Unni Leino

Overlap in present-day Finnish place names, given names, and surnames

1. Names and patterns

One of the fundamental questions – if not *the* question – in onomastics is the nature and origins of proprical expressions. Until the mid-20th century, the prevailing consensus was that names were descriptive in origin. While this was especially relevant in the case of toponyms, the origins of anthroponyms was viewed similarly. As a general onomastic principle this traces back to GOTTFRIED WILHELM VON LEIBNIZ (1710): "Illud enim pro axiomate habeo, *omnia nomina quæ vocamus propria, aliquando appellativa fuisse,* alioqui ratione nulla constarent."

Especially in the past few decades, scholars have increasingly started to disagree with this view, to the extent of suggesting that proper names are in fact primary and appellative nouns a secondary development (e.g. VAN LANGENDONCK 2007: 116–118), or that properhood itself is not a categorical property but rather a matter of pragmatics (e.g. COATES 2005, 2006). In any case, however, it is clear that there are interactions between appellative nouns and proper names, and between proper names in different categories.

The transition from seeing names as a development of descriptive appellations started in the latter half of the 20th century, when several onomasticians (notably VINCENT BLANÁR and RUDOLF ŠRÁMEK in Czechoslovakia, and KURT ZILLIACUS and EERO KIVINIEMI in Finland) began formulating a theoretical framework that saw naming as a pattern-based process. Most of this pioneer work was done in the context of toponyms (e.g. ZILLIACUS 1966, ŠRÁMEK 1972, KIVINIEMI 1977), but the model-based approach applies equally well to anthroponyms (BLANÁR 1990, PAIKKALA 2004: 44–51) and is compatible with a cognitive view on linguistics (SJÖBLOM 2006, LEINO 2007).

In toponomastics, model-based naming has typically been described as having separate models for structure and content. In ŠRÁMEK's (1972) view, the overall model can be described as a formula

$$ON=S\,\frac{AM+WM}{Z+L}$$

where the place name (ON, Ortsname) is derived from a complex interaction between the language (S, Sprache), the patterns that govern the motivations for the name (AM, Ausgangsstellungsmodell) and its structure (WM, Wortbildende

Modell); moreover, these patterns are specific to time (Z, Zeit) and place (L, geographische Lage).

Having separate patterns for the structure and content of the name can be used to describe different aspects of such pairs of names as *Myllyjärvi* 'Mill Lake' – *Kirkkojärvi* 'Church Lake', *Mäntyjärvi* 'Pine Lake' – *Mäntylampi* 'Pine Pond', *Iso Haukilampi* 'Great Pike Pond' – *Pieni Haukilampi* 'Small Pike Pond' or *Valkeinen* 'The White' – *Mustalampi* 'Black Pond' (LEINO 2005). In anthroponyms, the meaning of name elements is typically less relevant, and it is more useful to see names as forming a series, such as the female given names ending in -*nja* (SARTJÄRVI 2011): *Anja, Banja, Benja, Binja, Brenja, Brynja, Cenja, Cinja, Danja, Denja, Donja, Dunja, Enja, Fanja, Fenja, Finja, Fonja, Frinja, Genja, Granja, Henja, Inja, Janja, Jinja, Jonja, Kenja, Kinja, Kranja, Ksenja, Lenja, Lonja, Manja, Menja, Minja, Monja, Nanja, Nenja, Ninja, Onja, Panja, Penja, Pinja, Ranja, Renja, Rinja, Ronja, Sanja, Senja, Sinja, Sonja, Stanja, Svenja, Taanja, Tanja, Tenja, Tinja, Tonja, Tunja, Unja, Vanja, Venja, Vinja, Wanja, Xenja, Ynja, Zenja.*

In studying the use of names, the difference of the two viewpoints is not all that relevant: whether names have developed from appellative descriptions or appellatives from proper names, the difference between an appellative noun and a proper name is that the primary pragmasemantic content of the latter is a reference to an individual while the former has a more general meaning. However, in studying the various relationships between proper names and appellatives, and between different types of proper names, it would be misleading to start from the assumption that names are by nature appellative and descriptive in origin.

It's important to understand that pattern-based naming does not require that all patterns are already established, just as not all new sentences are completely based on established grammatical structures. Naming is innovation, and the 'patterns' that are used can be just single prior examples (see e.g. LEINO 2007). On the other hand, the more thoroughly a pattern has been entrenched as a part of the onomastic system the more likely it is to be used in coining new names.

2. Overlapping categories

A proper name can appear in several different categories. For instance, looking at Finnish toponyms and anthroponyms, *Kanerva* appears as a male given name, a female given name, a surname, and a name for a farm, a village, a field, and a small hill. There are several relatively common reasons for this kind of overlap in the Finnish toponomastic and anthroponomastic systems.

Toponyms are often taken from anthroponyms, typically by a farm being named for its owner (e.g. KEPSU 1981: 283). Earlier, the farmer's personal name was used for this, so that there are toponyms that look like given names; in the past century or so, though, it has become increasingly common to use a surname. These farm names can occasionally be transferred further, so that eventually the anthroponym ends up as being the name of an entire village, or in rare cases a natural feature such as a bay or a hill.

Similarly, toponyms can become anthroponyms. The most common way for this to happen is for a farm name to be adopted as a surname; this has happened regularly enough that the connection is used also to name farms after the owner families. In addition to this, since the 19th century some names of natural features – like the fell *Saana* or the lake *Inari* – have become given names.

There are also connections between surnames and given names. Sometimes given names have become surnames; historically, this isn't a case of an unmarked patronymic byname becoming hereditary, but rather a more complex development where often the surname has been taken from a farm name that can be further traced to a given name.

There are also cases where the names have been taken separately but from a common origin. For instance, *Kanerva* 'heather' has been adopted as both a given name and a surname, and also – likely through ellipsis – as the name of a couple of small hills. In some cases, the names have a similar-appearing form but different etymologies, so that for instance *Kaija* as a given name is a contraction of *Catharina* but as a surname derived from an appellative meaning 'gull'.

3. Material

To look at how common these different types of overlap are in present-day Finnish names, I have used two different data sources, both kept up to date by government offices.

For anthroponyms, I've used the Population Information System, kept by the Population Registration Centre; a snapshot with a selection of onomastically relevant fields was acquired by the Institute for the Languages of Finland in February 2002 and is available for research purposes. A summary of the names is shown as Table 1. This study uses the given names and surname given at birth to those people whose native language is listed as Finnish. The number of all given names is smaller than the sum of names given to girls and boys, as there is a number of names that are gender-neutral to some degree (see e.g. LEINO 2014: 806–807).

	Given names			Surnames	People
	All	Female	Male		
Finnish	63 564	38 565	27 360	93 706	6 822 100
All languages	100 488	58 556	47 414	153 893	7 539 588

Table 1: Anthroponyms in the Population Information System

For toponyms, I've used the Place Name Register kept by the National Land Survey for the purposes of compiling maps that cover Finland. Several snapshots of the register were acquired by the Institute for the Languages of Finland for onomastic research; I've used one taken in 2001 as the one that most closely matches the anthroponym data. The toponym data is summarised as Table 2.

	Names	Places
Finnish	303 626	717 747
All languages	359 394	800 924

Table 2: Toponyms in the Place Name Register

4. Overlapping names

While an in-depth research of the relationships between toponyms and anthroponyms would be a massive task, these data sets make it easy to compare the names and look for simple overlap between the categories. In this study, I look at identical forms – for instance, *Pentti* as a given name, surname, and place name, but not the related place name *Penttilä*. This is, of course, not an ideal solution but a necessary one, as finding etymological connections between names is difficult even for a well-trained onomastician. Nevertheless, there is a lot of overlap: as seen in Table 3, there are even a few names that have 50 instances in each of the three categories.

	All	≥ 2	≥ 5	≥ 10	≥ 20	≥ 50
Place / surnames	23 486	14 502	7 148	4 030	2 258	925
Given / surnames	4 762	2 555	1 215	759	467	261
Place / given names	1 855	742	201	79	23	4
All three	1 397	625	195	79	23	4

Table 3: Overlapping names

4.1. Surnames as place names

Surnames and place names are the pair of categories where overlap occurs most frequently, and 53 most common Finnish surnames all appear as place names as well. One can divide these into three or five distinct categories.

First, there are surnames that started out as farm names. This follows the traditional Western Finnish practice of a farmer using his farm name as a customary byname; this practice was common enough that in 1903–06 recommendations for adopting farm names as surnames were published in newspapers (MIKKONEN 2013: 71–77).

For the purposes of this study, the farm names that are used as surnames can be further divided into ones that contain an old personal name, like *Heikkilä* < *Heikki* (with 18 391 instances as a surname and 1 133 as a place name) or *Mattila* < *Matti* (16 094 and 1 028) and ones that are not anthroponymic in origin, such as *Mäkelä* < *mäki* 'hill' (26 759 and 2 199) or *Niemi* 'cape' (20 732 and 558). Both these categories contain names that are relatively common as toponyms, which is not surprising considering the toponymic origin of the names.

Second, there are old surnames. While surnames have been mandatory in Finland only since 1921, they have been used in Eastern Finland at least since the first Savonian tax records in the 16th century. These names can be further divided into ones that are patronymic in origin, such as *Heikkinen* (24 022 and 20) or *Manninen* (13 506 and 40), and those that are not, like *Korhonen* (32 632 and 18) or *Hämäläinen* (27 368 and 24). These are markedly less common as toponyms, as here the toponymic use is secondary.

Third, there are new surnames that originate from the mid-19th century or later. PAIKKALA (2004: esp. 400–415) has tracked the origins of these names and found two main types; both of her prototype names, *Virtanen* (36 493 and 53) and *Laine* (26 235 and 48), appear as place names as well. The frequency of these new surnames as toponyms is on par with that of the old surnames, which indicates that the toponymic use of surnames is mostly a young phenomenon. In addition to these, MIKKONEN (2013: 153–156) also lists surnames taken in late 19th and early 20th century that were based on toponyms other than farm names.

The most common toponyms that appear as surnames are either farm names like *Koivula* < *koivu* 'birch' (5 319 and 2 221, rank 1) or *Keskitalo* 'Middle House' (2 890 and 547, rank 51), or other old locatives like *Myllykoski* 'Mill Rapids' (1 110 and 373, rank 101) or *Lahti* 'Bay' (10 876 and 248, rank 201). These cover roughly the 250 most common toponyms in the surname set; after a a transitional zone of ranks 250–350 or so, other toponyms become the norm.

These include agrarian toponyms like *Isoniitty* 'Big Meadow' (6 and 159, rank 358), natural places such as *Kalliojärvi* 'Rocky Lake' (234 and 130, rank 452), names with pleasant connotations that make them appe alingas surnames, like *Mansikkamäki* 'Strawberry Hill' (861 and 186, rank 289), and modern surnames of the types discussed by PAIKKALA (2004: 212–222) such as *Aaltonen* (15 069 and 21, rank 2 645). Finally, there are occasional surnames of recent foreign origin, like *Andrejev* (80 and 1).

4.2. Surnames as given names

The overlap between surnames and given names is not nearly as large as that between surnames and place names. Nevertheless, most of the top male given names appear as surnames as well, like *Juhani* (31 880 instances as a given name and 1 as a surname) or *Tapio* (130 391 and 2 317). As seen here, some of these names are very rare as surnames, though. Moreover, only some of the common female given names do so, like *Helena* (152 377 and 1) but not *Maria* (301 211).

Likewise, many of the most common surnames occur in the data as given names. In many cases, this does not indicate real use as given names; rather, these can be explained as marital double surnames mistyped without a hyphen, so that one of the surname elements is incorrectly interpreted as a given name, or Scandinavian mellannamn (ENTZENBERG 2005, HARDING KÆLLERØD 2017) of people who have married in Sweden or Denmark and later moved back to Finland. Some, however, are genuinely used in both categories; these include both 19th century new-style surnames (MIKKONEN 2013: 248–254) such as *Laine* (199 and 26 235, rank 5) and old personal names like *Heino* (3 596 and 8 984, rank 37).

There are 26 names that occur at least 500 times in both categories. Some of these are medieval given names, like *Heino,* a Finnish form of *Heinrich*. While the historical Finnish name system did not include unmarked patronymic bynames that would have led directly to surnames, given names did occasionally become farm names (KEPSU 1981: 287–289).

Some of the names common in both categories date clearly from the 19th century. In that period, there was a major push towards Finnish names, and the new ones did not always establish themselves clearly in a single category. As a result, names like *Ahti* < 'pagan god of water / fishing' (9 511 and 731) or *Kanerva* < 'heather' (1 218 and 4 241) ended up common as both given and place names.

Finally, there are some names where the overlap is by chance: in some cases, a surname and a given name can look similar but have different etymologies.

Thus *Kaija* (18 447 and 641) is on the one hand a Finnish variant of the given name *Catharina* but on the other hand a surname ultimately derived from a dialectal word for 'gull'.

4.3. Given names as place names

The overlap between given names and place names is again less common than that between the two anthroponymic categories, as seen in Table 4. Moreover, the overlap of given names with, on the one hand, place names and, on the other hand, with surnames is clearly correlated, so that all given names which appear at least 9 times as place names also appear as surnames.

	Gender	Given name	Surname	Place name
Juhani	m	318 880	1	2
Maria	f	301 211	–	–
Johannes	m	212 790	11	–
Olavi	m	210 153	1	–
Antero	m	175 428	21	1
Helena	f	152 377	1	–
Tapani	m	148 040	463	28
Anneli	f	140 370	–	–
Kalevi	m	139 084	42	–
Tapio	m	130 391	2 317	35
Johanna	f	128 752	–	–
Kaarina	f	127 283	–	1

Table 4: Most common given names as both surnames and place names

There are 458 given names that occur as place names but not surnames. The oldest ones of these – those that were already used as given name before 1880 – are traditional given names. Many of these, such as *Eerikki* (3 724 and 3, two farms and an island, first occurrence as given name in 1859), are cases where a farm name was based on a personal name but didn't develop into a surname, while others, like *Saara* (17 185 and 1, brook, given name since 1862) or *Kaarlo* (25 678 and 1, island, given name since 1867) have various other origins. These names are all very rare as place names, regardless of their frequency as a given name.

Starting in the 1880's, new given names such as *Ainikki* (591 amd 1, woods, given name since 1881) or *Ihana* 'lovely' (63 and 4, all houses, given name

since 1882) appear. Many of these seem to be villa names that were motivated by the wishes of the villa owner.

Since the 1960's, the motivation for choosing either the given name or the place name have been less and less clear. In some cases, like *Manteli* 'almond' (10 and 1, farm, given name since 1972), the name appears semantically transparent; in others, like *Muska* (7 and 1, farm, given name since 1965), names is opaque. In yet others, the given name is clearly a pet form of an established given name, like *Hessu* (4 and 2, cape and lake, given name since 1989), usually a from of *Henrik* but listed by VILKUNA et al. (1984) also as used for *Hedvig, Helena, Helge* and *Hesekiel,* or *Jamppa* (5 and 1, village, given name since 1962), given by VILKUNA et al. as a form of *Jalmari < Hjalmar* and PAUNONEN–PAUNONEN (2000) for *Jarmo < Jeremias* and *Janne < Johannes.*

4.4. All three at once

As already seen in Table 3, a vast majority of the given names that appear as place names appear also as surnames. There are in total 1 397 of them; 10 names have at least 30 occurrences in all of the categories. These can be divided into four distinct groups.

First, there are the medieval given names *Pentti < Benedict* (58 572 given names, 1 097 surnames and 35 place names) and *Soini < Sven* (1 511 and 2 915 and 36). Here, the given name has first become a farm name, and the farm name subsequently adopted as a surname.

Second, there are names with associations to nature: *Alho* 'small marsh' (156 and 2 523 and 226), *Kanerva* 'heather' (1 218 and 4 241 and 91), *Laine* 'wave' (199 and 26 235 and 48) and *Siimes* 'calm shadowy place' (33 and 465 and 81). The vast majority of the toponymic occurrences are house names, although *Alho* appears often as a field name as well.

Third, there is one name that has its origins in Finnish mythology. *Tapio* (130 391 and 2 317 and 35) is mentioned as the name of a god of forests and hunting already by AGRICOLA (1551) in his list of pagan deities. As a given name it started gaining popularity in the early 20th century, with a peak just after the middle of the century; the toponymic occurrences are again mostly houses.

Fourth, there are three cases where the given name is etymologically different from the surname and place name. *Kari* (55 995 and 3 033 and 54) as a given name is a contraction of *Makarios,* but its toponymic use comes from the appellative *kari* 'underwater rock'. Similarly, *Maja* (64 and 552 and 47) is a Swedish variant of *Maria,* occasionally seen in Finnish as well, while the

toponym is based on *maja* 'hut'. These places, too, are mostly houses. Finally, *Roni* (2 140 and 119 and 53) as a given name is a contraction of *Hieronymus;* the origins of the similar-sounding toponym – in this case, evenly divided between small woods, hills, and houses – is less clear. There is a dialectal appellative meaning 'spring' (from Swedish *brunn* 'well'), but its georaphic distribution does not match that of the toponym.

5. Conclusion

There is a lot of overlap between surnames and place names. The overlap is mostly between surnames and farm names, and there are three historical developments that had led to this. In Western Finland, there was a long-standing tradition, since medieval times, to use the farm name as a byname in various documents, often in addition to a patronymic. Continuing this use, in the late 19th and early 20th century it was common – and even recommended – to adopt the farm name as a surname.

In Eastern Finland, hereditary surnames were in use already at the time of the earliest comprehensive records in the 16th century. In some cases, the surname became also the name of the farm. In a similar manner, farm names were given after surnames in the 20th century; this, however, can be seen as a distinct phenomenon, motivated after a more general awareness of the connection between surnames and place names. This connection has also led to new surnames being coined from existing toponyms, in modern times mostly other than farm names.

Given names have less in common with the other two categories. Many of the cases are due to farms, especially the oldest ones, being named after individuals, and for various reasons given names as such appear also as the names of other types of places. Most of the given names that appear as place names exist also as surnames, and there are no significant given names that are clearly toponymic in origin.

The rise of national awareness in 19th century led to new Finnish anthroponyms, both given and surnames. This trend continued several decades into the 20th century, and as new names were coined they were not always established in one category before being used in the other. As a result, despite a clearly stated prohibition in the Given Names Act of 1945, there is significant overlap between the two categories. In some cases, the given name and surname have different etymological origins.

References

Agricola, Michael 1551. *Dauidin Psaltari.* Stockholm, Amund Lauritzon.

Blanár, Vincent 1990. Das anthroponymische System und seine Funktionierung. In: Närhi, Eeva Maria ed. *Proceedings of the XVIIth International Congress of Onomastic Sciences.* Helsinki, University of Helsinki–Finnish Research Centre for Domestic Languages. I, 208–215.

Coates, Richard 2005. A Speculative Psycholinguistic Model of Onymization. In: Brozović Rončević, Dunja–Caffarelli, Enzo eds. *Naming the World: From Common Nouns to Proper Names. Proceedings from the International Symposium, Zadar, September 1st–4th, 2004, QuadRIOn 1.* Roma, Societa Editrice Romana. 3–13.

Coates, Richard 2006. Properhood. *Language 82:* 356–382.

Entzenberg, Sonja 2005. Släktnamnens sociala signifikans – gifta kvinnors val av släktnamn i modern tid. [The social significance of surnames – the choice of surname by married women in modern times.] In: Nyström, Staffan ed. *Namnens dynamik. Utvecklingstendenser och drivkrafter inom nordiskt namnskick. Handlingar från den trettonde nordiska namnforskarkongressen i Tällberg 15–18 augusti 2003.* Uppsala, NORNA-förlaget. 119–132.

Harding Kællerød, Lars-Jakob 2017. Det danske mellemnavn: et spørgsmål om identitet? Betragtninger over brugen af mellemnavne i Danmark i 1800-talet. [The Danish middle name: a question of identity? A look at the use of middle names in Denmark in the 19th century.] In: Leino, Unni-Päivä et al. eds. *Namn och identitet. Handlingar från NORNAs 46:e symposium i Tammerfors den 21–23 oktober 2015.* Uppsala–Tammerfors, NORNA-förlaget. 95–113.

Kepsu, Saulo 1981. *Pohjois-Kymenlaakson kylännimet. Village Place Names in Northern Kymenlaakso.* Helsinki, Suomalaisen Kirjallisuuden Seura.

Kiviniemi, Eero 1977. *Väärät vedet. Tutkimus mallien osuudesta nimenmuodostuksessa.* [Crooked waters. A study in the use of models in naming.] Helsinki, Suomalaisen Kirjallisuuden Seura.

Freiherr von Leibniz, Gottfried Wilhelm, 1710. Brevis designatio meditationum de Originibus Gentium, ductis potissimum ex indicio linguarum. In: *Miscellanea Berolinensia ad incrementum scientiarum, ex scriptis Societati Regiae Scientiarum exhibitis edita.* I, 1–16. Berolinum, Johan. Christ. Papenius. (http://www3.bbaw.de/bibliothek/digital/)

Leino, Antti 2005. In Search of Naming Patterns: A Survey of Finnish Lake Names. In Brozović Rončević, Dunja–Caffarelli, Enzo eds. *Denominando il mondo. Dal nome comune al nome proprio / Naming the world. From Common Nouns to Proper Names*, *Quaderni Internazionali di RIOn.* Roma, Societa Editrice Romana. 355–367.

LEINO, ANTTI 2014. Man, woman or me? Conflicting identities as evidenced by cross-gender name changes. In: TORT I DONADA, JOAN–MONTAGUT I MONTAGUT, MONTSERRAT eds. *Names in daily life. Proceedings of the XXIV ICOS International Congress of Onomastic Sciences.* Barcelona, Generalitat de Catalunya. 803–811.

LEINO, ANTTI 2007. *On Toponymic Constructions as an Alternative to Naming Patterns in Describing Finnish Lake Names.* Helsinki, Finnish Literature Society.

MIKKONEN, PIRJO 2013. *Otti oikean sukunimen. Vuosina 1850-1921 otettujen sukunimien taustat.* [Took a proper surname. The origins of surnames adopted in 1850–1921.] Helsinki, Unigrafia.

PAIKKALA, SIRKKA 2004. *Se tavallinen Virtanen. Suomalaisen sukunimikäytännön modernisoituminen 1850-luvulta vuoteen 1921.* [The modernisation of the Finnish surname use from the 1850's to 1921.] Helsinki, Suomalaisen Kirjallisuuden Seura.

PAUNONEN, HEIKKI–PAUNONEN, MARJATTA 2000. *Tsennaaks stadii, bonjaaks slangii. Stadin slangin suursanakirja.* [The grand dictionary of Helsinki slang.] Helsinki, WSOY.

SARTJÄRVI, MAIJA 2011. Aija ja Meija, Anja ja Minja: Kaksitavuisten ija-ja nja-päätteisten naistennimien yleistyminen ja suosionvaihtelu. [Two-syllable female names ending in *ija* / *nja*, their becoming more common, and their varying popularity.] Pro gradu – tutkielma, Helsingin yliopiston suomen kielen laitos. (http://urn.fi/URN:NBN:fi-fe201108112010)

SJÖBLOM, PAULA 2006. *Toiminimen toimenkuva. Suomalaisen yritysnimistön rakenne ja funktiot.* [The structure and functions of Finnish commercial company names.] Helsinki, Suomalaisen Kirjallisuuden Seura.

ŠRÁMEK, RUDOLF 1972. Zum Begriff "Modell" und "System" in der Toponomastik. *Onoma 17:* 55–75.

VAN LANGENDONCK, WILLY. 2007. *Theory and Typology of Proper Names.* Berlin–New York, Mouton de Gruyter.

VILKUNA, KUSTAA–HUITU, MARKETTA–MIKKONEN, PIRJO 1984. Etunimet. [Given names.] In: *Suomalainen nimikirja, Suomen kielen sanakirjat.* Helsinki, Otava. 5–226.

ZILLIACUS, KURT 1966. *Ortnamnen i Houtskär. En översikt av namnförrådets sammansättning.* [Toponyms in Houtskär. An overview of the composition of the onomasticon.] Helsingfors, Svenska Litteratursällskapet i Finland.

Abstract

Since the last third of the 20th century, naming has been seen as pattern-based. These patterns are specific to language, culture and time, and also to the name category so that anthroponyms and toponyms are typically different both in structure and content. The categories have overlapping names and elements, however.

In Finland, the National Land Survey has a database of every toponym on the 1:20 000 map, and similarly the Population Information System includes the names of every Finn was alive c. 1970 or was born thereafter. A snapshot of both of these registers from the start of the current century is available for onomastic research. Comparing these data sets confirms that there is a lot of overlap between surnames and place names; the overlap between given names and either of the first two categories is much smaller but nevertheless interesting. There are multiple reasons for the overlap, dating from the middle ages to modern times.

Keywords: toponyms, anthroponyms, cross-category names, data mining