

Construction Grammar in onomastics: the case of Finnish hydronyms

Antti Leino

13th January 2006

Abstract

Structural analysis of Finnish toponyms has been well established for the past three decades. Nevertheless, progress during the latter half of that period in cognitive linguistics offers the necessary theoretical background for a new view at toponymic structure. In addition, the emergence of computerised corpora and quantitative methods for analysing such corpora provide the methodological tools for such a new look.

This article is a concise study of Finnish hydronyms as analysed in the framework of Construction Grammar. The analysis covers a variety of onomastic questions, such as naming patterns, the importance of analogy, and inductive naming. It also reveals some features that require some refining of the theoretical base: in particular, phenomena like contrastive naming would seem to imply that partial productivity is much more common than practitioners of this theory have commonly thought.

1 Background

Systematic structural analysis of toponyms is by now well established in Finnish toponyms. The first treatises by ZILLIACUS (1966) and KIVINIEMI (e.g. 1971) are over three decades old, and since then no major paradigm shifts have taken place: later studies have added to the earlier research instead of proposing revolutionary changes to the theory. In fact, these theories might be considered *too* well established, as in recent years research interests have been drifting away from these themes into such fields as socio-onomastics (cf. e.g. AINIALA and PITKÄNEN 2002). However, recent years have seen advances in other sciences, and some of these can help shed new light on toponymic structure.

The traditional typology of Finnish toponyms, as seen by KIVINIEMI (1975), divides toponyms first to one-element and compound names; further analysis is then made on these elements. This further analysis is based on criteria like epexe-gesis, ellipsis, or the propriety of the elements. In this tradition the mechanics of naming has been explained mostly along the lines set by ŠRÁMEK (1972): naming is a process that combines patterns about the semantic content (Ausgangsstellungsmo-dell) with ones about the syntactic structure (Wortbildende Modell), and

relates the result with the surrounding linguistic environment and physical reality. The processes involved can be summarised as analogy; Šrámek's model is simply a more detailed analysis of what is going on. This model has been considered sufficiently good that little thought has been given to it in recent times. However, there are two reasons to take a new look at these issues.

The first such reason is the emergence of cognitive linguistics (e.g. LANGACKER 1987, 1991; FILLMORE and KAY 1995). Here, phenomena such as analogy and metaphor have been raised to the front of linguistic study; also, one of the main cognitivist claims is that a proper linguistic theory should be able to deal with the peripheral phenomena as well as more central ones. Recent onomastic studies (e.g. COATES 2005; BROZOVIĆ RONČEVIĆ and ŽIC FUCHS 2005; LEINO 2005) have started to apply these ideas to onomastics, and it seems that cognitivist theories are at least as good at explaining toponyms as more traditional approaches.

Second, there are also new opportunities for empirical research: computer corpora have been compiled at an impressive rate, and some of these can be used for onomastic studies. At first glance it is evident that these corpora make it possible to perform searches that would have been too cumbersome with paper files. Even better, though, computer scientists dealing with exploratory data analysis have been developing methods specifically for dealing with large masses of data. These can be applied to place name data to give new insights on the underlying onomastic processes.

This study is an attempt to combine the two. In the first stage I have used data-analytical methods to find significant co-location patterns between lake names in a computerised corpus of Finnish place names (LEINO et al. 2003; LEINO 2005). The second stage consists of using these co-location patterns as a starting point to analyse the structure of the lake names in terms of Construction Grammar.

2 Finnish Lake Names

The corpus used for this study is a subset of the Place Name Register maintained by the Finnish National Land Survey (LESKINEN 2002). The register contains all names — or, to be exact name/place pairs — that appear on the 1:20 000 Basic Map; in fact, the register is a part of the system used to produce these maps. For this study I selected all Finnish names¹ for lakes which appear at least 10 times, as shown in Table 1.

In a very crude typology, the lake names can be divided into two. On one hand there are stand-alone names — that is, names that do not require the presence of another place name to be understandable. These names are in the majority; a typical example is *Mustalampi* 'Black Lake'. It should be noted that for the purposes of this article the classification is made entirely on the outward appearance of the name, and these stand-alone names include a large number of names

¹The maps have names in Finnish, Swedish and various Sámi languages.

	Places	Names
All names	800 924	359 394
Names in Finnish	717 747	303 626
Finnish lake names	58 267	25 178
≥ 5 occurrences	29 170	1 492
≥ 10 occurrences	24 078	695
≥ 50 occurrences	12 580	111
≥ 100 occurrences	8 168	45
≥ 500 occurrences	522	1

Table 1: Number of names in the Place Name Register

that are actually based on other near-by names.

On the other hand, there are also names that are clearly inductive: the outward appearance of these names is such that it requires the presence of another name. A typical example of such a name is *Pieni Haukilampi* ‘Lesser Pike Lake’, which implies that there is either an *Iso Haukilampi* ‘Greater Pike Lake’ or a *Haukilampi* ‘Pike Lake’ in the neighbourhood. The frequencies of these types are shown in Table 2.

	Places		Names	
	Number	%	Number	%
Stand-alone	48 889	84	17 915	71
Inductive	9 378	16	7 263	29
Total	58 267	100	25 178	100

Table 2: Frequencies of the main categories

The current article is based on prior research on the same corpus. The first phases of the work (LEINO et al. 2003; LEINO 2004) were mostly focused on first proving that there are significant co-location patterns in the data and then developing methods for discovering such patterns. This resulted in pairs, and some larger groups, of lake names that are attracted to each other: these names occur next to each other at a rate that is notably larger than one would expect by chance, considering the frequency and overall geographic distribution of the names.

Recurrent co-location patterns do not tell very much in themselves, before they have been interpreted in terms of an onomastic theory. In this the first starting point was the traditional theory: my original goal was to see if the patterns could prove the claim, made explicit by PAMP (1991), that analogy plays an important role in naming places even in cases where the name is otherwise motivated as well. While the results of the analysis support this, it is also clear that the phenomena can be better described from a cognitive point of view than using the traditional concepts of naming patterns and analogy (LEINO 2005).

3 Construction Grammar

While the approach in this article is essentially based in cognitive linguistics, much of the tools have been taken from Construction Grammar. This family of theories, as proposed by FILLMORE et al. (1988) and later expanded by FILLMORE and KAY (1995) and others, is generative in its basic view of the role of grammar, in the sense that “[a]ny linguistic expression which is not ‘in’ the language is not licensed by the grammar; and any linguistic expression which is ambiguous can be analysed in more than one way by the grammar” (FILLMORE and KAY 1995: 1-15). This is somewhat weaker than the formulation of CHOMSKY (1957), and it can also be seen as compatible with a cognitive approach. In fact, as LANGACKER (1991: 8) points out, anything stated in terms of Construction Grammar has a direct analogue in Cognitive Grammar.

A common theme in the cognitive movement is that language is considered a collection of *constructions*. The term is used in roughly the same sense as in earlier linguistics: a construction is a unit that consists of form, meaning and a symbolic relation between the two, and it is often schematic to some extent — that is, not all of its components are fully specified. The claim is that everything in a language consists of such constructions, so that sentence schemas and individual lexemes are merely at opposite ends of a single continuum instead of being fundamentally different. Also, since a construction always has a symbolic relation between form and meaning, linguistic description cannot be divided into autonomous syntax and semantics, but rather these are intertwined.

Also, cognitive linguists reject the notion of core language and claim that a good linguistic theory should be able to cope with all phenomena, even those traditionally labelled peripheral. Proper names have often been labelled this way, so they are useful as a test case. Relatively early in the research it became apparent that Construction Grammar in the form presented by e.g. FILLMORE and KAY (1995) and GOLDBERG (1995) has problems in describing phenomena that are neither completely productive nor properly unproductive, such as naming patterns (LEINO 2005: 363). Radical Construction Grammar, as proposed by CROFT (2001), is somewhat more suitable for the current purpose, but it, too, needs some refining.

Radical Construction Grammar is, as the name implies, related to Construction Grammar, but with some differences. First, it takes a strictly taxonomic approach. A construction is not an abstraction in a grammar which is ultimately generative, but rather a generalisation of more specific linguistic units that are similar in some relevant respect; this was, to some extent, foreshadowed by BLOOMFIELD (1933: § 16.6). Such an approach also means that there are no linguistic categories in the traditional sense, but rather each construction has its own, although of course these can be further generalised.

An implication of such a taxonomic approach — which CROFT (2001) does not appear to make explicit — is that there is no fundamental distinction between on one hand a specific construct or utterance and on the other hand a schematic

construction generalised from such constructs. Instead, these are again different ends of a continuum. This, in turn, implies that language use is closer to prototypes than to the unification-based system of FILLMORE and KAY (1995). However, the use of prototypes requires some criteria to assess how far from the prototype one can diverge, and this is especially true when any actual utterance can act as a prototype. In the case of toponyms, one useful way to restrain prototypes can be based on entrenchment: the more something is used, the more firmly it settles as a part of the linguistic system. On the semantic side this is related to the concept of *salience*, as used by e.g. GIORA (1999).

While a specific utterance can be used as a prototype in creating new ones, this is still far less common than using a more schematic construction. This is natural: a construction represents a range of specific utterances, and so it is easy to create a new utterance that fits *within* this range. On the other hand, when a specific utterance is used as a prototype there is no such range, so the new one has to *extend* the scope of the prototype. This, in turn, makes the process less likely, and to offset this, the newly-coined utterance generally has to have a rather high level of similarity with the old one.

The mechanism of creating new utterances, either from constructions or from single prior examples, involves blending several existing elements much in the way proposed by FAUCONNIER and TURNER (2003). In normal language use some of these elements are schematic constructions and some closer to the lexical end of the continuum: for instance, the sentence structure and the accompanying conceptual framework typically are derived from a highly schematic construction, and this is blended with specific words and the corresponding concepts. However, there is no reason why a blend cannot consist of only specific — lexical or idiomatic — elements, and this in fact happens on all levels of language.

4 Constructions in Lake Names

The Finnish lake names behave like most other linguistic phenomena in that they follow the Zipf law: the frequency of the i th most common name is roughly $1/i^a$ of the frequency of the most common one. That is, as seen in Table 1, the most common names are very common but at the other end of the scale there are a very large number of names that occur only once or twice. A similar phenomenon can also be found in the different name constructions: over 85 % of the names can be considered variants of just one basic construction and 91 % fall under the two most common ones, as seen in Table 3.

The basic construction for stand-alone lake names has an identifying element followed by an element that marks the type of place, as seen in Figure 1. There are four relatively common subtypes:

1. The identifying element can be an adjective. In such cases the identifier is usually motivated by a notable feature of the lake, such as the darkness of its

	All		Stand-alone		In inductive			
	Places	Names	Places	Names	Places	Names		
	%	%	%	%	%	%		
<i>-lampi</i> 'pond'	35 626	61	11 975	48	65	53	37	33
<i>-järvi</i> 'lake'	14 095	24	6 951	28	24	29	25	25
<i>-vesi</i> 'water'	214	0	180	1	0	1	0	0
<i>-nen</i>	2 966	5	1 705	7	4	4	13	13
<i>-kkV</i>	634	1	511	2	1	1	4	4
Other	4 732	8	3 856	15	6	12	20	24
All	58 267	100	25 178	100	100	100	100	100

Table 3: Frequency of the most common constructions

water (as in *Mustalampi* 'Black Pond') or its shape (as in *Pitkälampi* 'Long Pond').

- The identifying element can be a noun. In these cases the identifier is often related to the use or shape of the lake (e.g. *Abvenlampi* 'Perch Pond', *Salmi-järvi* 'Strait Lake'), or it can be a reference to a near-by feature (e.g. *Riihilampi* 'Drying Barn Pond').
- As something of a special case of the above, the identifying element can be a noun in the genitive case. Such an identifier is often, but by no means always a personal name or the name of a near-by place.
- The identifying element can also be a verb stem, usually related to the use of the lake (e.g. *Likolampi* 'Retting Pond', referring to a step in the process of making linen).

In addition to the *identifier + type of place* construction there is another relatively common one, in which the identifying element is followed by a suffix. Historically, at least some of these cases are contractions: *Valkeinen* < **Valkeajärvi*. Not all such names can be explained in this manner, though, and even when they can such a diachronic explanation is somewhat problematic in itself: it is not clear that current users of the names view them in this manner. Synchronically, at least, the names can be interpreted as in Figure 2. Here the identifying element can be either an adjective or a noun, as seen in the examples in the figure. Also, the figure shows the two common suffixes; of these, the names ending in *-kkV* are rare and largely opaque, and perhaps should not be classified here.

These two main constructions cover roughly nine tenths of the lakes. The rest are mostly unique names, although there are a few names with as many as twenty occurrences. It is possible to classify these names into four very rough categories.

- There are several names that consist of simply a common noun. The most common name in this group is *Kaakkuri* 'Loon', with 8 lakes.
- Some names consist of just an adjective, like *Hoikka* 'Thin' (26 lakes). It is possible that names in these first two categories have been created by

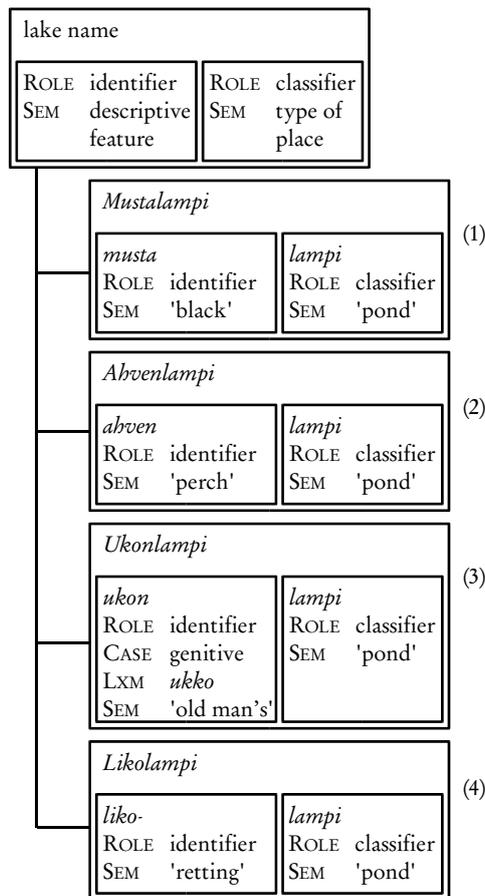


Figure 1: Basic construction for lake names

ellipsis, but from the point of view of a present-day name user these lakes just happen to be named like they are.

- Some names are compound, but do not quite fall into the *identifier + type of place* pattern. The most common name in this group, *Peipposenmeri* 'Chaffinch's Sea' (various spellings, 21 lakes in all), is formed after that pattern, but on the other hand it is clearly metaphorical: a chaffinch is a small bird, and these "seas" are very small ponds.
- Several names are also completely opaque to their present-day users, like *Päijänne* (21 lakes).

It is also possible to consider the *identifier + classifier* and *identifier + suffix* constructions related in some respects. In order to justify this, it is instructive to look at some common co-location patterns in the names. The corpus contains several hundred pairs of names that show a tendency to appear near each other (LEINO 2005), and it is clear that a large number of these pairs are a result of pro-

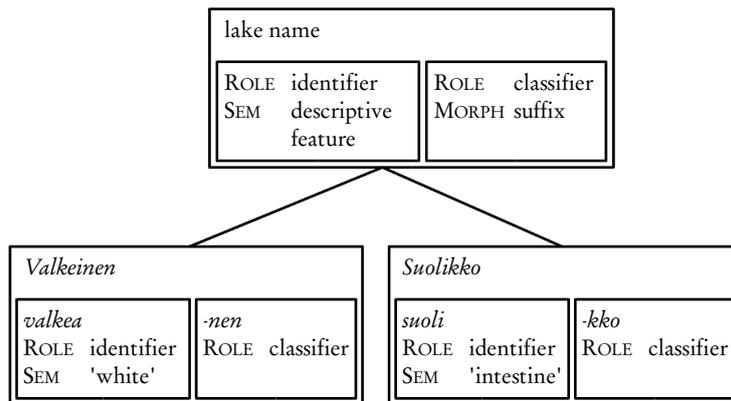


Figure 2: The *identifier + suffix* construction

cesses traditionally called naming patterns and analogy. Some examples of such pairs are shown in Figure 3; the direction of the arrow is hypothetical, especially in the case of pair (b).

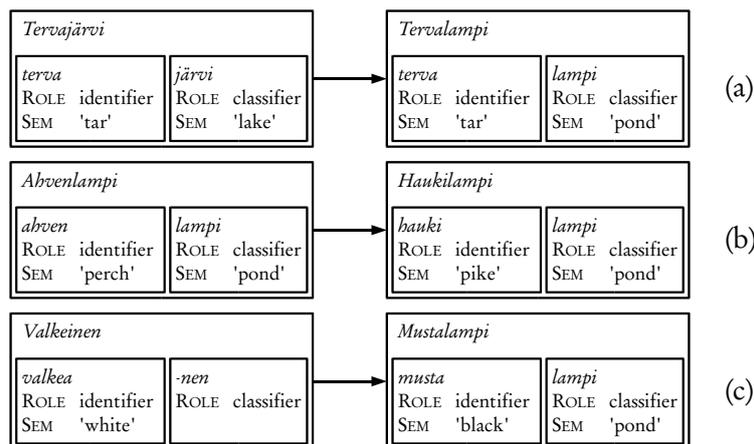


Figure 3: Some common co-occurring names

Two things can be inferred from pairs like these. First, they can be seen as one more piece of evidence against the claim that there is a qualitative distinction between on one hand constructions which are systematically productive and on the other hand patterns of coining which result in unique innovations. The co-occurring pairs tell another story: there is a large group of single innovations, each of which is productive in itself. Considering this, using single names as prototypes for forming new ones cannot be considered exceptional.

The second inference is closely related to cases like (c) in Figure 3, where one name in the pair is an instantiation of the *identifier + type of place* construction and the other of the *identifier + suffix* construction. There are some such cases,

and they look remarkably similar in other respects to cases like (b) where both names end in a *type of place* classifier. It is certainly possible that several of these cases date from an earlier time when the suffixation had not yet happened; still, from a present-day perspective it is also clear that these names will be interpreted as related. It seems, therefore, plausible to postulate a family of constructions that covers both cases, as in Figure 4. This is not an altogether new idea: KIVINIEMI (1975: 41) mentions that suffixes can replace the head of a compound noun, and PAMP (1973: 63–64) treats suffixes the same way as heads of compound nouns.

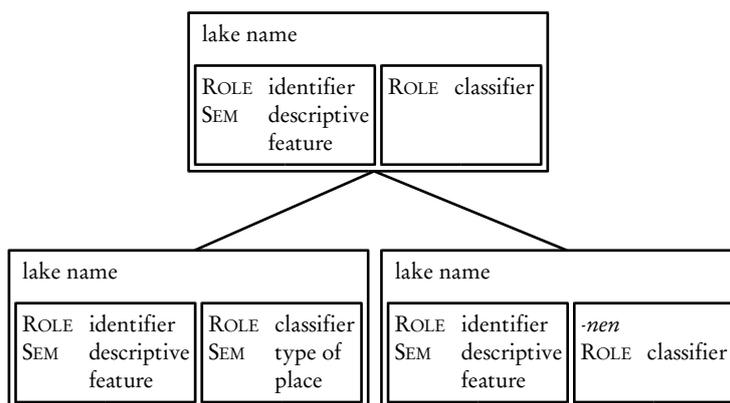


Figure 4: Overall construction *identifier + classifier*

Inductive names add another common construction. This one, as seen in Figure 5, consists of a modifier followed by a toponym. Such names often occur in pairs where the names have contrasting modifiers such as *Iso / Pieni* or other variants of 'Large' and 'Small', and the unmodified toponym does not appear at all. The size of the lake is the only criterion that appears frequently enough to be visible in the co-location patterns, but a large number of other modifiers exist in the corpus: such modifiers are usually adjectives, like the identifiers in subtype 1 of the basic stand-alone construction in Figure 1.

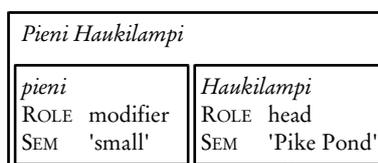


Figure 5: The *modifier + name* construction

Another common way to form a toponym from another is to use a proper name as the identifier part in a Figure 1 subtype 3 construction. However, in this study I have chosen to treat them as stand-alone names, since there is no obvious criterion for making a distinction between primary and secondary names of this form. A name like *Kalettomanlampi*, in Figure 6, may or may not be an inductive

name: it could be a smaller pond near a larger lake named *Kaleton*, or the pond could itself have been named **Kaleton* and the current name a result of epexegetis.

<i>Kalettonlampi</i>	
<i>Kaleton</i>	<i>lampi</i>
ROLE identifier	ROLE classifier
CASE genitive	SEM 'pond'
LXM <i>Kaleton</i>	
SEM 'Fishless'	

Figure 6: Stand-alone or inductive?

5 Conclusions

By now it should be no surprise that names are modelled after existing ones. However, it appears that it is often very difficult to distinguish between semantic and syntactic patterns in the way earlier, essentially structuralist, name typologies have attempted to do. A cognitive approach which recognises that the two are fundamentally intertwined is, in this respect, more true to the data at hand.

It is also remarkable how widespread certain types of naming patterns are. Even though every individual name has been uniquely given, the exactly same innovation can be — and has frequently been — made time and again. This is hard to explain if one considers productivity an either/or matter: rather, at least in naming places, partial productivity is the normal state of affairs. This can be explained rather well by changing to a prototype-based view of constructions. According to this view, new concrete utterances are not created from constructions by means of unification, so that the unspecified elements in one match the specified ones in the other, but rather by blending the conceptual spaces represented by these constructions. The end result is mostly the same, but now a new utterance can also be based on specific examples in a halfway-productive manner.

The resulting model for toponymic structure is somewhat simpler than the traditional one. However, according to a normal formulation of Ockham's Razor, it can be considered better only if the increase in simplicity is achieved without sacrificing predictive power. This means that there has to be some way of ensuring that one can't go arbitrarily far from the prototype when constructing a new name. The use of prototypes has to be restrained, but fortunately it seems that such criteria as salience and entrenchment can be used as a basis for constructing these restraints.

All in all, however, one important conclusion is that a cognitive approach results in a model that adequately describes a toponymic corpus. Another one is that onomastics is useful in refining cognitive theories, as there are phenomena such as partial productivity that are more easily apparent in names than in more traditional linguistic material. Proper names are a proper part of language and

should be considered as such.

References

- T. AINIALA AND R. L. PITKÄNEN: *Onomastic research in Finland*. “Onomastica Uralica” 2 (2002):81–121.
- L. BLOOMFIELD: *Language*. New York: Henry Holt & Co., 1933.
- D. BROZOVIĆ RONČEVIĆ AND M. ŽIC FUCHS: *Metaphorical processes as the basis of proper names*. D. Brozović Rončević and E. Caffarelli (ed.), “Naming the World. From Common Nouns to Proper Names. Proceedings from the International Symposium, Zadar, September 1st–4th, 2004”, QUADRION 1, pp. 33–44. Roma: Società Editrice Romana, 2005.
- N. CHOMSKY: *Syntactic Structures*. The Hague: Mouton & Co., 1957.
- R. COATES: *A speculative psycholinguistic model of onymization*. D. Brozović Rončević and E. Caffarelli (ed.), “Naming the World. From Common Nouns to Proper Names. Proceedings from the International Symposium, Zadar, September 1st–4th, 2004”, QUADRION 1, pp. 3–13. Roma: Società Editrice Romana, 2005.
- W. CROFT: *Radical Construction Grammar. Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press, 2001.
- G. FAUCONNIER AND M. TURNER: *The Way We Think: Conceptual Blending and the Mind’s Hidden Complexities*. New York: Basic Books, 2003.
- C. FILLMORE AND P. KAY: *Construction Grammar*. CSLI Lecture Notes. Stanford: Center for the Study of Language and Information, 1995.
- C. J. FILLMORE AND P. KAY AND M. C. O’CONNOR: *Regularity and idiomatity in grammatical constructions: The case of let alone*. “Language” 64 (1988)(3):501–538.
- R. GIORA: *On the priority of salient meanings: Studies of literal and figurative language*. “Journal of Pragmatics” 31 (1999)(7):919–929.
- A. GOLDBERG: *Constructions. A Construction Grammar Approach to Argument Structure*. Chicago and London: University of Chicago Press, 1995.
- E. KIVINIEMI: *Suomen partisiippinimistöä*. Suomalaisen Kirjallisuuden Seuran toimituksia 295. Helsinki: Finnish Literature Society, 1971.
- *Paikannimien rakennetyypeistä*. Suomi 118:2. Helsinki: Finnish Literature Society, 1975. Abstract: *Über die Strukturtypen der Ortsnamen*.

- R. W. LANGACKER: *Foundations of Cognitive Grammar*, volume I: *Theoretical Prerequisites*. Stanford: Stanford University Press, 1987.
- *Foundations of Cognitive Grammar*, volume II: *Descriptive Application*. Stanford: Stanford University Press, 1991.
- A. LEINO: *Pikes and perches go together. A data-analytical view on Finnish lake names*. M. Nenonen (ed.), “Papers from the 30th Finnish Conference of Linguistics”, *Studies in Languages* 39, pp. 79–84. Joensuu: University of Joensuu, 2004.
- *In search of naming patterns: A survey of Finnish lake names*. D. Brozović Rončević and E. Caffarelli (ed.), “Naming the World. From Common Nouns to Proper Names. Proceedings from the International Symposium, Zadar, September 1st–4th, 2004”, *QUADRION* 1, pp. 355–367. Roma: Società Editrice Romana, 2005.
- A. LEINO AND H. MANNILA AND R. L. PITKÄNEN: *Rule discovery and probabilistic modeling for onomastic data*. N. Lavrač and D. Gamberger and L. Todorovski and H. Blockeel (ed.), “Knowledge Discovery in Databases: PKDD 2003”, *Lecture Notes in Artificial Intelligence* 2838, pp. 291–302. Berlin Heidelberg: Springer, 2003.
- T. LESKINEN: *The geographic names register of the National Land Survey of Finland*. “Eighth United Nations Conference on the Standardization of Geographical Names”. 2002.
- B. PAMP: *Kring ortnamnens grammatik*. *Namn och Bygd* 61 (1973):55–82.
- *Onomastisk analogi*. G. Albøge and E. Villarsen Meldgaard and L. Weise (ed.), “Tiende nordiske navneforskerkongres, Brandbjerg 20.–24. maj 1989”, *Nornarapporter* 45, pp. 157–174. Uppsala: Swedish Science Press, 1991.
- R. ŠRÁMEK: *Zum Begriff “Modell” und “System” in der Toponomastik*. “*Onoma*” 17 (1972):55–75.
- K. ZILLIACUS: *Ortnamnen i Houtskär. En översikt av namnförrådets sammansättning*. *Studier i nordisk filologi* 55. Helsingfors: Svenska Litteratursällskapet i Finland, 1966.