

# Geographic Space in Pentti Haanpää's Novel *Korpisotaa* – where does the War Happen?

Kimmo Kettunen<sup>1</sup>

<sup>1</sup> University of Eastern Finland, Yliopistokatu 2, 80100 Joensuu, Finland

## Abstract

Pentti Haanpää (1905-1955) was one of the most important Finnish authors in the first half of the 20th century. His short stories and novels describe life in the north-western part of the Finnish countryside many times, but his collected works also include many other themes. Among his works are five books, three novels, and two short story collections, which describe either military life or war. His first war novel, *Korpisotaa* describes the Finnish Winter War of 1939–40. Haanpää wrote the novel based loosely on his own war experiences for a competition for the best winter war novel arranged in 1940 by Prentice-Hall together with the Finnish publisher Otava; the novel was ranked third best in the competition. The novel is generally considered the first realistic war novel published in Finland [1-3], and its reception was favorable in general [4].

In this study, we focus on the analysis of geographic space in *Korpisotaa*. We use a digital version of the novel to be able to easily search for all the relevant space and location words in the novel. The methods we use in the study are familiar from linguistic corpus studies, and they have been used to some extent in literary studies as well. Besides common methods like keyness and frequency counts, we can benefit from a lexical semantic tagger of Finnish. Usage of the tagger systematizes the finding of the geographic space words in the novel and the comparison texts and enables us to perform keyness counts for semantic word groups instead of single words. Our work contributes especially to the use of digital methods in literary analysis and the creation of literary study corpora. Even for a novel-length, the availability of a digital version of the studied text helps detailed analysis very much, as will be shown in the analysis of *Korpisotaa*.

## Keywords

Pentti Haanpää, *Korpisotaa*, war novel, keyness, semantic tagging, geographic space

## 1. Introduction

Pentti Haanpää (1905–1955) was one of the most important Finnish authors in the first half of the 20th century. His short stories and novels describe life in the north-western part of the Finnish countryside many times, but his collected works also include many other themes. In his biography of young Haanpää, Eino Kauppinen [5] mentions that Haanpää was not "a proper regionalist", who wrote only about regional themes and details. Among his works are five books, three novels, and two short story collections, which describe either military life or war. His first war novel, *Korpisotaa* (published in 1940, 'War in the Backwoods/Wilderness', translated only in French as 'Guerre dans le désert blanc'), describes the Finnish winter war of 1939–40. Haanpää wrote the novel based loosely on his own war experiences for a competition for the best winter war novel arranged in 1940 by Prentice-Hall together with the Finnish publisher Otava; the novel was ranked third best in the competition. The novel is generally considered the first realistic war novel published in Finland [1-3] and its reception was

---

The 6th Digital Humanities in the Nordic and Baltic Countries Conference (DHNb 2022), Uppsala, Sweden, March 15-18, 2022.

EMAIL: Kimmo.kettunen@uef.fi (A. 1)

ORCID: 0000-0003-2747-1382 (A. 1)



© 2022 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

favorable in general [4: 272]. Haanpää is critical or ironic towards the establishment and its institutions in many of his works, particularly in earlier descriptions of military life, but this is mostly absent in *Korpisotaa*. Martikainen [6: 248], who has studied war discourse in the Finnish literature of 1917–1995, states that *Korpisotaa* belongs to the category of "hegemonic discourse" in the publications of the 1940s. The Finns are seen as a nation of heroes in this discourse, and the "spirit of the Winter War" is not broken in the novel [6: 149]. After a long period of publishing problems in the 1930s, the novel established Haanpää as one of the major authors in Finland for the rest of his life and career.

## 2. *Korpisotaa* – background

*Korpisotaa* describes the Finnish Winter War that lasted for 105 days. It is a novel, but part of its descriptions is based on the experiences of the author in the 6th battalion of the infantry regiment 40 [2, 7, 8: 126–132]. The Finnish winter war was fought in the backwoods during the climatically worst part of the year in the North: late autumn and winter. The winter of 1939–1940 was also exceptionally cold, which is well depicted by Keskisarja [9]. *Korpisotaa* does not describe very much individual soldiers, only a few of them are even named. Jokinen [10] interprets Haanpää's description of war in *Korpisotaa* so that the novel describes a collective of soldiers, who live under the harsh winter conditions without a possibility to have any influence on the events. According to Jokinen *Korpisotaa* differs from the mainstream literary descriptions of the Winter War: there is no room for individual braveness or initiative in the novel.

The novel mentions only a few soldiers repeatedly, a young second lieutenant, whose name is not given and who dies, and a foot soldier named Puumi. Other soldiers are mentioned by name only occasionally, and they are almost exclusively foot soldiers, which is considered typical for Haanpää's military descriptions. The enemy is called mostly *vihollinen* ('enemy', 153 times), and eleven times *iivana* (a derivation from the Russian name Ivan). Word *ryssä* (a derogatory name for a Russian) was used more in the original manuscript, but due to censorship, Haanpää was made to change it [3: 207–208; 8: 138]. Still, five mentions of *ryssä* were left in the first printing of the novel, but they were removed from the second printing in 1941, as the letter from the publisher Hannes Reenpää shows [11: 248; 12]. Soldiers of the Finnish side are many times called collectively as *meikäläinen* ('one of us'), not as Finns or Finnish soldiers. Word *meikäläinen* has altogether 42 mentions in the novel. The use of the word brings a sense of collectiveness to the narration: Finns are among themselves; the Russians are outsiders.

Quite a lot of description in the novel is given to places where the war happens: forests, rivers, swamps, ditches, fields, wilderness, etc. There are over 500 mentions of different geographic space words in the novel, but names of exact locations are seldom mentioned. In this study, we concentrate on the landscape or geography of *Korpisotaa* and analyze the usage of the most frequent space or geography-related words and word classes of the novel.

## 3. Analysis methods for digital text

This study uses corpus linguistic methods in the analysis of *Korpisotaa* [see e.g., 13]. These include especially the usage of keyness analysis [14–17] and semantic tagging of the text. We use digital versions of Haanpää's books obtained from the Finnish classics library released by the National Library of Finland (NLF)<sup>2</sup>. We locate and analyze the spatial words and expressions of the novel with the help of corpus linguistic methods and tools [16]. Besides common corpus tools such as AntConc [18], we can benefit from a semantic tagger of Finnish in our text annotation [19–20]. Semantic annotation of the whole novel and other texts of Haanpää is especially useful in finding different semantic classes of words in the text. Analysis of the semantically annotated novel shows that words in the semantic category geographical terms (W3 in the semantic USAS schema) are the sixth top keyness group among the semantic categories of the novel after such thematically obvious semantic categories as war and army, weather, and temperature. The category of geographical or conceptual space (M7 in the USAS schema) is 31st in the list of keyword classes (see Table 1).

<sup>2</sup> [https://digi.kansalliskirjasto.fi/collections?id=21&set\\_language=en](https://digi.kansalliskirjasto.fi/collections?id=21&set_language=en)

### 3.1. Semantic tagging

Our main analytical method in the analysis of *Korpisotaa* is semantic tagging or marking of the literary text(s). Semantic tagging is defined here as a process of identifying and labeling the meaning of words in a text according to some semantic scheme. This process is also called semantic annotation, and in our case, it uses a semantic lexicon to add labels or tags to the words [21-23]. Our semantic tagger, FiST [19], is based on the USAS semantic annotation schema of Lancaster University. The lexical-semantic description of the USAS framework is based on the modified and enriched categories of the Longman Lexicon of Contemporary English [24].

Semantic tagging of FiST is based on the idea of semantic (lexical) fields. Wilson and Thomas [23: 54] define a semantic field as "a theoretical construct which groups together words that are related by virtue of their being connected – at some level of generality – with the same mental concept". According to Dullieva [25], "a semantic field is a group of words, which are united according to a common basic semantic component" [cf. also 26-27]. For example, words that are related to the notion of time belong to one semantic field, (T), in the USAS schema. This field is subdivided into four different meaning classes for words that describe time from different viewpoints. Figure 1 shows this semantic class. Alphanumeric abbreviations in front of the meaning classes are the actual hierarchical semantic tags used in the lexicon.

<b>T TIME</b>	
T1	Time
T1.1	Time: General
T1.1.1	Time: General: Past
T1.1.2	Time: General: Present; simultaneous
T1.1.3	Time: General: Future
T1.2	Time: Momentary
T1.3	Time: Period
T2	Time: Beginning and ending
T3	Time: Old, new and young; age
T4	Time: Early/late

**Figure 1:** Semantic class of time in the USAS classification

The descriptive approach taken in the USAS framework is quite generic: although lexical meaning classes in the semantic schema cover phenomena of the world quite extensively, the inner structure of the semantic classes may vary in specificity – some classes are more developed and fine-grained, some have only an elementary classification. The semantic lexicon of USAS is divided into 232 meaning classes or categories, which belong to 21 upper-level fields.

Löfberg [22] has compiled a Finnish semantic lexicon of 46 226 lexemes using the USAS annotation schema; her thesis also evaluates the lexical coverage of the lexicon with several different types of texts. Kettunen [19] introduced a prototype semantic tagger based on this lexicon and analyzed its lexical coverage with a variety of Finnish texts from different genres. At best lexical coverage of the tagger was 91–92%. With several high-quality fiction texts of the early 20<sup>th</sup>-century Finnish prose, the tagger achieved a lexical coverage of ca. 90–91% [19].

### 3.2. Data acquisition and semantic marking

Books that are available in the Finnish classics library can be studied either online or downloaded as pdf files. To be able to conduct this study we transformed the original pdf files of NLF's Haanpää digitizations into text files using the *pdftotext* utility<sup>3</sup>. We corrected the text files after *pdftotext* conversion by removing line-ending hyphens thus joining the beginning and end of the words on adjacent lines. Printing information in front and back of the books and extra empty lines were also

<sup>3</sup> <https://www.xpdfreader.com/>

removed. Already this improved lexical coverage of the tagger by some percent. After tagging the text files with FiST, we measured lexical coverage of the semantic tagging in the data: the tagger reached lexical coverage of ca. 79.7–87.5% in different texts of Haanpää. In the novel *Korpisotaa*, the lexical coverage is 87.15%. This can be considered adequate coverage remembering that the texts contain OCR errors and that Haanpää’s language is partly old-fashioned and slightly dialectal.

### 3.3. Corpus methods used in this study

To be able to systematically analyze *Korpisotaa*, we use corpus analysis methods in this study. Especially this means the usage of the keyness method introduced by Scott [14–15]. After its introduction, the method has been used mainly in corpus linguistics, but it has also gained some methodical status in general text analyses [28] and literary studies [16, 29–30]. Shortly put keyness is a statistical comparison method for texts. With the method, one text, usually called the study text, is compared to a larger text or group of texts - usually called the reference text. Keywords reveal the aboutness of the study text by highlighting its specific words in comparison to the reference text [15, 31]. The comparison uses statistical measures to distinguish meaningful differences in the texts on word, word cluster, phrase, or some other level if the texts have linguistic annotation. Many times, log-likelihood [32] is used as the statistical significance measure, but also other measures are used [28]. In corpus software AntConc several different statistical measures can be chosen, and we used log-likelihood for keyword statistics and Gabrielatos’s %DIFF measure as the keyword effect size measure [28].

An important methodological prerequisite for the use of keyness is the size of the reference corpus. Scott [14] did not specify this very accurately, but Berber-Sardinha [33] has shown that a reference text of ca. five times larger than the target text is enough to make statistical comparisons significant.

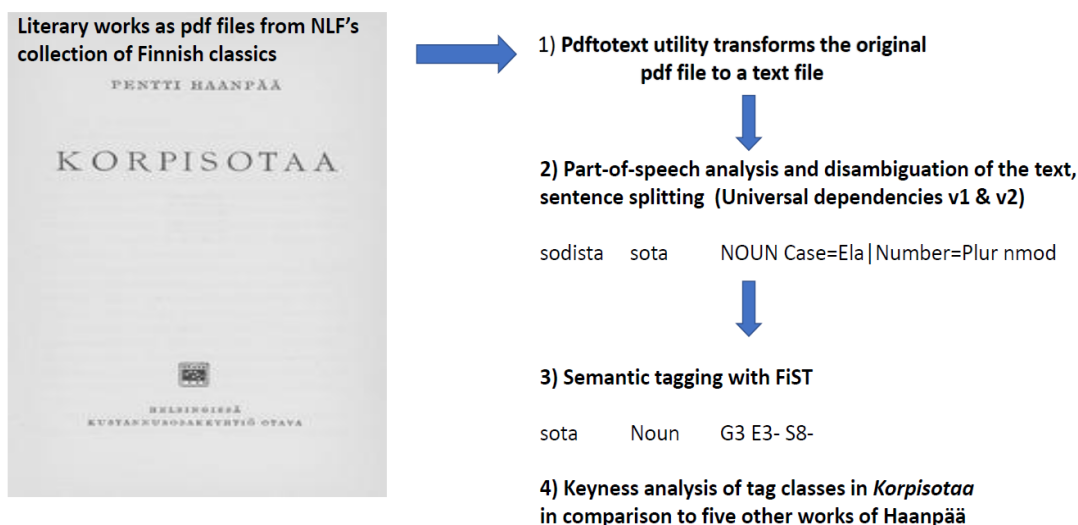
We chose five works of Haanpää from the 1920s and 1930s as reference texts for the keyness analysis. In the publication order, the works are the following: *Maantietä pitkin* (published in 1925, a short story collection), *Hota-Leenan poika* (published in 1929, a novel), *Isännät ja isäntien varjot* (published in 1935, a novel), *Lauma* (published in 1937, a short story collection), and *Taivalvaaran näyttelijä* (published in 1938, a novel). The works are from the same period or 10–15 years earlier than *Korpisotaa* and do not mainly describe anything related to war or army. The only exception is *Taivalvaaran näyttelijä*, where the main character is supposedly a person, who was lost in a battle and thought dead. Altogether these five reference works have ca. 133 670 words. The size of *Korpisotaa* is ca. 27 850 words, and thus the size of our reference corpus is along the lines suggested by Berber-Sardinha [33], the reference corpus being 4.8 times larger than the target corpus. A larger reference corpus could be used, but the keyword analysis method should be robust and produce plausible results anyhow [15]. A different reference corpus with a few more texts could perhaps bring a slightly different set of keywords, but it would produce a common core of keywords<sup>4</sup>. Thus, our selection should be representative of Haanpää’s writing of the time and large enough to fulfill the requirements of being a reference text collection in keyness analysis.

Figure 2 depicts the creation of the study corpus and its different representations used in the study. The same procedure was followed in the creation of the target corpus.

---

<sup>4</sup> We also created a keyness list with seven comparison texts adding two short story collections, *Karavaani* (1930) and *Ihmiselön karvas ihanuus* (1939) to the comparison texts. This changed the order of some of the top classes with a rank or two. W3 and M7 are still among the chosen classes. W3 is the seventh on the list and M7 the 31st. The size of keyword class set was also smaller with seven comparison texts: 33 versus 40.

## Phases of digital corpus creation for the study



**Figure 2:** Creation of the study corpora

Our most general analyses are based on semantic tag classes of the texts: text representations of the phase 3 in Figure 2. Besides keyness analysis, we can make word-level searches to the semantically tagged corpora. In addition to this, we have also sentence-by-sentence versions of the texts (results of Universal dependencies v2 analyses, phase 2), out of which we can locate original example sentences from the corpora.

In the word analysis sample of Figure 2 (phase 3) we can see several meaning tags marked for the word *sota* (‘war’) in the analysis result. Multiple tags are marked in the lexicon of the tagger for semantically ambiguous words, and FiST does not resolve ambiguity. In most of the cases, the first tag is probably the right one, as the most frequent tag for each word is the first one in the semantic lexicon [22: 74]. When we analyze the texts, we only use the first tags marked for the words. In the literature on word sense disambiguation, this is known as the most frequent meaning baseline, which is many times hard to outperform with disambiguation methods [34-35]. Many of the disambiguation methods also have a bias towards the most frequent sense of the word [36-37].

### 3.4. Semantic classes of space in the Finnish USAS schema

Our analysis of *Korpisotaa* concentrates on the geographic space of the novel. Two main semantic classes denote space in the USAS semantic classification: M7 (places) and W3 (geographical terms) [22]. Also, names of locations (Z2) can be considered as part of geographical space. Class M6 is a class of location and direction, but it contains very few interesting words for our analysis, and it is thus left out. This study analyzes only the usage of semantic classes W3 and M7 in *Korpisotaa*.

Semantic class W3 comprises words that denote geographical terms. Löfberg [22] mentions as prototypical examples of this class such words as *joki* (‘river’), *aallokko* (‘waves’), and *aarniometsä* (‘primeval forest’), among others. The words of this class describe mainly nature and its elements and formations. The Finnish semantic lexicon contains 330 words, which have as their first semantic tag W3.<sup>5</sup>

Class M7 contains words that refer to geographical or conceptual spaces. Examples of these are *kirkonkylä* (‘village center’), *mantere* (‘continent’), and *osavaltio* (‘state’). The Finnish semantic lexicon contains 294 words, which have as their first semantic tag M7.

Table 1 lists the 10 top keyness classes found in *Korpisotaa* when the first semantic tags of the novel’s words have been compared to the semantic tags of the reference texts using AntConc’s Keyword

<sup>5</sup> <https://github.com/UCREL/Multilingual-USAS/tree/master/Finnish>

List functionality. We have also added class M7 on the 31st place. Z2, the class of location names, is not among the keywords. G3, the class of warfare and defence, is the most distinctive class, as one would expect, as the class contains lots of occurrences of military ranks and war-related words. Mentions of temperature (O.4.6) and weather (W4) are obvious in the context of the Winter War. The class of sports sounds odd here, but it is natural, as skiing and skis are part of it, and the Finnish army moved by skiing in the war. Occurrences of the class S1.2.1 consist mostly of the word *vihollinen* ('enemy').

**Table 1**

The most important semantic classes in *Korpisotaa* according to rank and keyness using five comparison texts

Rank	Number of class occurrences	Keyness value	Semantic class <sup>6</sup>
1	585	+750.68	G3 Warfare, defence, and the army; Weapons
2	295	+169.78	O4.6 Temperature
3	112	+112.73	K5.1 Sports
4	185	+90.95	W4 Weather
5	167	+75.08	S1.2.1 Approachability and Friendliness
6	323	+72.99	<b>W3 Geographical terms</b>
7	555	+70.02	M6 Location and direction
8	120	+56.81	L3 Plants
9	673	+48.25	B1 Anatomy and physiology
10	213	+47.96	O2 Objects generally
...	...	...	...
31	200	+12.08	<b>M7 Places</b>

### 3.5. Semantic classes W3 and M7 – geographical terms and geographical space

In *Korpisotaa* class W3 is the sixth most frequent key semantic class with 323 occurrences and a keyness value of 56.43. M7 is the 31st most frequent key semantic class with 200 occurrences and a keyness value of 12.08, as was seen in Table 1.

Tables 2 and 3 show the top-10 words in these two categories in *Korpisotaa* with their rank in the frequency list, absolute frequency, and normalized frequency per 10 000 words.

**Table 2**

Top-10 words of the class W3

W3 word	Meaning	Absolute frequency	Normalized frequency (per 10 000)	Rank out of 5593 lemmas
metsä	forest	70	25.12	39
kuoppa	pit, foxhole	32	11.48	104
erämaa	wilderness	27	9.69	136
ranta	shore	20	7.18	192
järvi	lake	19	6.82	216
joki	river	17	6.10	243

<sup>6</sup> [http://ucrel.lancs.ac.uk/usas/usas\\_guide.pdf](http://ucrel.lancs.ac.uk/usas/usas_guide.pdf)

korpi	backwoods	15	5.38	274
pelto	field	11	3.95	375
rinne	hillside	9	3.23	454
meri	sea	8	2.87	518

**Table 3**

Top-10 words of the class M7

M7 word	Meaning	Absolute frequency	Normalized frequency (per 10 000)	Rank out of 5593 lemmas
maa	ground, soil, country	83	29.79	30
kylä	village	50	17.94	55
paikka	location, place	19	6.82	211
kaupunki	town	12	4.31	347
pohjola	North	9	3.23	456
raja	border	7	2.51	567
isänmaa	homeland	7	2.51	606
tila	space	6	2.15	639
alue	area	4	1.44	1147
sija	position	3	1.08	1284

As can be seen, words denoting forest and ground are the most frequent words of these two meaning classes. Their rank is also high in the whole vocabulary of the novel. After *forest* in the class W3 come *pit* and *wilderness*, and in M7 *village* and *location*.

Due to space restrictions, we shall analyze only the usage of the most frequent words in each of the two semantic classes. Further analysis of the words is left for later study.

#### 4. W3 words – metsä (forest)

Forest is a self-evident location for a literary description of the Finnish Winter War. The most common characteristics of the forest in the novel are darkness and snow, which is understandable. The Finnish winter is usually quite snowy in the areas where the Winter War was fought, and winter is the darkest time of the year, even if days begin to lengthen slowly after the winter solstice on December 21.

Fifteen mentions of forest are connected to darkness in the forest: either dark/darkness/black is an adjacent attribute of forest or mentioned in the same sentence with forest. Soldiers move in the dark forests. The attitude to the forest's darkness is twofold: either the dark forest is threatening

*"Tulijoitten ei sentään suoraa päätä tarvinnut lähteä tuonne mustiin metsiin, jossa laukaukset räjähtelivät ja luodit vingahtelivat."* ('... shots and bullets are heard in the dark forest')

or protective

*"Pimeys ja metsä suojelivat."* | *"Viimeinkin aava loppui, ja vainottu sotamies hiihtää hoippuroi metsän suojaan."* | *"He hiipivät suksilla eteenpäin tiheään metsän suojaan."* ('soldiers ski in the cover of the forest or get there after an open space')

When soldiers move in the dark forest, the forest seems sometimes endless:

*"Aina riitti pimeätä metsää ja upottavaa lunta." | "He marssivat kilometrimääriä pimeitä metsiä ja nevoja." | "Rannattomia metsiä..." ('The forest lasts for kilometers or seems endless')*

Snow is mentioned eleven times with forest, and it seems to have a quite general function in the forest description: forests are covered with snow. A few mentions are given to the whiteness or light of snow in the forest, but otherwise, it is not characterized much. Once sinking of snow underfoot is mentioned, and once the snow squeaks underfoot - it does this only when the weather is cold enough. Snow seems mainly to be a general element of winter and the forest and belongs to the time of the year.

Skiing was the main means of moving in the forest in the Winter War. However, it is mentioned only four times explicitly with forest in *Korpisotaa* - with the word forest in the same sentence with skiing. Altogether skiing is mentioned 70 times in the novel in different ways, and the context implies forest many times without mentioning it directly. Thus, the forest is present more in the novel than the plain word count reveals.

## 5. M7 words – maa (earth/ground/country/soil)

The word *maa* is the most common word of the meaning class M7 in *Korpisotaa*. *Maa* is a polysemous word, which has several different meanings. Nykysuomen sanakirja [38] lists four main meanings for it – 'globe', 'ground', 'soil', and 'area' (with one submeaning of 'country', 'state') – and states that these meaning groups have fuzzy boundaries. In *Korpisotaa* all these four main meanings are in use but in very different proportions. The novel uses the word *maa* only once in the meaning 'globe':

*"Ellei tästä sodasta mitään muuta hyvää siunautuisikaan, niin onhan komea maine kuulumassa ympäri maan piiriä."* ('A grand reputation will be heard all around the globe because of this war')

About a quarter of the usages of the *maa* is in the meaning of country or state:

*"Maassa tapahtui liikekannallepano." | "Maahan oli hyökätty, ja sellaisen seikan varalta oli suomalaisella ammoiset valmiit kaavat: taistella, vaikka joka kynsi kylmenisi ..."* | *"Mitä lienee siitä naurusta ajatellut väijyksissä makaava vieraan maan mies?"* ('there was a mobilization in the country' | 'the country was attacked...' | '...soldier of a foreign country')

Some of the uses of *maa* in the meaning country are fuzzy in their meaning, there is a hint of concreteness in these examples in their context (relating to ground):

*"Ja kuitenkin tämä oli hänen maansa." | "Ne luottivat siis siihen, että ensi kesänäkin on vilja heiluva tuulessa, että maa meidän on ja olla täytyy." | "Vihollinen on saava tästä maasta vain tulta ja tuhkaa."* ('Anyhow, this was their country|ground' | 'They counted on the fact that crops will sway in the wind even next summer and the country|ground is ours and it must be' | 'The enemy will only get fire and ashes out of this country|ground')

There is a concrete allusion to the ground in all these sentences, but at the same time, the meaning of country is present. It is hard to say which meaning is the prevalent one in these examples.

One sentence in the novel uses the word *maa* three times in three different meanings:

*"Maata ne lähtivät valloittamaan ja joutuivat itse maahan ja muuttuvat maaksi."* ('They came to conquer the country, but were put in the ground and will become dust')

Most of the meanings of *maa* belong to the group ground and soil. Soldiers dig the frozen ground; they lay on the ground looking for shelter: they feel that the ground is their protection, and the ground is characterized as good a few times in this connection. When ammunition hits the ground, the ground shakes. Both Finns and Russians have caved their dugouts to the frozen ground.



## 6. Discussion

We analyzed in this study usage of locations and geographic space words in Pentti Haanpää's novel *Korpisotaa*. We had available a digital version of the novel and could make systematic searches and analyses out of the novel's text. We used corpus software AntConc and a semantic tagger for Finnish to be able to easily locate expressions of location and geographic space in the novel. We used keyness analysis to extract the most distinguishing semantic classes out of the novel in comparison to a reference corpus that consisted of five other works of the author.

Our analysis concentrated on two different semantic classes in the USAS semantic schema: M7 (places), and W3 (geographical terms). We can summarize the usage of the location-related words in the novel as follows.

1) Words in the two locational USAS classes W3 and M7 in the novel describe either Finnish natural landscape, civilization, or space. Mentions of political space are not very frequent, but they exist (border, homeland/country).

2) The most frequent two words in the classes W3 and M7 are *metsä* (forest) and *maa* (earth/ground/country/soil). Forest is one of the main scenes of the war and it is described both as threatening and protective. Many times, the forest seems also endless. *Maa* is a polysemous word with four main meanings. Part of them relate to the country or state, but mainly *maa* is used in its concrete meanings of ground and soil.

The main contribution of this paper is methodological. We use a well-known corpus method, keyness analysis, with semantic annotation of a literary corpus and can use different textual representations of the literary text in the study. As the complete works of even one author can consist of thousands of pages, mere human reading of the works becomes challenging fast, let alone, when one wants to study large collections of fiction. Computer-aided ways of going through the works are thus needed, and corpus methods used in linguistics can offer 'semi-distant' reading aids for a literary scholar. We used keyness analysis for a small literary corpus, one novel. Even for a novel-length, the availability of a digital version of the text benefits detailed analysis very much. Usage of a semantic tagger of Finnish brought available a more general level of analysis than plain words. Methods like keyness analysis do not substitute for close reading of literary works, but they can help the reader to focus on the most relevant parts of the texts. Possibility and results of fully automatic literary analyses have been criticized heavily for example by Da [39] and Fletcher [40]. In keyness analysis computing works as a starting point for human analysis by pointing out interesting topics for study by using textual statistics. The actual analysis is left for humans, as it should be.

## 7. References

- [1] P. Haanpää, *Korpisotaa*, Third printing, Otava, Helsinki, 1999.
- [2] V. Karonen, *Haanpään elämä*, Suomalaisen Kirjallisuuden Seura, Pieksämäki, 1985.
- [3] V. Karonen, Pentti Haanpään talvisota ja Korpisotaa, in: P. Haanpää, *Korpisotaa*, Third printing, Otava, Helsinki, 1999.
- [4] J. Koivisto, *Leipää huudamme ja kiviä annetaan*. Pentti Haanpään 30-luvun teosten kytkentöjä aikansa diskursseihin, todellisuuteen ja Raamattuun, SKS 284, Helsinki, 1998.
- [5] E. Kauppinen, *Pentti Haanpää I. Nuori Pentti Haanpää 1905–1930*, Otava, Helsinki, 1966.
- [6] E. Martikainen, *Kirjoitettu sota. Sotadiskursseja suomalaisessa kaunokirjallisuudessa (1917–1995)*, Ph.D. thesis, Tampere University Press, 2013.
- [7] P. Haanpää, *Kirjeitä kahdesta sodasta ('Letters from two wars')*, Second printing. Otava, Helsinki, 1977.
- [8] E. Viirret, *Haanpään siivellä*, Pilot-kustannus Oy, 2005.
- [9] T. Keskisarja, *Raaka tie Raatteeseen. Suurtaistelun ihmisten historia*, 7th printing, Siltala, Helsinki, 2012.
- [10] A. Jokinen, *Isänmaan miehet – Maskuliinisuus, kansakunta ja väkivalta suomalaisessa sotakirjallisuudessa*, eds. Markku Soikkeli, Ville Kivimäki, Vastapaino, Tampere, 2019.
- [11] P. Haanpää, *Kirjeet ('Letters')*, eds. Vesa, Esko Viirret, Otava, Helsinki, 2005.

- [12] H. Pilke, *Etulinjan kynämiehet — suomalaisen sotakirjallisuuden kustantaminen ja ennakkosensuuri kirjojen julkaisutoiminnan sääntelijänä 1939–1944*, SKS, Helsinki, 2009.
- [13] N. Smith, S. Hoffmann, and P. Rayson, *Corpus Tools and Methods, Today and Tomorrow: Incorporating Linguists' Manual Annotations*, *Literary and linguistic computing*, 23 (2) (2008), 163–180.
- [14] M. Scott, *PC Analysis of Key Words – and Key Key Words*, *System* 25(2) (1997), 233–245.
- [15] M. Scott, M., *Problems in investigating keyness, or clearing the undergrowths and marking out trails*, in: M. Bondi, M. Scott (eds.), *Keyness in Texts*, John Benjamins, pp. 43–57.
- [16] B. Fischer-Starcke, *Corpus Linguistics in Literary Analysis. Jane Austen and her Contemporaries*, Continuum, London, 2010.
- [17] V. Brezina, *Statistics in Corpus Linguistics. A Practical Guide*, Cambridge University Press, 2018.
- [18] L. Anthony, *AntConc (Version 3.5.8) [Computer Software]*, Tokyo, Japan: Waseda University, 2019. URL: <https://www.laurenceanthony.net/software>.
- [19] K. Kettunen, K., *FiST – towards a Free Semantic Tagger of Modern Standard Finnish*, in: *IWCLUL2019*, <http://aclweb.org/anthology/W19-0306>
- [20] K. Kettunen, M. La Mela 2021, *Semantic Tagging of Political Concepts: the Case of Everyman's Rights*. *Digital Scholarship in the Humanities*. DOI: <https://doi.org/10.1093/llc/fqab052>
- [21] G. Leech, *Developing Linguistic Corpora: a Guide to Good Practice Adding Linguistic Annotation*, 2004. <https://ota.ox.ac.uk/documents/creating/dlc/chapter2.htm>
- [22] L. Löfberg, *Creating large semantic lexical resources for the Finnish language*, Ph. D. thesis, Lancaster University, 2017. DOI: <https://doi.org/10.17635/lancaster/thesis/3>.
- [23] A. Wilson, J. Thomas, *Semantic annotation*, in: R. Garside, G. Leech, T. McEnery (Eds.), *Corpus annotation: Linguistic information from computer text corpora*. New York: Longman, pp. 53–65.
- [24] T. McArthur, *Longman Lexicon of Contemporary English*, Longman, London, 1981.
- [25] K. Dullieva, *Semantic Fields: Formal Modelling and Interlanguage Comparison*, *Journal of Quantitative Linguistics* 24(1) (2017) 1–15. DOI: <https://doi.org/10.1080/09296174.2016.1239400>.
- [26] D. Geeraerts, *Theories of Lexical Semantics*, Oxford University Press, Oxford, 2010.
- [27] P. R. Lutzeier, *Lexical fields*, in: Allan, L. (ed.), *Concise Encyclopaedia of Semantics*. Elsevier, Oxford, 2006, pp. 470–3.
- [28] C. Gabrielatos, *Keyness Analysis: nature, metrics and techniques*, in: C. Taylor, A. Marchi, (eds.), *Corpus Approaches to Discourse: A critical review*. Oxford: Routledge, 2018, pp. 225–58.
- [29] M. Bondi, M. Scott, *Keyness in Texts*, John Benjamins, 2010. DOI: <https://doi.org/10.1075/scl.41>.
- [30] J. Culpeper, *Keyness: Words, parts-of-speech and semantic categories in the character-talk of Shakespeare's Romeo and Juliet*, *International Journal of Corpus Linguistics* 14(1) (2009) 29–59.
- [31] M. Bondi, *Perspectives on keywords and keyness. An Introduction*, in: M. Bondi, M., Scott, (eds.), *Keyness in Texts*, John Benjamins, pp. 1–18.
- [32] T. Dunning, *Accurate Methods for the Statistics of Surprise and Coincidence*. *Computational Linguistics*, 1993. URL: <https://aclanthology.org/J93-1003.pdf>
- [33] T. Berber-Sardinha, *Comparing corpora with WordSmith Tools: How large must the reference corpus be?* in: *CompareCorpora '00: Proceedings of the Workshop on Comparing Corpora, October 2000*, pp. 7–13. URL: <https://www.aclweb.org/anthology/W00-0902/>.
- [34] R. Navigli, *Word Sense Disambiguation: A Survey*, *ACM Computing Surveys* 41 (2009) 10–69.
- [35] M. Bevilacqua, T. Pasini, A. Raganato, R. Navigli, *Recent Trends in Word Sense Disambiguation: A Survey*, in: Z-H Zhou (ed.), *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21. International Joint Conference on Artificial Intelligence, Inc., Vienna*, pp. 4330–4338. *International Joint Conference on Artificial Intelligence, Montreal, Canada*, 21/08/2021. DOI: <https://doi.org/10.24963/ijcai.2021/593>
- [36] M. Postma, W. Izquierdo, E. Agirre, G. Rigau, P. Vossen, *Addressing the MFS Bias in WSD systems*, in: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pp. 1695–1700. URL: <https://www.aclweb.org/anthology/L16-1268>.
- [37] J. Preiss, *A detailed comparison of WSD systems: an analysis of the system answers for the SENSEVAL-2 English all words task*, *Natural Language Engineering* 12(3) (2006) 209–28. DOI: <https://doi.org/10.1017/S1351324906004281>.

- [38] Nykysuomen sanakirja, (Dictionary of Modern Finnish), Tome 3: L-N. Werner Söderström osakeyhtiö, Porvoo, 1954.
- [39] N. Z. Da, The Computational Case against Computational Literary Studies, *Critical Inquiry* 45:3 (2019) 601–639.
- [40] A. Fletcher, Why Computers Will Never Read (or Write) Literature: A Logical Proof and a Narrative, *Narrative* 29(1) (2021) 1–28. DOI: 10.1353/nar.2021.0000.