', 'nykyi'), compared with bourgeois newspapers, 1895–1917.

RANK	WORD	TRANSLATION	KEYNESS	EFFECT SIZE
1	työläinen	worker	1,917	19.8
2	yhteiskunta	society	1,586	5.8
3	oma	own (OCR error)	900	2.1
4	kapitalistinen	capitalist	598	38.5
5	köyhälistö	proletariat, the poor	587	28.7
6	yhteiskuntajärjestelmä	societal system	546	53.8
7	liitto	union	539	3.1
8	torppari	crofter	524	5.6
9	työmäki	working people (OCR error)	505	13.6
10	työ	work	504	2.0
11	työväki	working people	448	4.2
12	tilanne	situation	432	2.3
13	järjestelmä	system	422	2.3
14	työttömyys	unemployment	420	6.8
15	työnantaja	employer	414	5.8
16	työmies	working man	390	3.2
17	palkka	wage	368	2.3
18	moida	can (OCR error)	366	2.8
19	järjestö	organisation	357	5.3
20	ammatillinen	occupational, vocational	321	33.7

Source: Lemmatised raw text files of *Työmies*, *Kansan Lehti*, *Länsisuomen Työmies / Sosialisti*, *Helsingin Sanomat / Päivälehti* and *Uusi Suometar*, 1895–1917, distributed by the National Library of Finland, https://digi .kansalliskirjasto.fi/opendata.

However, scrolling further down the list shows that socialists indeed use words such as 'misery' (ranked 24 by keyness value), 'miserable' (27), oppression (87) and 'hunger' (111) in close proximity to the present much more often than their contemporaries, leading us to the conclusion that the level of negativity in the socialist discourse on the present was extraordinary. Socialists imagined the worst possible present.

What about the systematic nature of the present we encountered when quantifying collocates? It exists also in Table 17.3 in the form of 'society', 'capitalist', 'societal system' and 'system'. In addition, we have strong supporting evidence of 'social order' (76), 'economic system' (80), 'production system' (86), 'class society' (205) and 'system of oppression' (359).

The third feature of the socialist present, changeability, seems to be the least unique to labour newspapers. Apart from 'situation', words referring to the *changing* and *changeable* nature of the present are missing from Table 17.3. Some of them can be found in the key collocation list: for example, 'reaction' (70) and its counter-concept 'progress / development' (900). However, looking at the whole list of key collocates, it seems that this feature of the socialist present does not stand out in the context of Finnish newspaper discourse. Perhaps all the major political languages of the time—from liberalism, conservatism and socialism to Lutheran Christianity—believed that the world was changing, but they had different interpretations of what exactly was changing, how fast and, above all, if these changes happening in the present were leading to a better or worse society in the future.

It can be intellectually satisfying to find confirmation of prior interpretations, but nothing compares to finding something new. What is new is that the socialist protagonists of the present differ starkly from the bourgeois ones. The socialist version is based on the antagonism between good ('worker', 'working people', 'working man', 'proletariat / the poor', 'crofter') and evil ('employer') actors. This fundamental feature of the present was not found in the traditional collocation analysis, for the antagonism is so deeply rooted in the overall socialist discourse that it does not specifically stand out in the context of the socialist present. This socialist tendency to construct political agency through a vigorous repetition of collective singulars, especially 'the working people' and 'proletariat / the poor', can only be invoked when comparing socialism to other political languages of the time, in this case with the help of the key collocation method. Correspondingly, the trade union jargon ('union', 'organisation') escapes the collocation analysis, but it is clearly visible in the list of key collocations.

While collocation concentrates on the architecture of a concept in isolation, within only one discourse, key collocation can reveal the uniqueness or generality of these historical conceptual architectures. In the case of the socialist present, the latter method seems to confirm most of the findings achieved in the collocation analysis. Nevertheless, we should also respect the fact that an opposite result was possible. For example, we know that a negative present is not a feature confined to the socialist discourse of the early 20th century (old

people complained about present children and manners already in the days of Plato[35]), but the point is that key collocation gives us a comparative empirical context, against which we can measure how much (for example, 'negativity') is much. Constructing comparative contexts through traditional close readings is a labour-intensive task. Perhaps this is one reason why historical temporalities have been analysed based on a rather limited amount of sources.[36]

From Sources to Wisdom: DIKW for Digital Historians

In this final part before the concluding remarks and return to the sun of socialism, we rise from the empirical case study to a more abstract level, by providing a theoretical account of our intellectual journey so far with the help of the so-called DIKW pyramid, a concept that has been influential in the information sciences, knowledge management and systems theory for decades.[37] The pyramid describes the hierarchical relations between Data, Information, Knowledge and Wisdom. Although the pyramid has received criticism from several directions,[38] we believe that a slightly revised version of the pyramid can be useful for explaining not only the analytical process of this chapter, but also, more broadly, the idea and promise of distant reading in the context of digital history.

The vertical axis in the model represents what is usually described as 'connectedness'. Connectedness increases as we climb up the ladder towards wisdom.[39] The idea is not completely unfamiliar to historians, but we traditionally prefer the word 'context' when describing the process of historical analysis. In fact, the etymological root of context means weaving or joining together.[40] Thus, we can replace 'connectedness' with 'context' in the model without a bad conscience.

Then, we should add one layer below data, that is, historical sources.[41] The central difference between sources and data to digital historians is that the latter is machine-readable, and currently only a small part of historical sources is available for computational analysis as data. In the context of this chapter, physical historical newspapers are sources, whereas their digital representations—the PDF images and text files we downloaded from the National Library—belong to the category of data.

What is information, then? Here, we differ from general definitions of information as 'data + context', or 'data + meaning',[42] for the words 'context' and 'meaning' carry too much historical weight in the humanities. 'Information as processed data' is a more suitable definition for our purposes.[43] Examples of information would be simple word frequency time series (for example, Figure 17.1), word frequency tables (for example, Tables 17.1–17.3) or visualisations of words appearing close to one another (for example, Figure 17.2). In each of these examples, raw data has been computationally re-organised into low-level information.
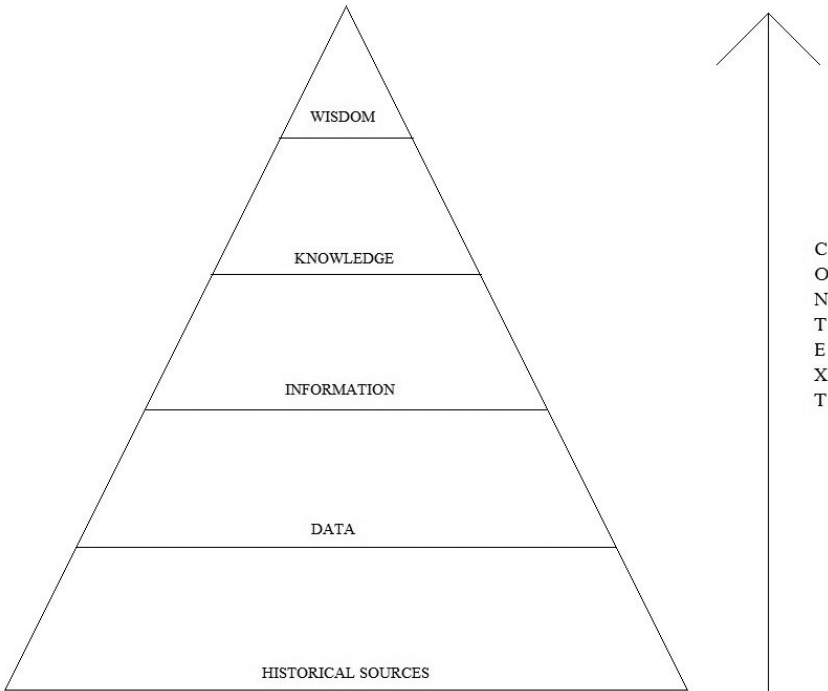
Figure 17.3: The revised version of the DIKW pyramid. Source: Author.

In our model, information does not include the historian's interpretation of information, which is located one step higher in the pyramid. Knowledge is information that is interpreted and contextualised by a human scholar. In this chapter, knowledge refers to assigning meaning to individual tables and figures, and then connecting these meanings not only to one another, but also to previous research—a difficult enterprise we undertook in the preceding pages.

The top of the pyramid is called wisdom, and it is the most controversial layer, for it escapes a clear definition.[44] Thus, it might be best to demolish it entirely. Russell Ackoff, often acknowledged as the founder of the pyramid, defined wisdom as evaluated understanding.[45] A historian could perhaps imagine wisdom as an ability to see which parts of the specific knowledge she has produced is relevant in answering the most complex questions of history. If knowledge means deciphering the meaning of the socialist sun, wisdom requires that a historian understands the meaning of this meaning under the aspect of eternity. (We have not reached an understanding this deep in this chapter.)

Now that we have reconstructed the pyramid, we can explain the point of distant reading in digital history. Distant reading aims at making the foundations of our historical explanations more solid, by piling up more stuff in the bottom of the pyramid. In other words, we want to increase the scale of historical sources, and this is possible as long as our sources are in a machine-readable

form. A historian can hardly read 100,000 words per day, whereas a computer can 'read' more than billions of words a minute. Of course, such distant reading is not reading with an understanding, but rather finding connections between words on a more primitive level.

However, we should not underestimate the fact that distant reading techniques, already in their current premature phase, can produce information that cannot be produced by human cognitive abilities alone. It is then up to a scholar to make sense of this elementary information. When a seasoned historian criticises your digital history research for 'lacking context', he or she probably means that you have not reached the level of historical knowledge, in other words, you have not paid enough attention to connecting your preliminary findings to one another and to previous research. Data nor information interprets itself. Nowadays, it is already a cliché that distant and close reading methods are complementary. With the visual aid of the pyramid, we could rephrase this idea: in digital history, the goal is to shift our limited cognitive energy upwards.[46] We distribute the most monotonous part of our historian's craft to the machines, in order that they could find patterns and trends that need human explanations. Machines are fast, precise and tireless, in other words, good at refining data to information, but, at least by now, only humans are able to refine information to knowledge.

Conclusions

What have we learned from our distant readings of the present, from turning our macroscope towards the sun of socialism? First of all, as expected, the General Strike of 1905 seems to be a pivotal moment, a measurable rupture, in the history of socialist temporality. The words referring to the present increase rapidly after the strike. Metaphorically, the strike meant a mental earthquake that had long-term consequences. The strike reshaped the political environment so dramatically that old ideological maps lost much of their ability to explain contemporary reality. In this new situation, the political language of Finnish socialism turned out to be a temporary winner, gaining most votes in each of the post-strike elections. According to Hannu Soikkanen, socialism gave working people a coherent and solid world view which stood in sharp contrast to their unstable conditions.[47] We could specify this argument from the perspective of temporality by adding that a new understanding of time formed one important part in the breakthrough of the socialist world view. In those turbulent times, the political language of Finnish socialism offered the most believable interpretation of the present for the working people. As demonstrated by our collocation analyses, labour newspapers convinced their readers that the present misery was caused by the system, and this system could and should be changed.

Thus, based on our distant readings, we argue that the meaning of the socialist sun in the early Finnish labour movement was not limited to the temporal

dimension of the future. The socialist sun affected the present too, by making it appear in a new, bad light. When one saw the red sun in the future, simultaneously the shackles of capitalism came out in the broad daylight of the present. We also learned that socialists did not see the present system as eternal or divine, but as historical and man-made. In fact, we could contrast the sun of socialism with the biblical sun which was the same for everyone, or in Jesus' words: 'He (Father in heaven) causes his sun to rise on the evil and good, and sends rain on the righteous and on the unrighteous.'[48] Our distant readings, especially the key collocation analysis, revealed that the socialist sun was shining exclusively for the 'working people', 'workers' and 'the proletariat', but not for 'the capitalist employers'. Thus, unlike the biblical sun that directed people's attention towards the hereafter, the red sun of socialism highlighted earthly problems.

If we wanted an even deeper understanding of the socialist sun and temporality, in other words, if we wanted to get closer to the top level of historical wisdom in the DIKW pyramid, these same analyses (comparable relative word frequencies, collocation, key collocation) should be performed for each dimension of the time: the past, present and future. In addition, we could broaden our quite narrow focus from word frequencies to richer forms of linguistic information. For example, experimenting with verb tenses (the socialist use of past, present or conditional forms) sounds reasonable when solving questions related to historical perceptions of time.

In the end, we could speculate for a moment: if the telescope and microscope changed the fields of astronomy and biology, will the macroscope, or computational methods in general, change our historical research?[49] We believe the answer is positive in the long term, but we are not quite there yet. According to Max Weber, a new 'science' emerges where new problems are pursued by new methods,[50] but in this chapter we have mainly answered old questions by using novel tools developed outside the community of historians. While not revolutionising the field, these tools can help us to improve our craft, in our everlasting quest towards historical wisdom.

Notes

1. 'Työwäestön olosuhteista Karjalassa', *Työmies*, 21 August 1903, p. 1, https://digi.kansalliskirjasto.fi/sanomalehti/binding/728854?page=1.
2. The quote is from Korff 1993: 124. On political symbols in the context of socialist movements, see, e.g., Steinberg 2002: esp. 224–246; Hake 2017: esp. 100–119.
3. See, e.g., Soikkanen 1961; Haapala 1986; Ehrnrooth 1992; Suodenjoki 2010; Rajavuori 2017.
4. On Marxian interpretation of history, see, e.g., Soikkanen 1961: 30, 91–92, 231–232; on socialist utopianism, see Ehrnrooth 1992: 169–177; on socialist eschatology, see Huttunen 2010: 57–65.

5 Kaihovaara 1986: 34–35; Kaihovaara 1991: 52.

6 Koselleck 2004: 263–267.

7 Tikka 2017.

8 Alapuro 1988: 115–117; Eley 2002: 66, table 4.2.

9 *Le Macrosope* was the title of Joël de Rosnay's classic introduction to systems theory (1975). Katy Börner reintroduced the concept in her often-cited article on plug-and-play macroscopes. See Börner 2011.

10 Available at https://digi.kansalliskirjasto.fi/opendata?language=en. For more information on the historical newspaper collection, see Pääkkönen et al. 2016.

11 The absolute numbers for the search string 'nyky*' in 1900–1917: 72,874 hits in 96,354,158 total words in *Työmies*; 121,473 times in 157,438,611 total words in *Uusi Suometar*; and 96,923 times in 135,166,576 total words in *Helsingin Sanomat / Päivälehti*. All the calculations in the chapter were performed using AntConc (3.5.7.), available from http://www.laurenceanthony.net/software.

12 Kujala 1995: 42–53.

13 Kaatra 1906: 4. Translation by the author.

14 This analysis is based on comparing the words appearing in close proximity to the present before (1.1.1904–30.10.1905) and after (7.11.1905–31.12.1907) the General Strike of 1905 in *Työmies*. In practice, all the words appearing in a window of five words to the left or right of the search string 'nyky*' were collected into one 'post-revolutionary' mini-corpus, which was then statistically compared with all the 'pre-revolutionary' neighbouring words of the same search string. The method is explained more carefully in the section entitled 'Key Collocation'.

15 Anderson 2006.

16 Kokko 2016: 24–25, 44–45, 413–425.

17 Anderson 2006: 24–26, 33.

18 On conceptual history, see, e.g., Koselleck 2004; on discourse analysis, Baker 2006; on the Cambridge school, Skinner 2002.

19 Finnish text files were lemmatised with the LAS command-line tool. See Mäkelä 2016.

20 There are different collocation metrics. We have used mutual information (MI) which measures the probability of whether or not the relationship between the search word and its neighbouring word is likely to exist by mere chance. The higher the MI value, the stronger the link between two words.

21 With the high cut-off point, many commercial adverts that contain words referring to the present 'drown' in the sheer mass of newspapers texts since not many of them circulate in each of the three chosen socialist newspapers. Thus, the ideological patterns of the socialist language use become more visible when the cut-off point is high.

22 The visualisation was achieved with LibreOffice (5.1.6.2), available from https://www.libreoffice.org/.

[23] de Bolla 2013.

[24] On tsarist repression in the socialist newspapers, see, e.g., 'Suomen suhde Venäjään', *Sosialisti*, 5 December 1908, p. 2, https://digi.kansalliskirjasto.fi/sanomalehti/binding/717172?page=2; 'Wenäjä ja me', *Kansan Lehti*, 13 December 1909, p. 1, https://digi.kansalliskirjasto.fi/sanomalehti/binding/649887?page=1; 'Wirkaryssät uusia eduskuntawaaleja hommaa massa', *Työmies*, 23 October 1912, p. 1, https://digi.kansalliskirjasto.fi/sanomalehti/binding/1187961?page=1. On domestic bourgeois oppression, see, e.g., 'Suhteemme hallitukseen', *Kansan Lehti*, 4 July 1907, p. 1, https://digi.kansalliskirjasto.fi/sanomalehti/binding/645861?page=1; 'Työväen suojeluslait', *Sosialisti*, 6 April 1911, p. 2, https://digi.kansalliskirjasto.fi/sanomalehti/binding/1177250?page=2; 'Budjettikeskustelu eduskunnassa', *Työmies*, 23 April 1913, p. 2, https://digi.kansalliskirjasto.fi/sanomalehti/binding/1187515?page=2.

[25] Soikkanen 1975: 120–197.

[26] 'Euran Pohjoispäästä', *Sosialisti*, 27 April 1907, p. 3, https://digi.kansalliskirjasto.fi/sanomalehti/binding/702923?page=3.

[27] 'Länsi-Teiskosta', *Kansan Lehti*, 12 January 1912, p. 3, https://digi.kansalliskirjasto.fi/sanomalehti/binding/1238591?page=3.

[28] 'Miksi me vaadimme järjestelmän muutosta', *Työmies*, 23 December 1902, p. 2, https://digi.kansalliskirjasto.fi/sanomalehti/binding/739481?page=2.

[29] Soikkanen 1961: 29.

[30] 'Hiukan sosialismista', *Kansan Lehti*, 22 June 1911, p. 1, https://digi.kansalliskirjasto.fi/sanomalehti/binding/1238279?page=1.

[31] 'Kollektiiviset työ- ja tariffisopimukset Englannissa ja Amerikassa', *Työmies*, 4 August 1909, p. 2, https://digi.kansalliskirjasto.fi/sanomalehti/binding/730033?page=2.

[32] 'Anarkia, pakkoluowutukset ja sosialidemokratit', *Sosialisti*, 28 September 1907, p. 2, https://digi.kansalliskirjasto.fi/sanomalehti/binding/703031?page=2.

[33] 'Sosialistin kirjapaino-osuuskunnan kirjapainon pakkokappaleasia', *Sosialisti*, 25 November 1913, p. 2, https://digi.kansalliskirjasto.fi/sanomalehti/binding/1178786?page=2.

[34] Baker 2006: 125–128. There are different keyness metrics. We have used the log-likelihood (four-term) test in order to measure statistical significance. In addition, we have measured effect size using ratio of relative frequencies.

[35] Quintelier 2007: 165.

[36] For example, the only monograph focusing on Finnish temporality in the late 19th century is based on a correspondence of one educated family in eastern Finland. See Ollila 2000.

[37] Rowley 2007.

[38] See, e.g., Tuomi 1999; Frické 2009; Jennex 2009.

[39] See, e.g., Bellinger, Castro & Mills 2003.

[40] Hyrkkänen 2009: 260.

[41] Perhaps we should add even one more layer to the very bottom of the pyramid, that is, life itself. People's rich lives leave only fragmentary sources for historians (and computers) to analyse.

[42] See, e.g., Worboys & Duckham 2004: 5; Floridi 2010: 20.

[43] Ackoff 1989.

[44] Hoppe et al. 2011.

[45] Bellinger, Castro & Mills 2003.

[46] This idea is taken from a very different academic field: plasma physics. See Carpenter & Cannady 2004: 4–6.

[47] Soikkanen 1961: 30.

[48] Matt. 5:45. On the many meanings of the sun in the Bible, see Patterson 2011.

[49] Graham, Milligan & Weingart 2015: 1–2.

[50] Franco Moretti made Weber's quote famous for digital humanists in his pioneering article on distant reading; see Moretti 2000.

References

Ackoff, R. (1989). From data to wisdom. *Journal of Applied Systems Analysis, 16*, 3–9.

Alapuro, R. (1988). *State and revolution in Finland*. Berkeley, CA: University of California Press.

Anderson, B. (2006). *Imagined communities: reflections on the origin and spread of nationalism*. London: Verso.

Baker, P. (2006). *Using corpora in discourse analysis*. London and New York, NY: Continuum.

Bellinger, G., Castro, D., & **Mills, A.** (2003, 7 May). *Data, information, knowledge and wisdom*. Retrieved from http://www.systems-thinking.org/dikw/dikw.htm

Börner, K. (2011). Plug-and-play macroscopes. *Communications of the ACM, 54*(3), 60–69.

Carpenter, S., & **Cannady, J.** (2004, July). *Tool for sharing and assessing models of fusion-based space transportation systems*. 40th AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit, Fort Lauderdale, Florida. DOI: https://doi.org/10.2514/6.2004-3535

de Bolla, P. (2013). *The architecture of concepts: the historical formation of human rights*. New York, NY: Fordham University Press.

Ehrnrooth, J. (1992). *Sanan vallassa, vihan voimalla: Sosialistiset vallankumousopit ja niiden vaikutus Suomen työväenliikkeessä 1905–1914*. Helsinki: Finnish Historical Society.

Eley, G. (2002). *Forging democracy: the history of the Left in Europe, 1850–2000*. Oxford: Oxford University Press.

Floridi, L. (2010). *Information—a very short introduction*. Oxford: Oxford University Press.

Frické, M. (2009). The knowledge pyramid: a critique of the DIKW hierarchy. *Journal of Information Science, 35*(2), 131–142. DOI: https://doi.org/10.1177/0165551508094050

Graham, S., Milligan, I., & **Weingart, S.** (2015). *Exploring big historical data: the historian's macroscope*. London: Imperial College Press.

Haapala, P. (1986). *Tehtaan valossa: Teollistuminen ja työväestön muodostuminen Tampereella 1820–1920*. Helsinki: Vastapaino and Finnish Historical Society.

Hake, S. (2017). *The proletarian dream: socialism, culture, and emotion in Germany, 1863–1933*. Berlin: De Gruyter.

Hoppe, A., Seising, R., Nürnberg, A., & **Wenzel, C.** (2011, July). *Wisdom— the blurry top of human cognition in the DIKW-model?* Proceedings of the EUSFLAT conference, Aix-les-Bains, France.

Huttunen, N. (2010). *Raamatullinen sota: Raamatun käyttö ja vaikutus vuoden 1918 sisällissodan tulkinnoissa*. Helsinki: Finnish Literature Society & Finnish Society of Church History.

Hyrkkänen, M. (2009). All history is, more or less, intellectual history: R. G. Collingwood's contribution to the theory and methodology of intellectual history. *Intellectual History Review, 19*(2), 251–263. DOI: https://doi.org/10.1080/17496970902722932

Jennex, M. (2009, January). *Re-visiting the knowledge pyramid*. Proceedings of the 42nd Hawaii International Conference on Systems Sciences. Waikoloa, Hawaii. DOI: https://doi.org/10.1109/HICSS.2009.876

Kaatra, K. (1906). *'Suurlakkokuvia' ym. työväenlauluja*. Hämeenlinna: Boman & Karlsson.

Kaihovaara, P. (1986). *Juurella lipun punaisen: Työväenjärjestöjen liput—osa työväenkulttuuria*. Helsinki: Yhteiskunnallinen Arkistosäätiö.

Kaihovaara, P. (1991). Työväenliikkeen lippukulttuurin symboleista. In U. Peltonen & K. Stenvall (Eds.), *Myytit ja symbolit: Kirjoituksia suomalaisista kulttuuritulkinnoista* (pp. 47–61). Väki Voimakas 5. Tampere: Finnish Society for Labour History.

Kokko, H. (2016). *Kuviteltu minuus: Ihmiskäsityksen murros suomenkielisen kansanosan kulttuurissa 1800-luvun puolivälissä* (doctoral dissertation, University of Tampere, Tampere). Retrieved from http://urn.fi/URN:ISBN:978-952-03-0282-5

Korff, G. (1993). History of symbols as social history? Translated by H. Drost. *International Review of Social History, 38*(Supp. 1), 105–125.

Koselleck, R. (2004). *Futures past: on the semantics of historical time*. Translated by and with an Introduction by K. Tribe. New York, NY: Columbia University Press.

Kujala, A. (1995). *Venäjän hallitus ja Suomen työväenliike 1899–1905*. Helsinki: Finnish Historical Society.

Moretti, F. (2000). Conjectures on world literature. *New Left Review*, *1*(January–February), 54–68.

Mäkelä, E. (2016). LAS: an integrated language analysis tool for multiple languages. *Journal of Open Source Software, 1*(6), 35. DOI: https://doi.org/10.1109/HICSS.2009.876 10.21105/joss.00035

Ollila, A. (2000). *Aika ja elämä: Aikakäsitys 1800-luvun lopussa*. Helsinki: Finnish Literature Society.

Patterson, R. (2011, 23 May). *Sun, light, and the Son*. Bible.org. Retrieved from https://bible.org/article/sun-light-and-son

Pääkkönen, T., Kervinen, J., Nivala, A., Kettunen, K., & Mäkelä, E. (2016). Exporting Finnish digitized historical newspaper contents for offline use. *D-Lib Magazine, 22*(7–8). DOI: https://doi.org/10.1045/july2016-paakkonen

Quintelier, E. (2007). Differences in political participation between young and old people. *Contemporary Politics, 13*(2), 165–180. DOI: https://doi.org/10.1080/13569770701562658

Rajavuori, A. (2017). *Esityksen politiikka: Sosialistinen agitaatio keskisuomalaisella maaseudulla 1906–1908*. Helsinki: Finnish Society for Labour History.

Rowley, J. (2007). The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of Information Science, 33*(2), 163–180. DOI: https://doi.org/10.1177/0165551506070706

Skinner, Q. (2002). *Visions of politics: regarding method*, Vol. 1. Cambridge: Cambridge University Press.

Suodenjoki, S. (2010). *Kuriton suutari ja kiistämisen rajat: Työväenliikkeen läpimurto hämäläisessä maalaisyhteisössä 1899–1909*. Helsinki: Finnish Literature Society.

Soikkanen, H. (1961). *Sosialismin tulo Suomeen: Ensimmäisiin yksikamarisen eduskunnan vaaleihin asti*. Porvoo: WSOY. DOI: http://doi.org/10.31885/2018.00017

Soikkanen, H. (1975). *Kohti kansanvaltaa 1. 1899–1937: Suomen Sosialidemokraattinen Puolue 75 vuotta*. Helsinki: Suomen Sosialidemokraatinen Puolue, Puoluetoimikunta.

Steinberg, M. (2002). *Proletarian imagination: self, modernity, and the sacred in Russia, 1910–1925*. New York, NY: Cornell University Press.

Tikka, M. (2017). Strike in Finland, revolution in Russia: the role of workers in the 1905 General Strike in the Grand Duchy of Finland. In M. Hilson, S. Neunsinger & I. Vyff (Eds.), *Labour, unions and politics under the Northern Star: the Nordic countries, 1700–2000* (pp. 197–217). New York, NY: Berghahn Books.

Tuomi, I. (1999). Data is more than knowledge: implications of the reversed knowledge hierarchy for knowledge management and organizational memory. *Journal of Management Information Systems, 16*(3), 103–117.

Worboys, M., & Duckham, M. (2004). *GIS: a computing perspective*. 2nd edn. Boca Raton, FL: CRC Press.