What was the Finnish *Hiisi*? Applying Computational Methods to the study of Folk Religion

R. Pulkkinen[1], M. Salmenkivi[2], A. Leino[2] and H. Mannila[2]

[1] Department of Comparative Religions, University of Helsinki, Finland
[2] Helsinki Institute for Information Technology, Department of Computer Science, University of Helsinki, Finland

Abstract. Place-names reflect the folk world view and thinking. Electronic maps enable the use of large-scale data analysis methods in the study of place-names. We consider some of the methodological opportunities. We utilize a corpus containing 717,646 Finnish place-names, their coordinates and type descriptions. The large corpus being available, it is reasonable to expect statistically significant attraction between name elements belonging to the same idea complex. The evaluation of the significance is not straightforward, however. We introduce computationally intensive methods for the purpose, and apply the methods to the problem of the meaning of *hiisi*, a key notion of the ancient Finnish folk religion. In earlier research *hiisi* has been seen, e.g., as a "holy grove", or a place for a cult of the dead. Our analysis gives some support for these hypotheses, though no striking attraction was detected between the name element *hiisi* and name elements refering to the holy, or to death.

1. Introduction

Place-names reflect the folk world view and thinking. Thus, they provide an important source of information for the research on folk religion. Electronic maps enable the use of large-scale data analysis methods in the study of place-names. In this article we consider some of the methodological opportunities, especially from the viewpoint of the study of folk world view.

We concentrate on the problem of the meaning of *hiisi*, a key notion of the Finnish folk religion. *Hiisi* has been considered, e.g., as a place of facing the holy in general ("holy grove"), or a place of a cult of the dead. Our goal is to test

Received xxx
Revised xxx
Accepted xxx

these hypotheses by investigating the occurrences of the name element *hiisi*, and name elements refering to the holy (*pyhä*), and to death (e.g., *kalma, kuolema*). We base our study on the 717,746 Finnish place-names in the National Place Name Register maintained by the National Land Survey of Finland.

Our main interest is finding evidence of the pairs of the name elements (*hiisi–pyhä*, *hiisi–kalma*) belonging to the same *idea complex*; i.e., that they, from the perspective of the members of the culture in focus, are closely associated with each other. We assume that significant attraction between name elements in a short distance, less than 3–4 kilometres, can be considered as evidence of the kind of semantic linkage.

Hence, we want to find methods for evaluating the significance of the attraction. The task is not straightforward. We introduce computationally intensive Monte Carlo methods for the purpose. Our analysis shows that the place-name data provide some support for the hypotheses, though no striking attraction was detected.

This paper is organized as follows. In Section 2 we describe the key notions and earlier work on them. The corpus and the distributions of the name elements of interest are described in Section 3. In Section 4 we first deal with the co-occurrences of the name elements in the corpus. After that we introduce computational methods for assessing the significance of them. In Section 5 we apply the methods to the occurrences of *hiisi*, *pyhä*, and *kalma, kuolema*. Section 6 discusses the results, and the restrictions and possible sources of error of the approach. Section 7 is a brief conclusion.

2. Key notions of *hiisi*, *pyhä* (holy) and *kalma, kuolema* (death)

When translating The Old Testament from Hebrew Bishop Mikael Agricola had to find Finnish expressions for a number of new things. The expressions were to be easy to understand for the still to a certain extent "pagan" Finns. Agricola made use of the Finnish word *hiisi* when translating *aeera*, the Cananian holy grove dedicated to the spouse of *Baal*. (Agricola 1931, e.g., 583; Deut. 16:21, II Kings 21:7, Is. 27:9). But what was *hiisi*, actually?

2.1. Holy grove

Distinguished areas in the forest, "holy groves" had a significant role as places of worship among several Finno-Ugric peoples. Maris, for instance, had (and still have) three kinds of forest shrines: the groves of *Jum* ("groves of the upper blessing") for celestial and nature deities, *keremet* groves, dedicated to certain subterranean deities ("groves of the lower blessing"), and the groves of the dead ones. In addition to the sky god *Jum* and his closest assistants, the numerous nature gods, which are rulers or "owners" of the natural phenomena were worshipped in the "groves of the upper blessing".

In a grove there was a sacrificial tree dedicated to each of the deities worshipped there, and these trees were standing in a long arched row at the Eastern or the Southern edge of the grove. The trees of the "Mother Earth" and the earth spirits, however, were located at the Western or the Northern edge of the area. The cult of these deities was, thus, addressed "downwards" unlike the rest of the

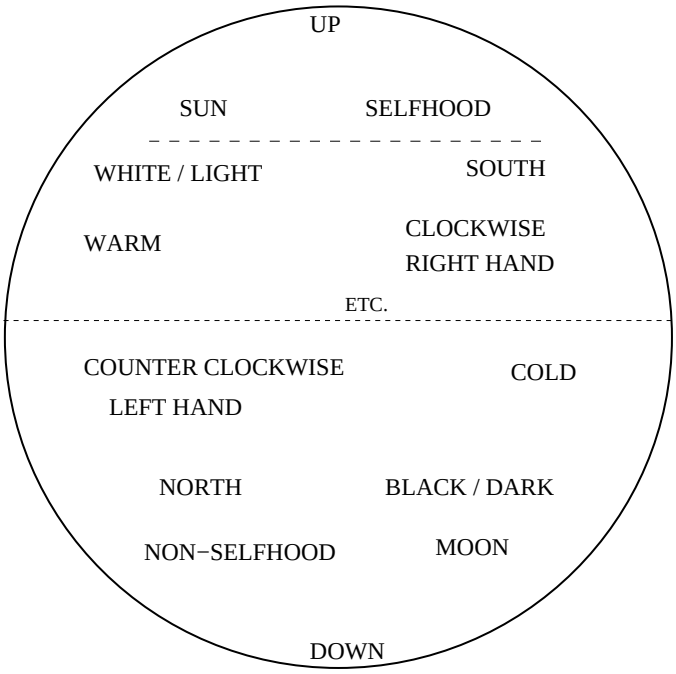


Fig. 1. Semantic field of the life (upper part), and the death (lower part). In the three-layer universe the fields correspond to the upper and the middle "worlds" (upper field), and to the lower "world", i.e., the realm of the dead (lower field).

cult that took place "upwards" (cf. Figure 1). The colour of the sacrificial animal was white and the sacrificial tree was a leafy one - linden, oak or birch. The cult in these groves took place in daytime, and it was open to both sexes. There were groves for smaller and larger cult communities (Holmberg 1914, 72-101).

The true nature of the *keremet* groves is more open to speculation. The name *keremet* itself is an old southern loan in Mari, and originally a Turkish word meaning 'holy'. Semantic elements linked to the *keremet* worship are opposite to the former. The sacrificial animal had to be black and the sacrificial tree was to be a coniferous one. The sacrifice took place in the evening, towards the West or to the North, "towards the night" and "downwards" – as the Maris say. The cult as well as the groves were taboo for women.

The cult in the *keremet* groves probably had something to do with the dead although the link is quite obscured. The cult was performed strictly in a family or kindred community. (Holmberg 1914, 102–113).

Although among the Maris the cult of the dead proper usually takes place in homes, it has been customary to sacrifice to the dead also in the cemetery groves, especially under a crisis. The sacrificial animal is a black ox that is eaten as a communion sacrifice in the cemetery (Holmberg 1914, 30).

Cult places corresponding to the three types of holy groves have been found among many Finno-Ugrian peoples. What is essential here, is the dichotomy between the groves of the "upper worship" and the groves of the "lower worship". The groves have to be seen as metaphysically different ones. Speaking about holy groves without further specification is, thus, misleading.

When it comes to the Finns, there is no doubt about the existence of the holy groves. Pope Gregory IX gave in 1229 an official letter by which he authorized the church to confiscate and destroy the holy groves of the pagan Finns. Many of the oldest churches in the province of Varsinais-Suomi were built on the sites of these groves or in the neighbourhood of them. Following the missionary instructions of the Catholic Church from the 8th Century the old pagan cult places were, thus, given new spiritual contents (Kuusi 1955, 147).

In the backwoods of Savolax there were, according to ecclesiastical authorities, holy groves still in the 18th century. For instance, the Vice-Minister of Kerimäki parish sent in 1738 a letter to the bishop's council in Porvoo telling about the "destroyal of suspicious groves and trees" (Siikala 1985, 341). In the Eastern Orthodox area in Karelia and Ingria holy groves were left untouched until very lately because the old cult had become entangled with the Christian elements into a kind of syncretism, which, in practice, was approved by the church.

2.2. Hiisi

But what was the true nature of the Finnish *hiisi*? Was *hiisi* something parallel to the holy groves of the other Finno-Ugrian peoples? Researchers have discussed these questions a lot without consensus.

To Academician Martti Haavio, *hiisi* was a place to meet the Holy, and, further, he identifies this holy grove with, e.g., the Paradise of Genesis. Haavio's point of view is an Eliadean one, in his phenomenology of religion he sees the holy grove as a manifestation of a prioric Holy. However, as shown by Tarkka and Anttonen, Haavio has a historical-sociological point of view along with the phenomenological one. (Tarkka 1987, 64-65; Anttonen 1996, 119). According to Haavio the common cultic activity in the holy grove maintained the profane social order of the community.

Anyway, in the *hiisi* groves sacred tales inherited from the fathers were solemnly recited in order to strengthen the original and good world order. The tales told about the origin of the world and the means of livelihood. Naturally, along with the development of the means of livelihood, new religious ideas became stratified on the basis of this original idea of *hiisi*, the most important of which was the cult of the dead. Under the late paganism, at any rate, *hiisi* probably was the "dwelling place of the good-willing departed ones" (Anttonen 1996,123).

The ideological background of the cult of the dead is the idea of the local realm of the dead that had become prevailing after the breakthrough of agriculture (Siikala 1992, 115-117). The dead were now living in the neighbourhood of the living ones while they in the earlier hunting community had been sent to the distant realm of the dead, a place somewhere in the North and behind the river of Tuonela. The dead became the guardians of the moral code and proper behaviour in general.

Today *hiisi* is more commonly interpreted as the centre of the Iron Age Finnish cultic community, a village or a clan, where the cult of the dead was practiced. According to this view it was not until then that the entire idea of *hiisi* was launched among the ancient Finns. As Veikko Anttonen writes on the basis of Mauno Koski's research:

> The Iron Age cemeteries were formed to the ritual centres of limited inhabited areas, in

Finnish *pitäjäs* [- - -]. According to the modern etymological research the name *hiisi*, based on a German loan word, was linked to these centres. (Anttonen 1996, 116)

The fact that not all the *hiisi* places were ancient cemeteries is explained by Koski, who suggests that later *hiisi* was generalized to refer to any anomalous nature place. Anttonen himself represents a third point of view according to which both *pyhä* (holy) and *hiisi* were primarily territorial border terms. According to Anttonen, *pyhä* was localized into the border areas of the inhabited and uninhabited territorial spheres, whereas *hiisi* was primarily located inside the inhabited areas.

Later on *hiisi* was made a symbol of territorial inner borders systems, which were of social and economic nature. Anttonen links this further to *pyhä*. He explains that along with the process in which activities associated with issues of ownership and usage of nature resources were concentrated to *hiisi* places, the notion of *pyhä* was also transformed. It lost its nature as a territorial border term and got a new, more abstract and spiritual meaning (Anttonen 1996, 122-123). Also in this third case it is natural to think that during the Iron Age *hiisi* finally attracted the then prevailing cult of the dead.

Irrespective of what line of explanation we choose to follow, along with the Christianization *hiisi* came to mean "the dwelling place of hostile beings" (Anttonen 1996, 123). This finally resulted in a profound confusion of concepts. For instance, according to Ganander *hiisi* meant Hell (Ganander, 1984, 14). *Hiisi* was also personified and it came to mean a giant as an aetiological being, an evil goblin etc.

It has to be discussed, thus, whether *hiisi* originally meant an extraordinary place to face the holy (in the Eliadean sense of the notion), i.e., a holy grove in a general meaning, or whether it has been primarily an extraordinary place for the cult of the dead. Still another possibility is that *hiisi* was transformed from a territorial term into a cultic one, the cult of the dead having been adopted into its sphere secondarily. Perhaps ancient Finns had two or three kinds of holy groves as the Maris still do.

Although there is no doubt about the cult of the dead on the Finnish territory during the Iron Age, some facts speak against the interpretation that *hiisi* was a place for the cult of the dead. In Estonian *hiis* is generally a "holy grove" without any special relationship to the cult of the dead. *Aeera*, the Cananian holy grove dedicated to the spouse of Baal, which Agricola translated *hiisi*, was definitely not a place for the cult of the dead but for ritual practices aiming at assuring and strengthening the fertility of the land, cattle and man. Agricola may have thought of rituals like drinking the *Ukon vakka*, and drinking beer in honour of the thunder god *Ukko* to get more rain at the right time. In that connection Agricola emphasizes the licentiousness associated with the ritual. On the other hand, Agricola wrote in the foreword of the very same translation (Agricola 1931, 212-213):

```
Tapio / Metzest Pydhyzet soi /
ia Achti / wedhest caloia toi.
          [ - - - ]
Hiisi / Metzeleist soi voiton /
Wedhen Eme / wei Calat wercon.
```

(Tapio / gave the hunting game / And Achti / the fish from the water. Hiisi / gave victory over the big game / Wedhen Eme [Mother of the Water] / brought the fish into the nets.)

According to this quotation *Hiisi* was the master of the game animals among the Eastern Finns, corresponding to *Tapio* of the Western Finns.

On the other hand, *Tapio* was primarily the forest itself, its personification or roundabout expression. If *hiisi* had originally [and also] been a grove, an extraordinary place in the forest, semantic shifts like this become easy to understand. Without doubt people living in a hunting culture asked for success in hunting in their holy groves. It is difficult to see why Agricola would have used the term *hiisi* in two totally different meanings.

In this study we address the question of the meaning of *hiisi* by analyzing the Finnish place-names appearing on the 1:20,000 Basic Map. In the case that *hiisi* and the cult of the dead were semantically linked, we should find signs of attraction between place-names containing *hiisi* as a name element, and the place-names with a name element referring to death (such as *kalma, kuolema*). We also test the hypothesis of attraction in the case that *hiisi* was a holy grove in a general meaning by investigating the place-names with a name element meaning the holy (*pyhä*).

In the next section we describe the corpus, and the occurrence of the key terms *hiisi, pyhä*, and *kalma, kuolema* in the corpus.

3. Occurrence of *hiisi*, *pyhä* and *kalma, kuolema* in the corpus

We investigate the Finnish names occurring in the National Place Name Register, a part of the Geographic Names Register kept by the National Land Survey of Finland. The register contains all place-names that appear on the 1:20,000 Basic Map in electronic format, and it is maintained for the purposes of creating these maps. In addition to 717,746 Finnish place-names, the data include the coordinates of the locations, types of the objects, and the names of the municipality and the region where the object is located. The main types include, e.g., natural, water system, and cultural names with several subtypes. Below we do not separate between the natural and the water system types. The term 'natural' is used for refering to the locations that are of type natural or water system in the original classification.

Each object is represented by one pair of coordinates. This is, of course, unrealistic in case of large lakes, long rivers etc. However, the number of very large entities among all the locations is small. Thus, the pointwise representation is unlikely to cause significant distortion.

The name element *hiisi* occurs in 690 names. In 572 cases the named object is of type natural, and the remaining 118 cases are cultural names. The base forms of the name elements are *hiisi* (497 cases natural /73 cultural names), *hiis* (71/47), *hiie* (11/10), *hiije* (1/0), *hiisilä* (5/3), *hiitola* (22/13), *hiitelä* (7/0). Out of 690 occurrences 27 are in the plural (*hiitten*). These names are excluded from the analysis since the plural form refers to the naming having taken place at a later time, when the notion of *hiisi* did no longer have its supposed original meaning (see the previous section). Figure 2, left panel, shows the locations of the remaining 673 occurrences.

Further, we have 384 occurrences of *pyhä*, 285 of them nature type, and 99 cultural locations. The basic forms are *pyhä* (267/83), *pyhi* (2/0), *pyhinöinen* (1/0), *pyhiö* (1/0), *pyhät* (such as *pyhätlammi)*) (4/0) , *pyhänen* (2/0), *pyhäs* (3/0), *pyhälys* (2/0), *pyhälystä* (1/0), *pyhälö* (1/0), *pyhälöinen* (1/0), *pyhälä* (0/16). We excluded forms *pyhälä*, 16 occurrences all of which are cultural names, mainly

Fig. 2. Occurrences of *hiisi* (left), *pyhä* (middle), and *kalma, kuolema* (right).

buildings, as well as *pyhitys, pyhitty*, which most likely are of Christian origin. Figure 2, middle panel, displays the locations.

The following name elements were considered as referring to death in a useful sense from the point of view of our study. The total of 541 occurrences, 508 of them type nature, include the following words as basic forms of the name elements: *kalma* (140/7), *kalmo* (214/12), *kalmoo* (6/0), *kalme* (4/0), *kalmee* (0/2), *kalmi* (1/0), *kalmu* (21/4), *kalmas* (3/0), *kalmus* (3/0), *kalmos* (1/0), *kalmukki* (1/0), *kalmut* (2/0), *kalmot* (5/0), *kalmonen* (1/0), *kalmila* (0/1), *kalmala* (0/1), *kuolema* (84/2), *kuolemainen* (10/1), *kuolemais* (1/0), *tuoni* (8/1), *tuonela* (1/2), *manho* (2/0). Figure 2, right panel, shows the locations.

4. Evaluation of attraction

Our major interest is finding evidence for attraction between pairs of name elements. We first give a summary of distances between the place-names including one of the name elements *hiisi, pyhä*, and *kalma* or *kuolema*. After that we formulate the hypothesis of attraction between name elements more accurately and discuss methods for testing the hypothesis.

4.1. Co-occurrences of *hiisi*, *pyhä* and *kalma, kuolema*

Table 1 gives a summary on the distances between the name elements *hiisi*, *pyhä* and *kalma*. For instance, the table shows that there are 22 occurrences of *hiisi* from which the closest occurrence of *pyhä* resides in a distance of less than 2000 metres, and 26 occurrences of *pyhä* such that the closest occurrence of *hiisi* is located within 2000 metres.

It is not very easy to assess whether these numbers are significantly large or

Table 1. Co-occurrences of *hiisi, pyhä* and *kalma, kuolema*.

max distance	*hiisi* and *pyhä*	*pyhä* and *hiisi*	*hiisi* and *kalma*	*kalma* and *hiisi*
500 m	1	1	0	0
1000 m	7	8	8	7
1500 m	13	16	17	16
2000 m	22	26	28	25
2500 m	29	32	46	38
3000 m	37	34	65	51
3500 m	43	38	76	60
4000 m	52	41	90	66
4500 m	61	56	107	73
5000 m	75	67	123	86

not. The assessment should, of course, be related to, e.g., the total numbers of occurrences of the name elements, and the common area of distribution of them.

The word "significantly" implicitly assumes some standard of comparison. The numbers being significantly large means that they are large compared to some other numbers derived from some expected conditions. We now move on to introducing methods for evaluating (statistical) significance of the co-occurrences.

4.2. Randomness as a standard of comparison

It is often difficult to verify whether a tendency or regularity appears in data. We illustrate this by an example. In the left panel of Figure 3 there are 100 points. Is there regularity in the distribution of the points? The human eye seeks structure, and indeed, it seems that there are different kinds of areas in the figure: empty regions, areas with slight accumulation of points, and areas between these extreme cases. In reality, however, the location of each point was generated randomly and uniformly in the sense that choosing any location for a point was equally probable, and the points were generated independently of each other. Thus, the clusters of points, and the empty areas are due to pure chance.

In addition to the 100 dark points of the left panel, the right panel of the figure includes 50 white points. The white points are generated randomly as well, and independently of the locations of the black points. Thus, any closeness of a dark and a white point is due to chance.

Randomness is often adopted as a conservative *null hypothesis* in data analysis. The key idea is to investigate whether the observations could be explained by pure chance with any reasonable probability. For instance, according to Table 1 we have 37 occurrences of *hiisi* from which the closest occurrence of *pyhä* resides at a distance of less than 3000 metres. Is it possible, with any reasonable probability, that the explanation is chance; that is, as we have 690 cases of *hiisi*, and 384 cases of *pyhä*, some of them happen to occur close to each other, and thus, there is no need for an explanation based on, e.g., common meaning.

Considering randomness as the null hypothesis can be seen as applying the famous principle of *Occam's Razor*, that is, rejecting unnecessary explanatory factors. If we can show that the probability of observing the data is very small, when assuming pure chance as the explanation, we can conclude that another reason for the observations should be found.

We can utilize computers in assessing the probability by generating fictive

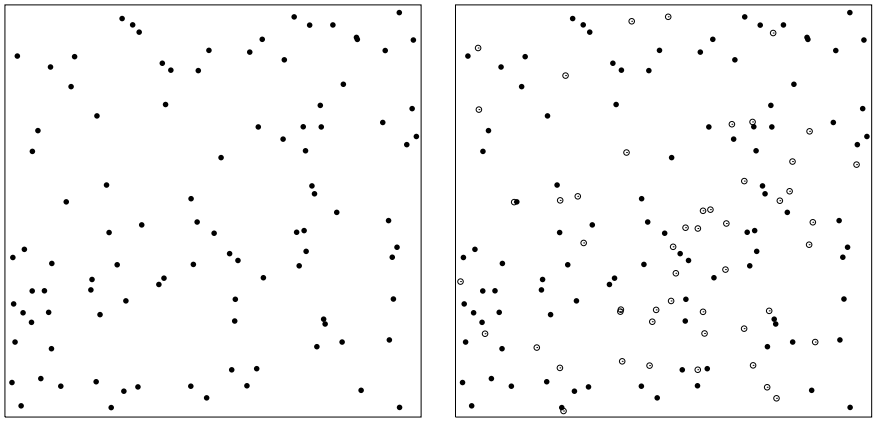


Fig. 3. 100 randomly generated points (dark circles), and 50 randomly generated points (white circles).

data which we know to have random origin, analoguously to the situation of Figure 3, where the points were generated by a computer uniformly and independently of each other.

To continue with the example of *hiisi* and *pyhä*, we can, for instance, generate 384 occurrences randomly, and compute how many occurrences of *hiisi* has a generated point within less than 3000 metres. If we obtain a number close to 37, we have shown that chance can explain the number of co-occurrences of *hiisi* and *pyhä*.

Of course, to make the conclusions more reliable, we need to generate more than one set of fictive data points. The larger the number of similar trials is, the more confident the results are. These kinds of procedures are called Monte Carlo tests, the name referring to the connection of the well-known casino and exploitation of random numbers (see, e.g., Cohen 1995).

Complete spatial randomness The simplest version of the hypothesis of randomness is assuming *complete spatial randomness* (CSR) (see, e.g., Bailey and Gatrell 1995). This was the case in the example of Figure 3. Under the condition of CSR each data point is generated uniformly randomly within the area of investigation; i.e., any location has equal chances to be selected. All the generated points are independent of each other.

For our purposes the CSR hypothesis has at least the following drawbacks. First, name elements do not usually occur everywhere, and the frequency of occurrence varies in different areas. Often the reason is that occurrence is connected with the distribution area of a specific dialect.

Second, names in general are distributed unevenly. There are more names in the south than in the north, for instance. In addition to large-scale variation, remarkable local variation appears due to different geographical or cultular conditions. Figure 4 illustrates this by showing all the place-name occurrences in the corpus in a part of Finland (surroundings of Oulu). Rivers are discerned from the surroundings with higher intensity. Lake Oulujärvi is the white area in the bottom right square of the figure. Very white areas along the coast in the bottom

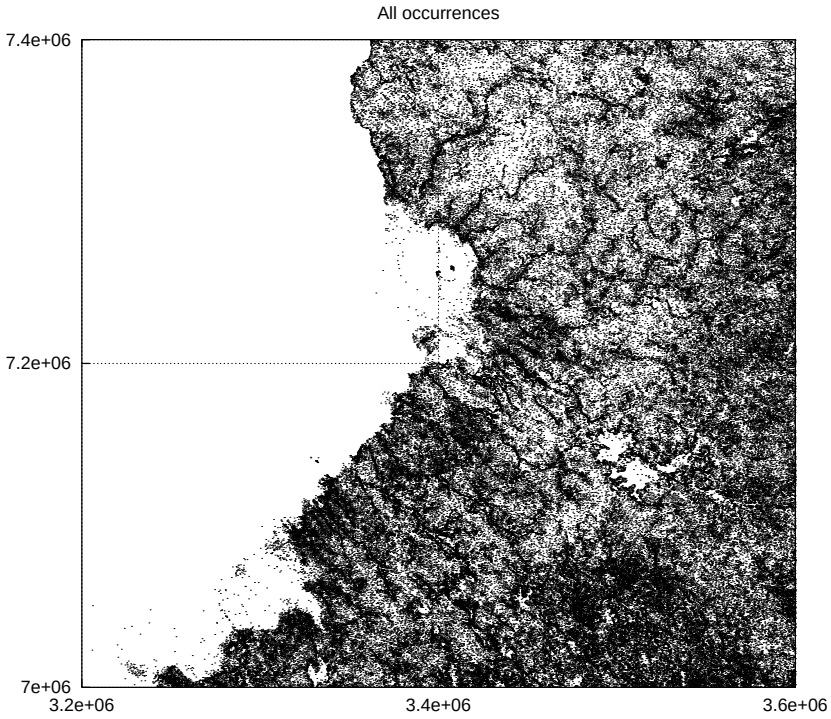


Fig. 4. All place-name occurrences of the corpus in the surroundings of Oulu.

left are the Swedish-speaking areas of Ostrobotnia with remarkably less Finnish names.

Completely random generation of fictive data does not, however, make any separation between different areas. The implied problem is that pairs of different name elements may reside closer to each other than predicted by the CSR hypothesis, because they both occur more frequently in the regions of high overall intensity, e.g., along rivers, and less frequently in the regions of low intensity, e.g. in the middle of lakes.

Third, since we are testing the attraction of two *different* name elements, we should, instead of complete randomness, allow accumulation of occurrences of the same name element in generated data. In reality, if a place is named with a name element, there are often other objects within a short distance named with the same name element (e.g. *Pyhäjärvi, Pyhäniemi* etc.) For instance, in 44 % of the occurrences of *pyhä*, and 41 % of those of *hiisi*, there is another occurrence of the same name element within 2500 metres These proportions are high, compared against the corresponding proportions of any name elements. The accumulation is, of course, what one expects, since *pyhä* and *hiisi* are name elements that were connected to holy or extraordinary places. Accordingly, they imply that the place and its immediate surroundings had a very special status.

In the next section we modify the condition of CSR in order to obtain a more confident standard of comparison for attraction between name elements. Before

the problems of our actual interest, attraction of *hiisi*, *pyhä*, and *kalma, kuolema*, we take a look at an introductory case for illustrating the methods.

4.3. Introductory case: attraction between *ukko* (old man, also thunder god) and *akka* (old woman)

We start with an introductory problem – the attraction between name elements *ukko* (old man, thunder god) and *akka, akko* (old woman) – for two reasons. First, the correct solution looks, intuitively, quite obvious. Thus, we can compare the results given by our method to the intuitive answer. Second, there are a lot of data points available, and they are smoothly distributed. This helps to overcome some difficulties in applying the method.

There are 1385 occurrences of *ukko* and 693 occurrences of *akka* in the corpus. The occurrences are shown in Figure 5.

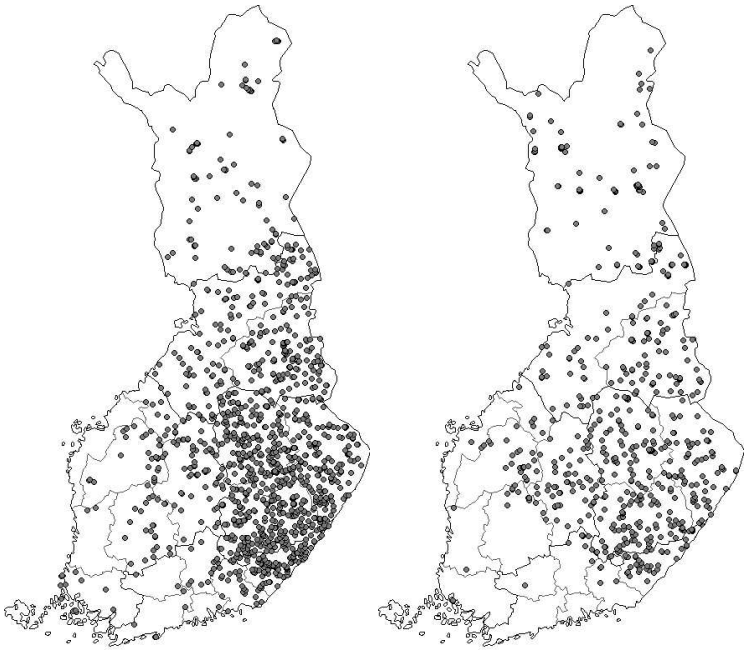
We address the three problems of CSR stated in the previous section. First, we take into account such areas only where both the name elements under investigation occur relatively evenly. For this purpose, we search for areas where for each name element of one type there is at least one occurrence of the other type within a pre-defined distance. An algorithm for this task was introduced by Ester et al. (1996).

Figure 5 (bottom) shows a more detailed picture of the occurrences of the largest cluster with the pre-defined maximum distance being 15 kilometres. This area includes 752 occurrences of *ukko* and 358 occurrences of *akka*. It covers roughly the provinces of South and North Karelia, North Savolax, and most of South Savolax and Central Finland. The background of the figure indicates the locations of all the place-names within at most 15 kilometres from an occurrence of *ukko*. Now, by choosing the fictive points from the background area, we guarantee that any random point resides, similarly to the real occurrences of *akka*, in a distance of at most 15 kilometres from the closest occurrence of *ukko*.

Instead of assigning equal probability to any pair of coordinates, we make use of the distribution of all the place-names, that is, the 240,308 dots of the background. If we now generate the fictive sets of occurrences of *akka* by choosing randomly any of the 240,308 points, our random data points are not distributed completely randomly spatially but they follow the distribution of all the place-names. This approach is based on the plausible assumption that the distribution of all the place-names provides a better approximation for the overall intensity of occurrences of *akka* than the spatially uniform CSR.

Accumulation of the occurrences of a single name element may still be a problem in using the distribution of all the place-names. In our example, let us consider a group of three occurrences of *ukko* such that there is another occurrence of the group within a "small" radius, 1500 metres, for instance. Further, let us assume that within 2 kilometres from all the three there is an occurrence of *akka*. It is very unlikely that the *ukko* names are independent of each other, and, thus, recording three occurrences with an occurrence of *akka* within 2 kilometres results in a misleading amount of close occurrences.

Accordingly, the following modification, though not very elegant from a theoretical point of view, can be accepted for practical purposes with good reason. We only count once the occurrences of a single name element within a "small" pre-defined radius; that is, we compute the distances to the closest occurrence of the other name element from each of the occurrences of the group, and select



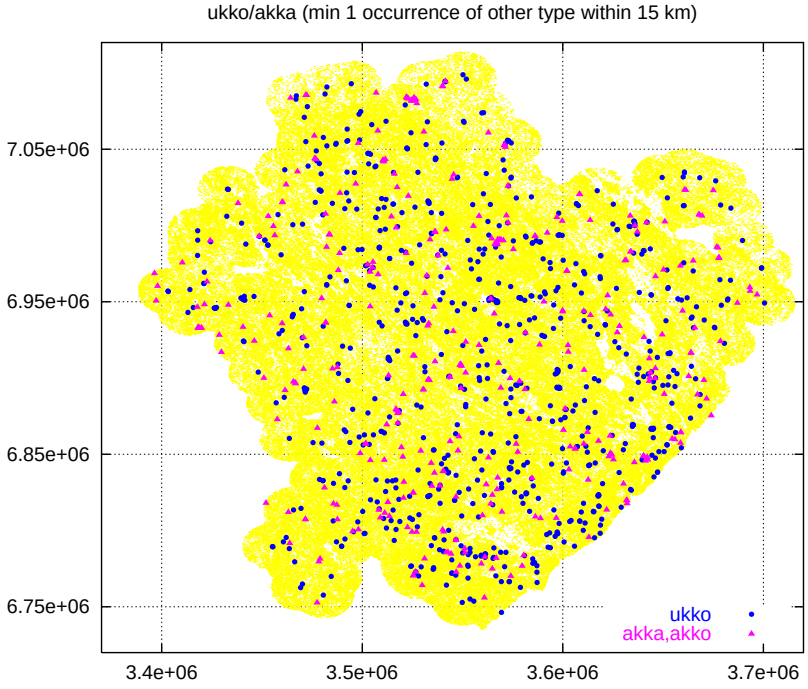


Fig. 5. Occurrences of *ukko* (top left) and *akka, akko* (top right). Largest cluster where each occurrence has an occurrence of the other name element within a distance of 15 kilometres (bottom).

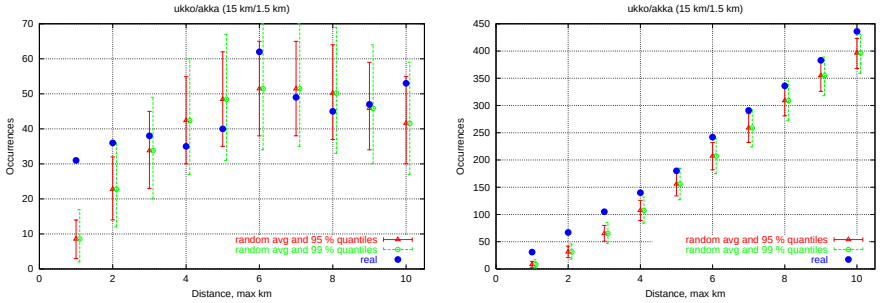


Fig. 6. Summary of closeness of *ukko* and *akka* in the largest cluster. Real distances are indicated by the points. Bars show the range where 95 %/ 99% of the distances settled when, instead of the real occurrences of *akka*, fictive (computer-generated) occurrences were used.

the minimum distance. Further, we generate only one fictive point for a group of occurrences within the radius. This induces a small error in the results. We conduct trials with different values of the radius for testing the sensitivity of the results for selection of the value.

Figure 6 gives a summary of the results. The points indicate the numbers of occurrences of *ukko* with the nearest occurrence of *akka, akko* at a distance of less than 1 km, 1–2 km, etc. We see, for instance, that there are a little more than 30 occurrences within a distance of 1 km, a little less than 40 occurrences at a distance of 2–3 km (left panel), and approximately 100 occurrences within 0–3 km (right panel). We generated 10,000 random data sets of 358 occurrences of *akka, akko*. Similarly to the real data, we computed for each randomly generated data set the numbers of occurrences of *ukko* with the nearest occurrence of *akka, akko* at a distance of less than 1 km, 1–2 km, etc. Small triangles in the middle of bars in the figure show the average of these values in the 10,000 trials. The bars indicate the range of values where 95 % (solid bar), and 99 % (dashed bar) of the values computed from random data settled.

Now, comparing the values computed from the real data against those from the fictive data clearly suggest that the number of co-occurrences within 1 km cannot be explained by pure chance. It is obviously (practically) impossible to obtain randomly the degree of closeness of occurrences of the real data. There are also many occurrences with a distance between 1 and 2 kilometres. The most extreme results from random trials, however, match the number computed from the real data. When investigating the distances between 2 and 3 kilometres and so on, no significant differences between the real and fictive occurrences can be detected.

5. Attraction of *hiisi*, *pyhä*, and *kalma, kuolema*

Next, we apply the methods described in the previous section to studying the attraction between name elements *hiisi* and *pyhä*, and *hiisi* and *kalma, kuolema*. In this section we concentrate on describing the results of the trials, while the next section discusses the interpretation of the results in more detail.

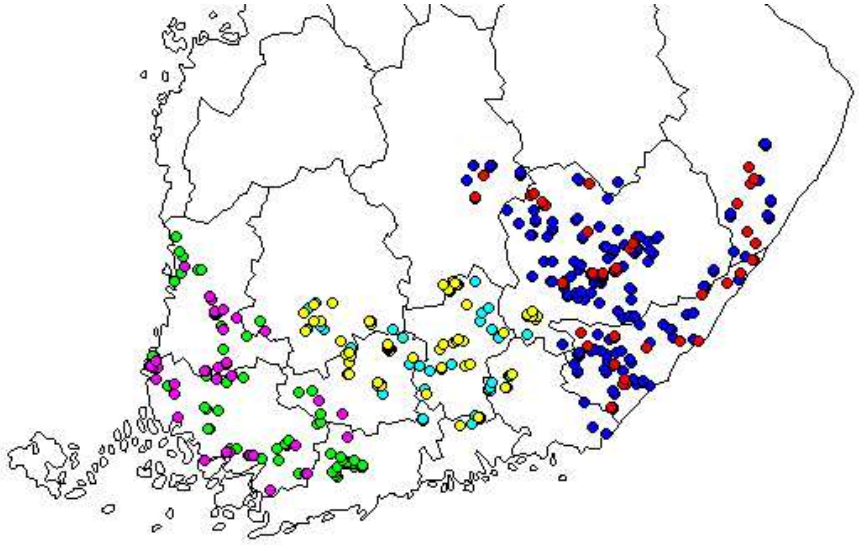


Fig. 7. Largest clusters of *hiisi* and *pyhä*, min 1 occurrence of the other type within 25,000 metres. In the easternmost cluster: hiisi (blue), pyhä (red), middle cluster: hiisi (blue), pyhä (yellow), westernmost cluster: hiisi (green), pyhä (violet).

5.1. *Hiisi* and *pyhä*

The emphasis of the overall distribution of *pyhä* is in the south-west, whereas the intensity of occurrence of *hiisi* is strongest in the south-east (see Figure 2).

We searched for common areas of occurrence as described in the previous section. Figure 7 displays the three largest groups for name elements *hiisi* and *pyhä* such that for each occurrence at least one occurrence of the other name element resides within 25 kilometres. The largest cluster in the south-east, which covers roughly the provinces of South Karelia, South Savolax, and a part of North Karelia, includes 238 (188 *hiisi* + 50 *pyhä*) occurrences, the second largest (Päijät-Häme, Kanta-Häme, Pirkanmaa, the westernmost part of Kymenlaakso) consists of 118 (51+67) occurrences, and the third largest (Varsinais-Suomi, Satakunta, the municipality of Vihti in Uusimaa) in the south-west, has 101 (57 + 44) occurrences. In the east the intensity of the occurrence of *hiisi* is almost four-fold compared to that of *pyhä*, whereas in the middle cluster, 55 % of the occurrences are of *pyhä*. In the westernmost cluster the proportion of *pyhä* is 45 %. Thus, the proportion of the densities of occurrence is clearly different in the east.

We first analyzed the three clusters together (i.e., fictive points were generated separately in each cluster, the resulting distances from each cluster were combined when computing the average distances, 95 % and 99 % quantiles, and the real distances). Figure 8 (top) shows a summary of the trials. Moderate attraction can be observed in short distances. Differences were small when using 1500 m instead of 2500 m as the threshold valued for considering occurrences of a single name element as separate occurrences, or when generating random occurrences of *hiisi* instead of *pyhä*.

The summaries of the co-occurrences in the real and random data in each

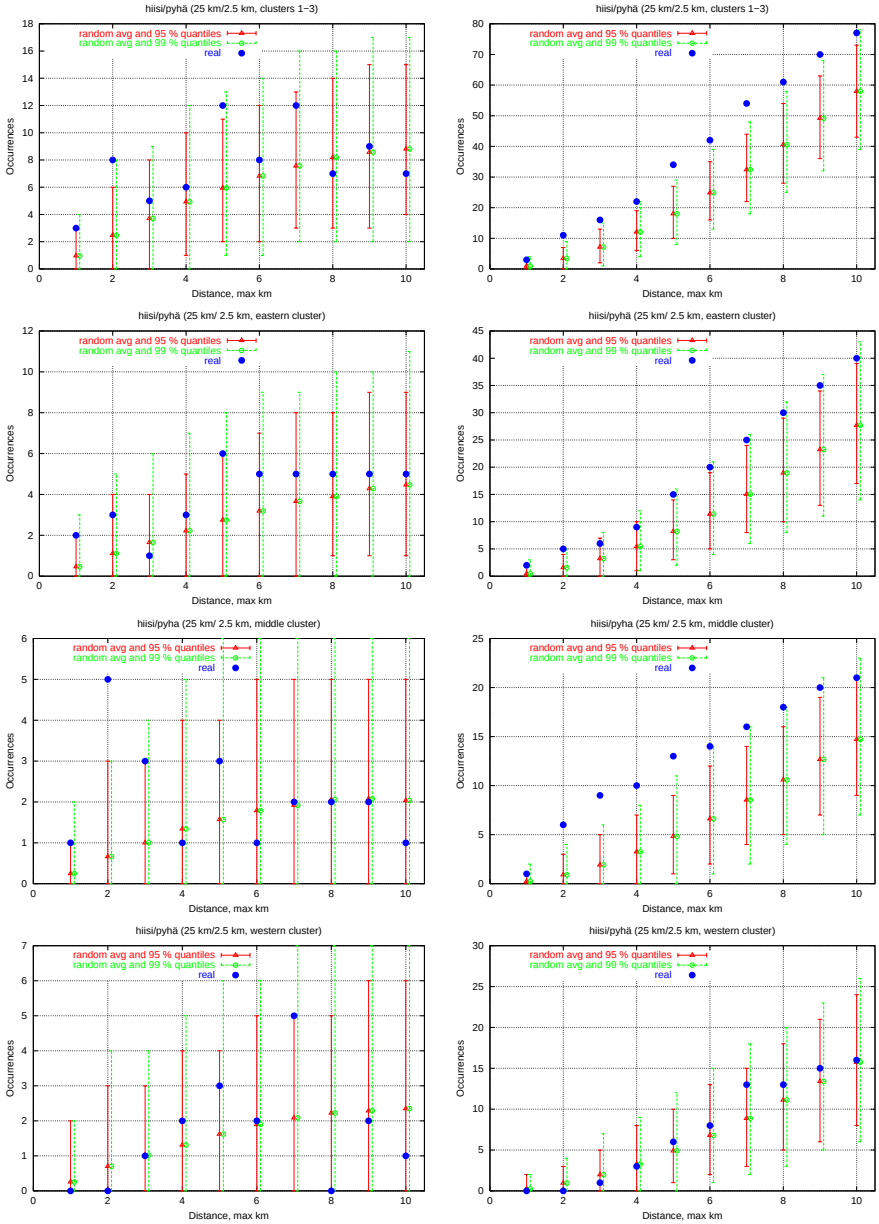


Fig. 8. Summary of closeness of *hiisi* and *pyhä*; three largest clusteres combined (top); eastern cluster (second row); middle cluster (third row); western cluster (bottom). Real distances are indicated by the points. Bars show the range where 95 %/ 99% of the distances settled when, instead of the real occurrences of *pyhä*, fictive (computer-generated) occurrences were used. Cumulative numbers of occurrences are shown in the right-hand panels.

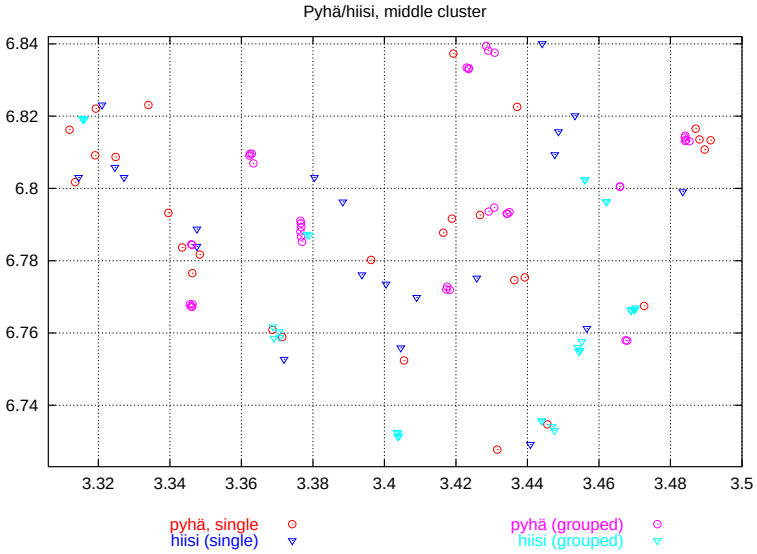


Fig. 9. *Hiisi* (circles) and *pyhä* (triangles), occurrences in the middle cluster, covering (roughly) the provinces of Kanta-Häme, Päijät-Häme, Pirkanmaa, and westernmost parts of Kymenlaakso. Occurrences with another occurrence of the same name element at a distance less than 2500 m are indicated by different colours. The squares in the background of the figure are of size 20 x 20 km.

cluster separately are also shown in Figure 8. In the eastern cluster the results indicate slight attraction, which might be accidental, though. In the middle cluster the attraction of the name elements is strong. The occurrences of this cluster are shown in detail in Figure 9. We observe that there are several co-occurrences in the southern and western parts of the area. However, in the western cluster no attraction can be detected. There is only one (!) occurrence of *hiisi* that has an occurrence of *pyhä* within less than 3 kilometres.

5.2. *Hiisi* and *kalma, kuolema*

Figure 2 reveals that the core of the distribution area of *kalma, kuolema* resides in Eastern Finland close to that of *hiisi*. However, the emphasis of occurrences of *hiisi* is slightly more south-western than that of *kalma, kuolema*. This difference complicates the analysis in the south-east, due to the difficulty of finding areas of even distribution. The largest area that contains at least one occurrence of the other type within 15 kilometres includes 235 points: 134 occurrences of *hiisi*, and 101 occurrences of *kalma, kuolema*. The area covers the provinces of South Savolax and South Karelia. Figure 10 presents the data.

Since the other clusters were small when using the distance of 15 kilometres, we applied the clustering again with the distance of 25 kilometres, excluding the 235 data points obtained in the first round. This yielded a relatively large cluster consisting of 616 points: 323 occurrences of *hiisi* and 293 occurrences of *kalma, kuolema*. It covers a large part of Finland from Varsinais-Suomi to Kainuu. The occurrences are shown in Figure 10.

Summaries of the trials on both the clusters separately are shown in Figure 11.

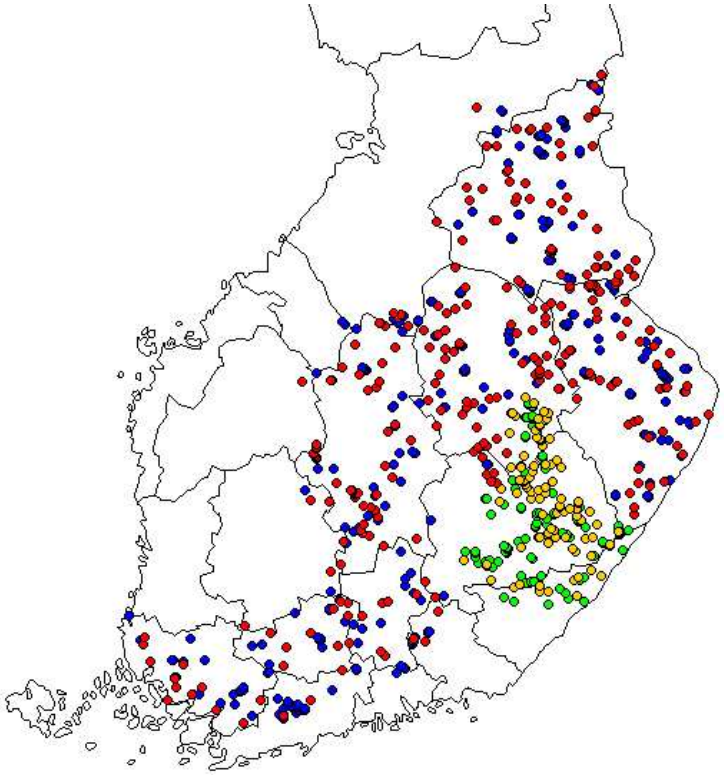


Fig. 10. Occurrences of *kalma, kuolema* (yellow), and *hiisi* (green) in the largest cluster with the maximum distance of 15 km to the closest occurrence of the other name element; *kalma, kuolema* (red), and *hiisi* (blue) in the largest cluster with the maximum distance of 25 km, the occurrences of the first cluster excluded.

We do not combine the clusters, since the areas overlap slightly (see Figure 10). In the south-eastern cluster the real values exceed the averages of the randomized trials when it comes to the distances between 1 and 3 kilometres. The results from the trials on the other cluster indicate that there are relatively many occurrences with a distance of less than 1 km, and, in particular, 2–3 km in the real data. This causes the cumulative number of occurrences to exceed the 99 % interval when investigating the distances less than 3 kilometres. Instead, in the case of 0–2 km the real value does not even reach the higher end of the 95 % interval.

6. Discussion

The distribution of occurrences of *hiisi* is concentrated to South-Eastern Finland, the provinces of South Savolax and South Karelia. *Hiisi*, at any rate, is a Baltic Finnic word known also in Estonian. Thus, the emphasis of the distribution needs to be explained.

Based on Bishop Mikael Agricolas list of Finnish gods in the foreword of the translation of the Old Testament (see Section 2.2), we could speculate on some eastern connotation (master of game animals?) that *hiisi* may have had. Another,

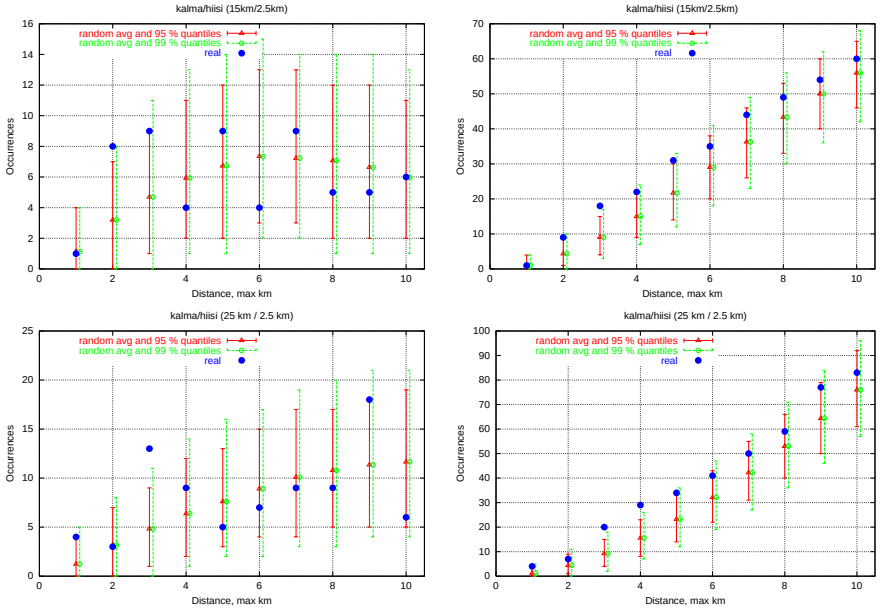


Fig. 11. Summary of closeness of *hiisi* and *kalma, kuolema* in the south-eastern cluster (top), and the cluster covering a large area from Varsinais-Suomi to Kainuu (bottom). Real distances are indicated by the points. Bars show the range where 95 %/ 99% of the distances settled when, instead of the real occurrences of *hiisi*, fictive (computer-generated) occurrences were used. Cumulative numbers of occurrences are shown in the right-hand panels.

perhaps more convincing, explanation for the larger number of occurrences in Eastern Finland could be the weaker and later ecclesiastical control in the east compared with that in the west.

When it comes to the attraction of *hiisi* and *pyhä* the differences are dramatic between the – for methodological purposes formed – middle and western clusters (see Figure 7). While we detected a remarkably strong attraction in the middle cluster, no attraction whatsoever could be found in the western cluster.

The border line between the clusters seems to be about the same as the border between the historical provinces of Varsinais-Suomi and Häme. There are 18 occurrences of *hiisi* in the small *Hiidenvesi* area belonging to the western cluster, but actually, this area was a part of ancient Häme.

The Christian church gave a new content for the concept of holy, *pyhä*, and it is natural to think that *hiisi* names, at least close to *pyhä* names were removed. It is tempting to explain the difference of the clusters by the historical fact that the inhabitants of Häme made hard resistance to the new religion, even returning to paganism in a rebellion in 1237 (Salo 2003, 292-293). This would be an oversimplification, however. According to Salo (2003, 286) at the end of the 12th century the Suomi diocese only consisted of the parish of Kalanti and a part of the province of Satakunta. During the following century the diocese was gradually expanded in several phases. The border of the province is not confluent with that of the diocese in any of the phases (see Salo 2003, map on p. 291). It is, thus, difficult to say whether the differences between the clusters tell more about the

regional understandings of *hiisi*, or about differences in historical backgrounds. Further study on a more regional level is one of the future challenges.

The overall results with the three main clusters combined indicate a significant attraction between *hiisi* and *pyhä*, though far from the significance of the attraction in our introductory example, *ukko* and *akka*. In case of *hiisi* and *kalma, kuolema* we detected attraction as well, but, again, no striking effect was found. In the trials on the occurrences of Southeastern Finland, an assumption of the method, namely, a rather even distribution of at least one of the name elements, was not met. Thus, the results in that case are not as confident as in the other cases.

One of the future challenges is to expand the investigation to studying Anttonen's theory on *hiisi*. We excluded it from this study, because testing the theory would have required a large amount of preliminary linguistic work, which was beyond our expertise.

Restrictions and possible sources of error There are several types of issues to be considered as possible sources of error: we discuss those concerning the data first, next those concerning the method, and finally those concerning the interpretation.

Most of the names are not very old, and this, of course, obscures the results. The data set does not include even all the current Finnish names. It is, however, a large sample; according to Kiviniemi (1990), the total number of Finnish names is approximately 2.5 millions, including the areas, e.g., in Karelia, nowdays belonging to Russia.

An interesting question is whether the names of the corpus are biased. In other words, how realistic is it to assume that the corpus is a sample from the set of all Finnish names such that all name elements occur in the corpus with the same proportions as in reality? This question is closely related to the discussion above; the ancient folk religion was a false religion from the perspective of Christianity. Thus, the place-names, such as *hiisi*, which were closely related to the cult of the folk religion were subject to substitution for new Christian or more neutral names.

Further, it is not straightforward to select all the relevant place-names from the available data. Neither are all the selected names correct. Linguistic expertise is needed for the task. This study is preliminary in the sense that it should be supplemented by a detailed linguistic analysis of the names (and the named objects) included in the study, as well as those excluded from it.

The applied computational methods unavoidably need to rely on assumptions. The assumptions can never be absolutely correct, but the results may still be useful. The restriction posed by the requirement of finding clusters of relatively smooth distribution could be relaxed by employing more complex statistical modeling. The problem caused by the accumulation of the occurrences of a single name element could also be treated that way [1]. More complex approaches

[1] More precisely, the density of occurrences of the name element to be randomized could be modeled as a spatial point process, e.g., as a heterogeneous Poisson process, the intensity of the process being estimated based on the observed occurrences. The kernel density estimation, for instance, could be applied for the task. The estimate could then be used as the distribution from which the fictive data are drawn. The problem of accumulation could be approached by modeling the occurrences, e.g., as a clustered Poisson process. For spatial point processes, and the kernel density methods, see, e.g., Cressie 1993, Bailey and Gatrell 1995.

are not free of assumptions, either. They may also raise some new problems (e.g., the requirement of edge corrections near the border of the country).

Finally, the introduced methods can only be used for detecting attraction. Even if attraction of name elements is clearly observed, the question of the reason for it remains. The observed attraction can be explained by another variable, not the semantic connection of *hiisi* and *pyhä*, or *hiisi* and *kalma*.

7. Conclusion

Computational methods may provide a significant set of tools for addressing some research problems in the field of folk religion. This viewpoint is probably not reachable by other means. The results, of course, must be supplemented by other methods. In any case, a lot of expert knowledge on the many relevant research fields, e.g., comparative religions, linguistics and history, is needed.

References

Agricola, M. (1931), Mikael Agricolan teokset, III. Helsinki: WSOY.

Anttonen, V. (1996), Ihmisen ja maan rajat. 'Pyhä' kulttuurisena kategoriana. (Suomalaisen kirjallisuuden seuran toimituksia 646). Helsinki: SKS.

Bailey, T. C., Gatrell, A. C. (1995), Interactive Spatial Data Analysis. Harlow, UK: Longman Scientific & Technical.

Cohen, P. R. (1995), Elements of the Artificial Intelligence. Cambridge: MIT Press.

Cressie, N. A. C. (1993), Statistics for Spatial Data. New York: Wiley.

Ester, M., Kriegel, H., Sander, J., and Xu, X. (1996), A density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, pp. 226–231.

Ganander, C. (1984), Mythologia fennica, eller förklaring öfver de nomina propria deastrorum, idolorum, locorum, virorum &c eller [etc.]. 4. p. Helsinki: Suomalaisen Kirjallisuuden Seura.

Haavio, M. (1961), Kuolematonten lehdot. Sämpsöi Pellervoisen arvoitus. Porvoo & Helsinki: WSOY.

Holmberg, U. (1914), Tsheremissien uskonto. (Suomen suvun uskonnot 5). Porvoo: WSOY.

Kiviniemi, E. (1990), Perustietoja paikannimistä. (Suomalaisen Kirjallisuuden Seuran Toimituksia 516). Mänttä: SKS.

Koski, M. (1967), Itämerensuomalaisten kielten hiisi-sanue. Semanttinen tutkimus I. (Turun yliopiston julkaisuja, sarja C, osa 5). Turku: Turun yliopisto.

Kuusi, M. (1955), Pakanuuden ja kristinuskon murros suomalaisen kansanrunouden kuvastimessa. Novella plantatio. Suomen Kirkkohistoriallisen Seuran juhlakirja Suomen kirkon juhlavuonna 1955. Toim. Maliniemi et al. (Suomen Kirkkohistoriallisen Seuran toimituksia 56); pp. 145–164. Helsinki: Suomen Kirkkohistoriallinen Seura.

Salo, U. (2003), Suomen ja Hämeen synty. Kauas missä katse kantaa. Unto Salon puheita ja kirjoituksia. Päätoim. Matomäki, T. (Satakunta. Kotiseutututkimuksia XXI). Vammala: Emil Cedercreutzin Säätiö.

Siikala, A. (1995), Mitä kansanuskosta tiedetään? Suomen historia 3. Toim. Blomstedt & al. Espoo: Weilin + Göös.

Tarkka, L. (1987), Tutkimuksia 'puiden varjojen muistoista'. Martti Haavion Kuolematonten lehtojen metodien arviointia. Suomen Antropologi 12; pp. 60–72.

Acknowledgements. We thank National Land Survey of Finland for the place-name data.