On the grammaticalization of Finnish colorative construction

Markus Veli Juhani Hamunen
University of Helsinki

This paper concentrates on the diachronic development of the so-called Colorative Construction (CoC) in Finnish, a two-verb expression consisting of an A-infinitive and an ideophonically based descriptive (or 'colorative') finite verb, e.g. *susi **juos-ta jolkottele-e*** [wolf run-INF COL-PRS.3SG] 'wolf runs trotting'. The paper combines variationist dialectal data, grammaticalization theory, and Construction Grammar formalization. The detailed diachronic description demonstrates that the development from proto-CoC to modern CoC is the epitome of constructionalization, i.e., a gradual process of grammatical changes whereby both the form and the function of an existing construction are altered, creating a new expression type. Major changes in the Balto-Finnic case system were the primary force behind this process. Constructionalization of the CoC itself included the first syntagmatic changes through reanalysis. This gradually created a new paradigmatic expression type, followed by paradigmatic extension through analogy, which widened the frame semantics of the newly coined type.

Keywords: Construction Grammar, constructionalization, infinitives, expressive verbs, corpus syntax, dialects, variation, diachronic syntax

1. Introduction[1]

1.1 Overview of the Finnish Colorative Construction

The rich derivative suffix system of Finnish allows for a strong tendency to lexicalize multiple ideophonic roots (Erhard Voeltz & Kilian-Hatz 2001: 1–8) as verbs or nouns. Particularly in dialectal speech, there are countless verbs that denote,

---

for example, the manner of running: *hampsii* 's/he runs with haste', *jolkottaa* 's/he runs jogging without haste', *kaahottaa* 's/he runs fussing', *lipottelee* 's/he runs lightly with small steps', *paarustaa* 's/he runs in a burdersome way', *pinkoo* 's/he runs fast', *taapertaa* 's/he runs toddling', *väännättelee* 's/he runs using much energy', all of them being structurally [run.3SG.PRS] (see Abbreviations and symbols at the end of the paper). These and similar kinds of verbs, which generally fall loosely under the term ideophonic vocabulary, have traditionally been referred to as onomatopoetic, descriptive, or expressive words in Finnish language studies (Mikone 2001; Anttila 1977). Their lexical meaning focuses on the manner of action. This vocabulary is often semantically more or less vague.

Finnish has a specific two-verb construction type called the Colorative Construction (hereafter, CoC). It was so named by the Finnish dialect researcher Ahti Rytkönen (1937). Even though the CoC is relatively infrequent in texts (see below), it has some identifiable structural characteristics. In example (1), we can see that the infinitive (*mennä*) is a semantically 'neutral' verb denoting the type of action in question, while a 'colorative' (COL) or ideophonic finite verb (such as *körötteli*) denotes how the motion is performed.

(1) *Ukko*   [*men-nä körötteli*]_{COC} *hevose-lla kirko-lle*   (made-up example)
     old.man   go-INF   COL-PAST.3SG horse-ADE church-ALL
     'an old man went to church on horseback in a leisurely manner and without haste'

The infinitive is (almost) always the A-infinitive form, which is the verbal form used in Finnish dictionary entries. Grammatical (inflectional) markings are associated with the colorative verb, depending on the syntactic position of the CoC.[2] The most usual word order is INF + COL, as in (1). Other formally parallel examples would be *seistä törrötti* 'stood jutting', *potkaista sätkäisi* 'kicked suddenly', *juosta jolkotteli* 'ran leisurely (without any hurry)' (see also Jarva & Kytölä 2007).

---

2. In standard Finnish, there are three synchronic morphological markers for infinitives: the A-infinitive (e.g. *osta-a* 'to buy'), the E-infinitive (e.g. *istu-e-n* 'by sitting', as in *hän odottaa istuen* 's/he waits sitting ~ sits and waits'), and the MA-infinitive (e.g. *syö-mä-ssä* '(to be) eating', as in *hän on syömässä* 's/he is eating'). Traditionally, Finnish grammarians have used ordinal numbers for labeling infinitives and the number of infinitival forms on the whole has been under debate. See more in Karlsson (1999) and ISK (§492) for an appropriate overall view of the Finnish infinitive paradigm. However, a typologically more plausible and cogent description is Ylikoski's (2003), where non-finite forms are divided into action nominals, participles, converbs, and infinitives in terms of their word-classes and syntactic functions. In this respect, the Finnish A-infinitive form *ostaa*, for example, would be a true infinitive, whereas the E-infinitive form *istuen* and the MA-infinitive form *syömässä* would be converbs (Ylikoski 2003: 203–204).

As Jarva & Kytölä (2007) and my own data show, the CoC principally has four potential synchronic variants, depicted in Table 1. The rows in Table 1 differ according to finiteness and columns differ according to word order. The word order may differ (cf. i, iii vs. ii, iv) as well as the inflection of the neutral verb, according to finiteness (cf. i, ii vs. iii, iv). From now on, I will refer to the type (i) as INF + COL, to the type (ii) as COL + INF, to the type (iii) as FIN + COL, and to the type (iv) as COL + FIN. Moreover, examples like *ukko men-i köröntel-lä* [ukko. NOM go-PAST.3SG COL-INF] are ungrammatical and they do not emerge from the empirical data since only the neutral verb (*mennä*) can be inflected for finiteness. The colorative verb is always in a finite form (whenever the CoC is a predicate).

Table 1. Four potential variants of the CoC

(i)	INF + COL		(ii)	COL + INF	
	men-nä *köröttel-i*			*köröttel-i*[…]* *men-nä*	
	go-INF COL-PAST.3SG			COL-PAST.3SG go-INF	
(iii)	FIN + COL		(iv)	COL + FIN	
	men-i […] *köröttel-i*			*köröttel-i* […] *men-i*	
	go-PAST.3SG COL-PAST.3SG			COL-PAST.3SG go-PAST.3SG	

* […] = other constituents are possible between the verbs

From the cognitive-constructional point of view, the peculiarity of the CoC is that the grammatically prominent verb is semantically background-like and vice versa (cf. a parallel phenomenon in this respect in Kim 2012). The colorative verb is grammatically prominent because it takes the morphological coding, that is, suffixes generally associated with finite verbs; yet semantically, it is more or less backgrounded since it primarily designates the manner of action. Meanwhile, the infinitive is grammatically less prominent because it is a non-finite form. However, semantically, the infinitive denotes the type of action in a neutral and general way, e.g. motion in (1). It represents the overall action frame and, hence, can be understood as semantically prominent. The colorative verb characterizes or specifies and usually emphasizes in some way adverbial manner meanings of that very same action, e.g. 'leisurely and without haste', as in (1). In this sense, there is a rather clear division of functional labor between the verbal elements of the CoC. In sum, this synsemantic nature is synchronically very much a semantic motivation of the CoC as a productive and conventional verb construction in Finnish.

Nominal expressions like *poja-n.kloppi* [boy-GEN.boy.NOM] 'a young boy who behaves badly' function in a similar way. This example consists of a neutral noun in the genitive case (*pojan*), designating the entity type in question, that is [+human, +male, +young], while a colorative noun (*kloppi*) designates with its lexical meaning the kind of boy in question, i.e. 'a young, badly behaving lad'. The

synsemantic character of the construction is due to the nature of the colorative vocabulary. Denotatively, such expressions are so clearly manner-related or, in terms of nouns, adjective-like, that usually they are more comprehensible when they occur in the company of other words. Exactly in this relative sense, A-infinitive verbs in CoCs can be comprehended as neutral. These verbs name an action frame that is semantically modified by manner meanings of a colorative verb. The lexical semantic relationship between the A-infinitive verb and the colorative verb is that of hyponymy and hyperonymy.

Finnish colorative verbs are very often lexicalizations and lexical conventionalizations from sound-symbolic or ideophonic roots, being onomatopoetic, imitative, and somewhat descriptive in nature. For example, the verb *hihi-ttä-ä* means 'to laugh in a snickering or giggling way', where *hih(i)* represents an imitation of laughter (*hi-hi-hi-hi*) and *-ttä-* is a derivational causative verbal suffix (*-ä* is an infinitive suffix). In particular, dialectal data show that colorative verbs can be relatively stable and are widely recognized by native speakers when conventionalized (2) or they may be more regionally distributed when bounded by a dialect area (3). According to the SMS corpus (see Section 2), the verb *kitvetellä* seems to be regionally bound to eastern dialects or Northern Savolax. Sometimes, colorative verbs are even occasional lexical formatives because of the relatively free word formation schemas available in the colorative vocabulary, e.g., *hehettää ~ hahattaa ~ hohottaa* 'to laugh' < $hV_ihV_ittä$- (V = vowel). Both (2) and (3) have slight semantic differences and idiosyncrasies, which are not always easy to define out of context.

(2)   *jolkottaa* 'to run slow and in a relaxed way'

(3)   *kitvetellä* 'to run slow and in a relaxed way'   (SMS, s.v. *kitvetellä*)[3]

The other form of occasionality in colorative verbs comes from their manner-focusing semantics. For example, the colorative verb *ahkuta* (see SMS, s.v. *ahkuta*) forms a CoC with verbs of motion (4a) and devouring (4b), and with a verb describing weather (4c).

(4)   a.   *Savu-a   tul-la   ahkua-a   uuni-sta*   (made-up)
            smoke-PAR come-INF COL-PRS.3SG oven-ELA
            'Thick smoke comes from the oven.'

      b.   *Se   syö-dä ahkua-a   kaikki ruoa-t*   (made-up)
            s/he eat-INF COL-PRS.3SG all   food-NOM.PL
            'S/he eats greedily all the food.'

---

3. SMS refers to one of the corpora, *Suomen murteiden sanakirja* (Dictionary of Finnish Dialects), used in this paper (section 2). The abbreviation s. v. means 'sub voce' (plural: 'vocibus') or 'sub verbo' (plural: 'verbis'), a conventional way to refer to dictionary articles.

c.   *Ulkona tuul-la   ahkua-a   kovasti*   (made-up)
outside blow-INF COL-PRS.3SG greedily
'It is blowing hard outside.'

Basically, *ahkuta* vaguely denotes the manner of action in general, and only be-
cause of that is it able to form a CoC with verbs of agentive and non-agentive
events, as well as with motion and non-motion verbs. In this respect, *kitvetellä* (3)
is similar: *juos-ta ~ pala-a ~ sata-a kitvettele-e* [run-/burn-/rain-INF COL-PRS.3SG].
This is the very reason or the functional explanation for the CoC-construction
to exist. The neutral A-infinitive verb specifies the overall frame, which is then
semantically modified by the colorative verb.[4]

1.2 Aims and organization

The aim of this paper is to describe and demonstrate the grammaticalization of the
CoC. The genesis of the CoC is an example of constructionalization, in the sense
that a new construction emerges as both the form and the function of the exist-
ing expression type changes (see Rostila 2004; Traugott & Trousdale 2014: 22). In
this respect, constructionalization of the CoC differs from constructional changes
that concern only certain features of an existing construction, such as only form
or only meaning. The process of constructionalization is gradual in a two-fold
way: first, syntagmatic reanalysis goes through a semantic ambiguity stage, since
the morphophonological changes are themselves gradual. Second, as reanalysis is
brought to a close, it establishes a new paradigmatic slot in a grammar that can
then expand gradually via analogy. This two-fold gradual constructionalization of
the Finnish CoC is the primary focus of this paper.

The organization of the paper is as follows. In Section 2, the dialectal data on
which the research is based will be introduced. Traditionally, grammatical varia-
tion has been seen as a marker of potential change in a language, but in a similar
way it can be seen as evidence for past changes as well. This is obvious in the
current variation of the Finnish CoC and for that reason, data from dialectal and
colloquial registers are essential. In section 3, the theoretical basis of the research
will be introduced. There are two central issues here: (1) Construction Grammar
offers excellent opportunities to describe explicitly the gradual nature of gram-
matical changes including the relevant formalization and (2) grammaticalization
in general can be seen as based on the cyclic character of changes, i.e. changes usu-
ally affect several ranks in a grammar (see Figure 1). Section 4 presents the current

---

4. Examples (4a–c) are all made-up, representing standard Finnish, but I have modified them
from examples on the SMS corpus; s. v. *ahkuta*.

research and is divided into subsections: in 4.1, previous suggestions on the possible diachronic development of the CoC will be discussed (4.1.1), after which I will present my proposal for a potentially reconstructed proto-CoC (4.1.2). In section 4.2, the constructionalization process of the CoC will be explained. Section 5 summarizes the outcome of the analysis.

2.   Data

The data for this paper come from three different sources of dialectal Finnish. These can be seen in Table 2 along with the number of tokens of the CoC gathered from each corpus.

Table 2. Tokens of the CoC in the LA, DMA and SMS corpora

Corpus	n (CoC)
Finnish Syntax Archive (LA)	84
Digital Morphology Archive (DMA)	408
Digital Dictionary of Finnish Dialects (SMS)	1249
Σ	1741

The total number of CoCs in these data consists of 1,741 examples. Because of the search criteria, the LA corpus (*Lauseopin arkisto*)[5] and the DMA corpus (*Digitaalinen muoto-opin arkisto*)[6] contain mostly CoC types (i) INF + COL and (ii) COL + INF (see Table 1). In the SMS corpus (*Digitaalinen Suomen murteiden sanakirja*), there are only a few effective search constraints available for the present

---

5. The LA corpus is syntactically and morphologically coded but the CoCs have to be manually separated from other syntactic structures that include an A-infinitive as an adverbial. The corpus is based on the XML database, which operates with xpath expressions. All the CoCs from this corpus have been separated from the raw data (338 hits) based on the following xpath command:

XPath: //cl[w[contains(@mrp, 'inf1') and @fun='advl:v']]/ancestor-or-self::s.

The LA consists of two subcorpora: a dialect corpus (spoken language) and a corpus of standard Finnish (written texts); only the first has been used here.

6. There are 408 manually gathered examples of the CoC from the DMA corpus. The total number of A-infinitive expressions amounted to 15,192 instances (at the search moment). In the DMA, CoCs can be found under the signum 740 (= active 1. infinitive lative forms).

topic.[7] The SMS is actually a digitized version of the printed volumes of *Suomen murteiden sanakirja* (Dictionary of Finnish dialects). I have used volumes 1–8 of the SMS, which cover the alphabetical range *a-kurvottaa* ('a-to laugh-COL/to lie-COL'). Consequently, all CoC types (i–iv) are in principle available but the two finite variants (FIN + COL ~ COL + FIN; cf. Table 1) are found only rarely (see also Jarva & Kytölä 2007: 242).

By and large, each of the three corpora relies on very different compositions and usage purposes. However, they have been founded on principles common in traditional dialectology, namely, informants have largely been elderly people in order for the dialectologist to be able to gather sufficiently 'original, proper and unbiased' dialect speech, which is also 'rich in content and fluent without articulatory setbacks' – to use the common characterization in traditional dialectology. The LA, DMA, and SMS corpora represent all dialects spoken in Finland in the twentieth century. In this respect, the SMS represents the oldest material, gathered mostly during the 1920s–40s by linguistic fieldworkers using printed questionnaires, while the other two resources are based on interviews with dialect speakers, largely during the 1950s–70s (LA) and the 1960s–80s (DMA). In addition, most of my data, i.e. DMA and SMS, lack a broad linguistic context. The genre in LA is basically interview material consisting of long narrative episodes.

The other crucial notion is that of the three corpora, only one, LA, is suitable for making proper statistical analyses. With data from the LA corpus, for example, it is possible to count frequencies of a certain language phenomenon or even to use other, more sophisticated quantitative methods. The size of the LA corpus is 1,194,843 words. Neither the DMA (841,755 words) nor the SMS corpus can be used with such methods. However, I find it necessary to represent in numbers how many authentic examples (instances not made up by the author's introspection) there are as well as their distribution with respect to the basic types of CoC (Table 1). Basic quantitative descriptions provide necessary elementary knowledge of the phenomenon, even if elegant statistical methods (that is, explanatory and predictive) cannot be applied.

---

7. Because the SMS is built on a Filemaker database, it accepts only simple and cumbersome regular expressions for data gathering. Thus, several ways have to be used to search and even then the result will not be exhaustive. The search expression <*koloratiivi.**> was used for the CoCs in volumes 1–8 of the SMS. With this method, 220 lexical entries were found and 1,249 examples were manually separated from these entries. Basically, this means that 220 lexemes (in volumes 1–8) have the following definition: "[the word/verb used] usually in Colorative Constructions". In general, the SMS is suitable enough for current purposes, particularly because it usually gives important semantic explanations for otherwise uncommon colorative verbs and it also reports geographical information.

According to previous research, however, narrative texts seem to be a kind of default context for the CoC as a prominent idiolectal choice of the individual language user (cf. Heikkinen & Voutilainen 2009 and the references therein). What is worth pointing out in particular is that the CoC appears more frequently in the spoken language than in the written language (Ikola et al. 1989: 304). Thus, LA, DMA, and SMS are the most relevant corpora available. From the more general point of view, the dialect corpora are representative, especially when it comes to diachronic studies, as they show the full extent of variation within certain grammatical structures. However, it is obvious that the CoC is highly infrequent, even in dialects, despite its otherwise prominent character in narratives. According to the LA corpus, in dialect speech only 0.05 percent of all clauses (one clause out of 2,000) include the CoC, while with modal verbs, the A-infinitive (e.g. *pitää mennä* 's/he has to go-INF') occurs in approximately 0.25–0.35 percent of all clauses (2.5–3.5 clauses out of 1,000) (see also Ikola et al. 1989: 304; Herlin et al. 2005: 19).

3.   Theoretical basis

In this paper, Construction Grammar (CxG) is used as the main theoretical framework (e.g. Fillmore & Kay 1995; Fried & Östman 2004; Boas & Fried 2006; Goldberg 1995, 2006). The following general characteristics favor applying the constructional framework here with certain ideas of the grammaticalization theory (e.g. Lehmann 2002; Traugott & Heine 1991; Heine et al. 1991; Bybee 2010). First, CxG is a sign-based grammar model that takes the notion of a grammatical construction (that is, the CoC) as the basic unit of linguistic analysis (see Fried & Östman 2004: 12). Thus, constructions are form-meaning pairings; hence, change in any aspect of a construction changes the whole linguistic pattern. Second, CxG is also a usage-based model (see Barlow & Kemmer 2000) that does not avoid *parole*. Here, usage data are important precisely from the point of variation, as variation is generally considered a prerequisite for a linguistic change. Third, CxG's formalism is able to show many subtle features explicitly and demonstrate detailed piecemeal mechanisms involved in such parts of grammaticalization as bleaching, reanalysis, and analogical extensions (e.g. Fried 2009: 263).

Combining principles of grammaticalization and CxG is not unheard of (as examples, see Traugott & Trousdale 2010, 2014; Bergs & Diewald 2008; Diewald 2007; Traugott 2008a, 2008b; Noël 2007; Rostila 2006). However, it is still a relatively new approach. CxG, as mentioned, offers some explicit formalization tools for performing language change. Moreover, grammaticalization principles are needed to give theoretical support to internal reconstruction (Ringe 2008; Anttila

1973: 317), which needs a full account of synchronic structural variation as shown by dialect data.

Here, I will use the cyclic character of grammaticalization as a general frame when describing the development of the CoC. Figure 1 shows that the basic idea of grammatical change lies generally in the fact that changes tend to proceed throughout a grammatical system and exert their influence at different levels (ranks). Here it is important to understand that, in principle, the outset for a change can be at any level in this cycle and that changes can go forward and have their influence on any level 'above' or 'below' the outset rank. However, according to the constructional approach, a change at any level of the cycle affects a construction holistically. Moreover, it is not generally necessary to consider grammatical changes dichotomously as either external or internal, that is, based always either on language usage and objectives in discourse or on an (autonomous) grammatical system. There are many kinds of grammatical changes, such as morphological and semantic changes, syntacticization, lexicalizations, and so on; these can have different 'vegetative points' in the rank hierarchy, depending on the individual language and phenomena in question.

> . . . → discourse → syntax → morphology → morphophonemics → (zero) →
> morphophonemics → morphology → syntax → discourse → . . .

Figure 1. Grammaticalization cycle

Figure 1 does not represent any one model of grammatical change, but as can be seen, it is reminiscent of Givón's (1979: 208–209) discourse-based notion of the cyclic character of syntacticization. However, Givón's idea is that syntactic structures are always based on pragmatics, that is, they are crystallized discourse patterns emerging through routine usage. As Givón puts it (1979: 232), "pragmatics gives rise to syntax, syntax in turn gives rise to grammatical morphology, which then decays via phonological attrition." From Givón's point of view, grammatical change always proceeds in the same order, from discourse to the lower ranks. However, as I see it, there is no single or strict model for grammatical change as a gold standard except perhaps some general principles of grammaticalization (e.g. unidirectionality, structural bleaching, and the like). Instead, some changes may be seen more as discourse-based (such as clefting in syntax), while others may be seen as more system-based (for instance, changes in case systems). These perspectives on language change represent different sides of the same coin. Different changes also involve different time scales. In principle, most historical processes are ongoing, yet language users are at any given point in time only subconsciously aware of these processes, driving synchronic (natural) systems to infinite variationist dynamism (see also, e.g., Bybee 2010).

I will use the grammaticalization cycle proposed in Figure 1 to present my theory of the development of the CoC. The benefit of using the cycle is that it is neutral with regard to different theoretical orientations (such as formalism and functionalism), but principally it takes the whole language into account. The idea of a cycle is general enough both to cover all the ranks in a language organization and to enable the description of different linguistic phenomena.

4.   Grammaticalization of the CoC

4.1 Historical background: the proto-CoC

In this section, I will lay out some of the relevant background for the grammaticalization process. First, in section 4.1.1, I will present Jarva & Kytölä's (2007) ideas on the development of the CoC and then in 4.1.2, I will formulate my own suggestion for the proto-CoC.

4.1.1   *Previous research*

At the end of their article, Jarva & Kytölä (2007) briefly "attempt to sketch out some factors which may explain how the CC [= CoC] has developed". Their line of reasoning (as set forth on pp. 268–269) is two-fold. First, the authors compare three different variants of the CoC: type (i) INF + COL, type (ii) COL + INF, and type (iii) FIN + COL (see Table 1). Their main idea is to determine which of these types could historically be the original variant. Since asyndeton is known to be old and widely distributed among Finno-Ugric languages (Karelson 1958), type (iii) FIN + COL, would for that reason be a good candidate for the most archaic variant.

However, there are some reservations about drawing this conclusion. The most obvious one is the question of how the other two variants would have developed on the basis of (iii) FIN + COL, as, according to Jarva & Kytölä (2007: 268), "some explanation is needed why a neutral verb has changed to a non-finite form"; in other words, why type (iii) FIN + COL changed into type (i) INF + COL and then further into type (ii) COL + INF. I suggest that this was not the case. First, that kind of development is not very natural in language change. Natural grammatical changes seldom make such leaps. Second, there is no empirical evidence for such a progression.

Although it is true that there can be syncretism between some A-infinitive forms and 3rd person singular present verbs in some Finnish dialects, especially in the written language (see example 5a, where *kantaa retuuttaa* can be interpreted either as type (i) INF + COL or type (iii) FIN + COL CoC), syncretism cannot be regarded as a historical link demonstrating the grammaticalization of the CoC.

This syncretism is a much later development which occurred long after the actual grammaticalization of the CoC. In fact, it is not clear to what degree the asyndetic two-finite-verb variants, that is, types (iii) FIN + COL and (iv) COL + FIN, have ever been established, as expressions with loose paratactic verb chains also occur in the spoken language (see example 5b, where the comma after *näkkyy* indicates a short pause in speech).

(5) a. *Karhu kanta-a retuutta-a pentu-j-a-an pesä-lle*
bear.NOM carry-INF COL-PRS.3SG cub-PL-PAR-3PX den-ALL
carry-PRS.3SG (made-up)
'A bear carried his/her cubs to a den dangling.'

b. *mikä kumma siellä näkky-y, huamotta-a* (SMS)
what strange.NOM there be.visible-PRS.3SG COL-PRS.3SG
'What on earth is [it] there, looming like a ghost.'

Therefore, I would argue that variant (iii) FIN + COL, even though it can be seen as connected to the variants (i) INF + COL and (ii) COL + INF, would not be the original outset for the grammaticalization of the CoC.

Second, after Jarva & Kytölä (2007: 269) propose that variant (ii) FIN + COL is older than variant (i) INF + COL (see below), they compare variant (i) to other Finnish infinitive structures and suggest that the CoC "has its origin in infinitive clauses rather than in asyndetic juxtaposition". The next quotation elucidates their argument (Jarva & Kytölä 2007: 269):

"How is it possible that an expressive construction [= CoC] has adopted the same syntactical structure as modal, permissive and otherwise abstract infinitive clauses? To some extent the infinitive verb could be understood as an explanation or a specification of the finite verb; therefore the infinitive clause *minä tahdoin mennä* 'I wanted to go' [– –] could be paraphrased as

(62) ˀ*minä tahdoin niin, että menisin*
I want-PAST-1SG so that go-PAST-COND-1SG
?'I wanted it to be so that I'd go.'

Similarly, the CC [= CoC] *susi **lotkotti** mennä* 'the wolf-COL [went relaxed and sluggishly, transl. M.H.]' [– –] could be paraphrased as

(63) *susi lotkotti niin, että se meni*
wolf COL-PAST-3SG so that it go-PAST-3SG
'the wolf COL-ED so that it went.'

> Whatever the development has been, variant (a), *kaatua tupsahdin* [= type (i) INF + COL], is its endpoint."[8]

Now, as reasonable as it would be to compare syntactic similarities between the infinitival structures, in light of the above quotation it is hard to see the diachronic connections between them. The Comparisons are simply too granular and general. Instead, Jarva & Kytölä's comparisons represent synchronic paraphrases more than a diachronic development. As early as 1880, the renowned Finnish scholar E. N. Setälä, in his *Suomen kielen lauseoppi* (*Finnish syntax*), had a similar idea about "an infinitive of relationship" in his analysis of the CoC: *Hän juosta hölköttää* 'S/he runs in a slow and relaxed manner' = *hölköttää juoksemisen suhteen* 'as for running, s/he lollops' (Setälä 1880, s. v. Nominaalimuodot [Nominal forms]). It is possible that other infinitive structures have had their impact on the development of the CoC at some point, but it is unlikely that this occurred in the early phases of the process. It is more likely that the Finnish infinitive system developed as a whole and for this reason the same infinitive functions in a similar way in relatively different infinitive structures, which enables paraphrasing as well.

All in all, the above-mentioned descriptions are reasonable enough but they still leave many open questions, such as how the current CoC might have gradually developed structurally and functionally from clausal expressions. Related constructions (such as types (iii) FIN + COL and (iv) COL + FIN) and synchronically comparable paraphrases cannot, on their own, adequately demonstrate the actual historical development of the CoC. Possible historical links between paraphrased expressions remain primarily implicit. Thus, even though Jarva & Kytölä (2007: 269) show genuine interest in a historical explanation, their purely synchronic description still remains too brief and speculative from a diachronic point of view, as they themselves admit.

4.1.2   *Reconstruction of the proto-CoC*

Jarva & Kytölä's (2007: 269) idea that variant (ii) COL + INF is the oldest (although without the inference that the CoC would have its origin in other infinitive clauses) is supportable. This fact is a suitable starting point for outlining the relevant

---

8. Jarva & Kytölä (2007: 269) have glossed the verb form *menisin* 'I would go' as past tense even though morphologically, it is present tense (PRS). In this particular context, however, *menisin* might be interpreted temporally as past because the finite verb is in the past tense. Nevertheless, morphologically, conditionals are considered not to have simple past tense forms in Finnish but only present (*menisin*) and perfect forms (*olisin mennyt* 'I would have gone'). Jarva & Kytölä refer to their own examples (11) and (6), which are as follows: (example 11, on p. 245) *minä tahdoin mennä sen ojan ylitse* 'I wanted to go across that ditch' and (example 6, on p. 242) *susi lotkotti mennä erellä* (LA) 'the wolf went-COL ahead'.

background for the development of the CoC. The first piece of evidence for the archaic character of the COL + INF variant is that there can be other constituents between the verbs (6a–b).

(6) a.   *si-llä sitä nylkytet-tiin **sittek** kääntä-äs   sitä pelto-o*   (LA)
it-ADE EXPL COL-PASS.PRT then turn.over-INF that field-PAR
'with that it was turned over bit by bit, that field'

   b.   *siält viilettä-ä   **yks niim punaise-s pusero-s**   tul-la*   (LA)
there COL-PRS.3SG one very red-INE   sweater-INE come-INF
'here comes rapidly (straight ahead) someone wearing a very red sweater'

As can be seen in examples (6a) and (6b), word order with COL + INF can be less coherent because it allows other elements between the verbs and, therefore, it is a less grammaticalized variant as well. I find this to be one piece of evidence for the fact that type (ii) COL + INF is an older variant than the type (i) INF + COL.

Table 3. Word order variations of the CoC in the data

	word order				
CORPUS	(i)	*men-nä köröttel-i* go-INF COL-PAST.3SG	(ii)	*köröttel-i* […]* *men-nä* COL-PAST.3SG go-INF
	(iii)	*men-i* […] *körötteli* go-PAST.3SG COL-PAST.3SG	(iv)	*körötteli* […] *men-i* COL-PAST.3SG go-PAST.3SG
LA	67 out of 84 hits		17 out of 84 hits	
DMA	359 out of 408 hits		47 out of 408 hits	
SMS	975 out of 1,249 hits		274 out of 1,249 hits	

* […] = other constituents are possible between the verbs

The word order variant COL + INF represents a minority portion of my data, as can be seen in Table 3. Most of the dialectal material represents variant (i) INF + COL (rarely also (iii) FIN + COL), while variant (ii) COL + INF (and rarely (iv) COL + FIN) is clearly in the minority. The variant (i) INF + COL is even more dominant in standardized written Finnish, while the inverse order is marginal, if not absent altogether, in standard registers (see ISK §450; Saukkonen 1966: 134). Moreover, the same can be said for variants (iii) COL + FIN and (iv) FIN + COL, which both have verbs in a finite form (e.g. *siit ko **maattii pössöteltii*** [PASS.PRT] 'then we rested (and) chilled out', DMA). In the LA corpus, there are no examples of this type. In the DMA corpus, there are four instances of the variant (iii) FIN + COL and none of the variant (iv) COL + FIN. Likewise, in the SMS corpus, a total of 200 cases can be found and the majority of them are type (iii) FIN + COL.

The argument concerning word order can be strengthened with evidence from other Balto-Finnic languages, as shown in the following examples (7a–7e):

(7) a. *siel   vain hapute-ttih   juosta*   (Karelian 2 out of 19) (KKSK)
         there just COL-PASS.PRT run-INF
         'there we/they just ran quickly'

     b. *paukaiž   ambu-da*   (Ludian 7 out of 21) (Virtaranta 1986: 154)
         COL.PRT.3SG shoot-INF
         's/he shot with a blast'

     c. *röpöta-b   bas-ta   bepsa-ks*   (Vepsian 10 out of 15)
         COL-PRS.3SG speak-INF Vepsian-TRA
         's/he speaks Vepsian in an unclear way'   (Kettunen 1943: 155–158)

     d. *hepoin hötsütä-b   joos-sa*   (Votic 10 out of 15)
         horse   COL-PRS.3SG run-INF
         'the horse ran in an easy manner'   (Posti 1980: s.v. *hötsütäB*)

     e. *sumista-b   lau-lda*
         COL-PRS.3SG sing-INF (standard Estonian 27 out of 27) (Põlma 1967: 89)
         's/he sings humming'

Examples (7a–e) represent the variant (ii) COL + INF. The numbers after the name of the language show how many cases of this variant exist compared to all instances of the CoC in that reference. Even though this material is scarce, it can be found in most Balto-Finnic languages: especially in Estonian (7e), and mostly in Vepsian (7c) as well as in Votic (7d). That Karelian (7a) and Ludian (7b) seem to have less of the variant (ii) COL + INF is probably because the other instances are all finite variants (mostly FIN + COL). (See also the data in Saukkonen 1966: 134–138.) One can thus conclude that the variant (ii) COL + INF might be, historically, the original one: this variant can be found in all Balto-Finnic languages in which the CoC occurs, whereas the type (i) INF + COL cannot (cf. Saukkonen 1966: 137).

Concerning the outset of the grammaticalization of the CoC, I have excluded the finite verb variants (iii) FIN + COL and (iv) COL + FIN as unlikely original structures. Instead, I have argued that the less cohesive infinitive variant (ii) COL + INF without any necessary connection to infinitive clauses represents the oldest type. Jarva & Kytölä (2007: 268–269) attempted to outline some factors for the development of the CoC but they did not take into account a particular diachronic aspect of the A-infinitive form. This concerns a certain morphophonemic phenomenon of the A-infinitive in current spoken Finnish. In examples (8a–b), the zeros (-ø) after the A-infinitive forms allude to this issue. As can be seen, in standard written Finnish (8a), there is nothing orthographically distinctive nor, in certain dialects (8b), is there anything phonetically discernible.

(8) a. *Koirat juos-ta-ø jolkutt-i-vat   takaisin koti-in.*
dogs   run-INF   COL-PRT-3PL back   home-ILL

(standard Fin., made-up)

'The dogs ran slowly and [in a] relaxed [manner] ~ trotted back home.'

b. *ihmist naura-a-ø rähäht-i-it*   (DMA)
people laugh-INF COL-PRT-3PL

'people laughed out loud'

However, in some contemporary dialects and in spoken Finnish in general, these zeros are still widely realized phonetically, as seen in examples (9a–b). Although this phenomenon is usually more or less inconspicuous for most speakers of present-day Finnish, it represents a diachronically significant grammatical trace of the development of the CoC (as well as other infinitive clauses with an A-infinitive form).

(9)   *se kuol-ta-**k** kupsah-t*   (DMA)
it die-[INF]^x COL-PRT.3SG

'it ~ s/he died suddenly'

savu   nous-ta-ʔ ʔängött-i   (DMA)
smoke rise-[INF]^x COL-PRT.3SG

'smoke rose very slowly'

The morphophonemic zeros of the A-infinitive suffix in examples (8a–b) are realized in examples (9a–b). They are glossed with the superscript x. As a historical trace, this x-feature is realized in a way that depends on the phonetic context that follows. Briefly, when a subsequent morpheme (e.g. a lexical unit) begins with a consonant, the zero occurs as phonetically homorganic (9a), whereas when a subsequent morpheme begins with a vowel, the zero is realized as a glottal stop (9b).

This can be considered a standard example of synchronic phonetic assimilation.[9] (For more on this issue, see Itkonen 1964.)

The morphophonemic x-feature described above is in fact assumed to be a historical trace of the Proto-Finnic terminal *k-lative case (see Hakulinen 1961: 76–78; Leino 2005: 109; Lehtinen 2007: 93; Ylikoski 2003: 213, 225; however, cf. Ylikoski 2011: 240, 248, 255–256). Hakulinen describes a few suffixes (*-k, *-s, *-n, and *-i), all of which have been considered as previous motion or direction implying lative case forms. Furthermore, Hakulinen (1961: 76) presents examples such as *toi vettä janoisten juoda*$_{INF}$'s/he brought water for the thirsty **to drink**', *sopii tehdä*$_{INF}$ 'it is all right **to do** so', *paneutui maata*$_{INF}$ 'put him-/herself **to lie down**', *oli pudota*$_{INF}$ 'was **to fall**' [boldface added]. Hakulinen argues that the original lative function is still evident here even though a clear morphological marker no longer exists.

With this historical information, we are able to suggest some potential reconstructions to represent the putative original structures of the CoC. As claimed above, the original word order variant seems to be type (ii) COL + INF with *k-lative in the A-infinitive form. Hence, examples like (10a–c) would act as plausible candidates for reconstructed proto-CoCs (SWF = standard written Finnish, LPF = Late Proto-Finnic).[10]

(10)   a.   *koira pyr-äšt-i   joos-ta-k   (LPF)
            koira pyr-äht-i   juos-ta^x (SWF)
            dog   COL-MOM-PRT.3SG run-INF-LAT
            'the dog suddenly started to run'

---

9. The structural assimilation rules are as follows: $^x \to C_i$ / __ # C_i (9a) and $^x \to \textipa{P}$ / __ # $^{\textipa{P}}$V (9b). At the very end, x does not occur at all. Even clitic suffixes (e.g. [nousta:kin] *nousta* 'to rise' + *kin* 'also') and composed words (e.g. [herne:keitto] *herne* 'pea' + *keitto* 'soup', 'peasoup') trigger the assimilation. In the Finnish linguistic tradition, this phenomenon is called *rajageminaatio* 'juncture gemination', *loppukahdennus* 'end duplication', and *alkukahdennus* 'initial duplication', among other designations. In fact, there are several specific morphological forms in current spoken Finnish, both nominal (i.e. allative case, 3. person possessive suffix, certain nouns ending in -*e*) and verbal (i.e. A-infinitive, 2. person singular imperative, negative present forms), which are related to this assimilation phenomenon. There is a relatively well-acknowledged consensus on the phenomenon's historical origins but synchronically, it is not always clear what kind of feature it actually is. In the glosses for (9a–b), the x-feature is attached with brackets to the A-infinitive morpheme specifically to indicate that it is a morpheme-based relic. As the synchronic issue is not relevant from a historical point of view I will not deal further with it here.

10. It is crucial to note that examples (10a–c) do not necessarily represent accurately any definite proto-linguistic stage but merely demonstrate the characteristics that the proto-CoC probably had. In this paper, the same principle applies to all the reconstructions addressed in the text.

b. **mees roj-ašt-i*   *puðo-ta-k*   (LPF)
    mies roj-aht-i   *pudo-ta^x*   (SWF)
    man COL-MOM-PRT.3SG fall.off-INF-LAT
    'the man fell off suddenly, helplessly, in a drooping or slouching manner'

c. **Puu rom-ašt-i*   *kaatu-ða-k*   (LPF)
    puu rom-aht-i   *kaatu-a^x*   (SWF)
    tree COL-MOM-PRT.3SG fall.down-INF-LAT
    'the tree suddenly began to fall down'

These examples are reconstructions and with features assumed to be the most relevant ones according to the grammaticalization of the CoC. As tokens of the proto-CoC, they represent the syntactic outset for the change and the starting point for proceeding along the grammaticalization cycle laid out in Figure 1. Syntactically, examples (10a–c) represent intransitive structures and semantically, they represent concrete spatial motion. Motion is manifested by lexical meaning of the infinitival verb roots and by the **k*-lative case. As a whole, the proto-CoC refers to the change of state of the subject argument and its transition from one state to another; in other words, motion.

In this context, it is reasonable to mention that North Sami has a CoC-like asyndetic verbal concatenation in which both verbs are finite in form: *nieida* ***bijadii njaccuhii*** [PRT.3SG] 'a girl went scurrying' (Jomppanen 2011: 107, 131; Nielsen 1979: 398, 405; Nickel 1994: 399–400). According to Bergsland (1994: 48), even South Sami has a similar pattern called *doppelt predikatverb* 'double predicate verb'. These are reminiscent of the variants (iii) FIN + COL and (iv) COL + FIN of the Balto-Finnic CoC. However, these finite expressions in the Sami languages do not have any genetic connection to the Balto-Finnic CoC, since these asyndetic structures represent looser syntagmatic verb expressions that do not build on constituency as do the prototypical Balto-Finnic CoC types (i) INF + COL and (ii) COL + INF. This also implies that the diachrony of the CoC would not go back earlier than Middle Proto-Finnic. Nevertheless, when it comes to ideophonic verbs in general, it is somehow symptomatic that they often turn up with another explanatorily neutral verb.

4.2 Constructionalization: from proto-CoC to the current CoC

In this section, I present my view of the constructionalization of the CoC. The discussion proceeds diachronically from the proto-CoC to the pre-CoC and eventually to the modern CoC. Sections 4.2.1–4.2.4 follow the grammaticalization cycle and fit the concepts of reanalysis, analogy, and diachronic micro-steps in the general frame of change with CxG formalization.

4.2.1   *Syntax → morphology*

Now that some relevant historical characteristics of the CoC have been intro-
duced, it is possible to present an abstract structure representing the assumed
original proto-CoC (Figure 2). Some basic principles of CxG's formal notation are
adopted here from Fillmore & Kay (1995) and Fried & Östman (2004). According
to Figure 2 and the reconstructed examples (10a–c), the inventory of common
characteristics of the proto-CoC are as follows: the word order was COL + INF, the
lative case of the infinitive indicated direction, the A-infinitive was an intransitive
motion verb, and the colorative verb was semantically momentative. Here, I will
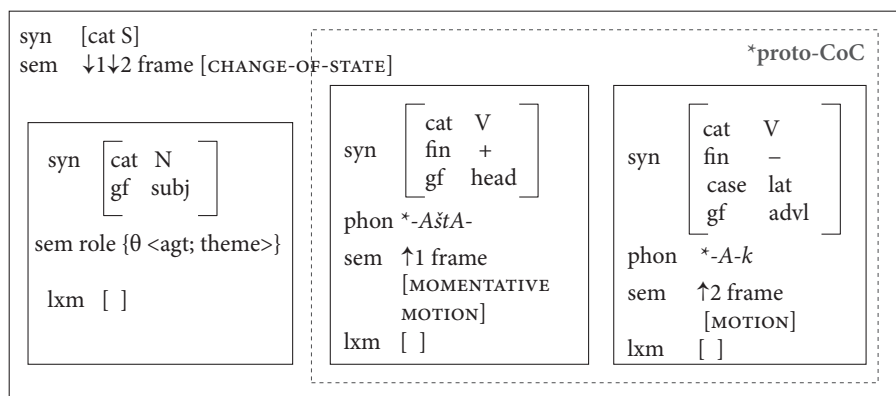use Figure 2 to go through these features along with some others.

syn   [cat S]
sem   ↓1↓2 frame [CHANGE-OF-STATE]   *proto-CoC

syn [cat N / gf subj]
sem role {θ <agt; theme>}
lxm [ ]

syn [cat V / fin + / gf head]
phon *-AštA-
sem ↑1 frame [MOMENTATIVE MOTION]
lxm [ ]

syn [cat V / fin – / case lat / gf advl]
phon *-A-k
sem ↑2 frame [MOTION]
lxm [ ]

Figure 2. The proto-CoC[11]

The proto-CoC was not a fixed construction of a kind similar to the current CoC.
It probably permitted other words between the colorative verb and the A-infinitive
in the same way as the variant (ii) COL + INF permits other words in current dia-
lects and spoken Finnish (e.g. *käire vaa, painettiin* **sinnem** *mennä* 'by walking, we
pushed our way **there**', DMA). In a formalization, this could be indicated by an
empty box between the verbal constructions with the so-called Kleene star (*),[12]
but for ease of presentation, this has been left out in Figure 2. The box with the
broken gray line denotes the proto-CoC as a whole. In addition, the syntactic ana-
lyzability between the elements was still apparent. The *k*-lative case was functional
and productive (syn [case lat]) and the infinitive was a free adverbial (syn [gf advl]).

---

11. It is essential to note that actually only the inner box (marked by the dotted line) in the for-
malization refers to the CoC, but as the construction is verbal I have kept its subject argument
along in all formalizations (cf. Figures 2–5).

12. So-called Kleene star (*) is a symbol in the Construction Grammar box notation indicating
that there might be zero or more constituent(s) in the slot where the symbol is used.

Perhaps the most important features of the proto-CoC as a construction were in its semantics. Originally, finite colorative verbs represented a morphologically closed verb class, namely, the momentative. This can be seen in the derivative suffix for the momentative (phon *-AštA- > -AhtA-* in current Finnish; e.g. *istahtaa* 'to sit down for a short while', *pyrähtää* 'to sprint') and its semantic frame MO-MENTATIVE MOTION. The presumption that the colorative verb was momentative in the proto-CoC is based on the fact that non-momentative coloratives would have formed odd or ungrammatical syntagms with A-infinitives in *k*-lative (e.g. **körötteli mennäk*). Accordingly, the original semantic frame of the infinitive was presumably MOTION, since different motions and movements represent concrete, frequent, and perceivable events in the world. Thus, the proto-CoC as a whole was about intransitive motion of a volitional AGENT or a non-volitional THEME (note the value pool in the sem role {θ <AGT; THEME>}. What distinguished it from ordinary intransitive motion frames was the integration of the semantic frames of both the colorative verb (↑1 frame) and the A-infinitive (↑2 frame). In Figure 2, this is indicated in the largest box by the notation ↓1↓2 frame. This frame integration takes place already within the proto-CoC box, but owing to lack of space, this information is represented only at the sentence level.

What is remarkable is that the semantic composite of the colorative and the A-infinitive frames was not (intransitive) MOTION nor was it MOMENTATIVE MO-TION but, rather, CHANGE-OF-STATE. One explanation for this would be that the finite colorative verb focused semantically on the precise culmination point between two subsequent states – immobility (or stationary state) and motion, alternatively called motion$_1$ and motion$_2$, respectively. Moreover, the momentative verb was inchoative by its *aktionsart* when combined with a motion-denoting A-infinitive in the *k*-lative case. Thus, when conceptualizing the proto-CoC sequence by sequence, we perceive a prototypical event in which a subject-NP is transferred from one state to another. As a state-of-affairs in the real world, these two stages are separated by an abrupt, colorative or ideophone-based process. At the constructional level described in Figure 2, this 'chunk of the world' was conceptualized as the semantic composition of the change (colorative) and the latter state (A-infinitive); hence, CHANGE-OF-STATE in the semantic frame.

This was the most characteristic semantic feature of the proto-CoC as a grammatical construction. First, it distinguished the proto-CoC semantically from other formally equivalent infinitival constructions like V_{FIN} + INF_{LAT}, such as *tahtoo mennä* 's/he wants to go', *lupaa mennä* 's/he promises to go', *pitää mennä* 's/he has to go', and so on. Second, it fits well with Goldberg's (1995: 4) general definition of a construction: semantic aspects of the proto-CoC cannot be predicted from its parts, but rather emerge from the whole. Third, the word order COL + INF_{LAT} of the proto-CoC was iconic in the sense of motivation; i.e., the linear order of two parts

of this syntagm reflects the natural order of events in reality (Haiman 1980). The colorative verb represented a perceptually prominent momentary event between two different states, and the A-infinitive in the lative case designated the latter event, usually motion. The function of the *k-lative case was to indicate the subject's transition to the action or motion designated by the A-infinitive.

It is crucial to understand that the interpretation of the current CoC did not fit the proto-CoC, because the *k-lative case was productive and no caseless infinitive form yet existed. That is why the colorative verb and the A-infinitive form constituted a motivated linear syntagm whereby two successive events were conceptualized as separate. The reason that the MOTION frame was very likely the original context for the proto-CoC can be inferred from two factors. First, concrete spatial motion is perceptually prominent and experientially a very frequent event in the world. Second, the corpus data show that motion is also a single clear event type with the current CoC. In the LA corpus, 37 out of 84 A-infinitives are motion verbs and even though the DMA corpus does not allow statistical methods, it is worth noting that almost half of its CoC examples are about motion. Common infinitives in the corpora are *mennä* 'to go', *tulla* 'to come', *kävellä* 'to walk', *ajaa* 'to drive', *lyödä* 'to hit', *juosta* 'to run', *käydä* 'to visit', *istua* 'to sit', *pudota* 'to fall', and *tehdä* 'to do, to make'.

Based on examples (10a–c), I have also argued that in the proto-CoC, colorative verbs were originally momentative-inchoative. Since the function of colorative verbs in the proto-CoC was to express a sudden change between two states – one inferred from the context and the other explicated by the A-infinitive – it is natural that onomatopoetic or other expressive verbs designating momentary events would have fulfilled that function. The derivative suffix *-AštA- (currently -AhtA-) with its momentative meaning simply emphasized those characteristics. Consequently, colorative verbs focused on the perceptually most prominent part of the MOTION activity, namely, the starting point of movement or the change of state.

In sum, as described above and formalized in Figure 2, the proto-CoC was a syntactic expression that consisted of a momentative colorative finite verb and a motion verb in the infinitive form with *k-lative case. The meaning of this motivated (iconic) syntagm was a sudden change of state. It cannot be categorically denied with positive evidence that the proto-CoC would not have had historically motivated connections to the other infinitive constructions. Rather, the clearest evidence for grammaticalization from the proto-CoC to the current CoC comes from changes in the Balto-Finnic case system morphology. In the following section, I will clarify this development.

4.2.2 *Morphology → morphophonemics (→ zero)*

In order to be able to give a plausible historical description of the CoC, we need a wider perspective on the Proto-Finnic non-finite paradigm. That perspective is represented in Table 4. The table is largely based on Lehtinen (2007) but uses modified terminology by Ylikoski (2003) in order to move in a more typologically plausible direction (see also Haspelmath 2010).[13]

Table 4 presents only those reconstructed non-finite forms that are relevant to the current topic. These are infinitives in directional, internal, and general locative case forms. Moreover, what emerged during the development from Middle Proto-Finnic (MPF) to the latest period of Late Proto-Finnic (LPF) and finally up to modern Finnish were new infinitival construction patterns. It would go too far afield from the central topic here to sketch even a general view of the development of those constructions, but a brief overview of the diachronic progression of non-finite case forms is nonetheless presented. In the MPF era, there were presumably fewer grammatical, non-finite forms, which had multiple functions, i.e., they were more polysemous according to current diachronic theories.

Table 4. The reconstructed historical paradigm of Balto-Finnic A-infinitive and MA-infinitive forms, example *tulla* 'to come' (see Lehtinen 2007: 93, 122, 124, 134; Ylikoski 2003: 203–205, 208, 212–213, 216–217)

Middle Proto-Finnic (MPF)		Late Proto-Finnic (LPF)	
tulða-k	come-LAT	INF	*tulða-(k)* come-Ø
	come-TRA	CONV	*tulðak-kse*(+PX) come-TRA
tulema-nA come-INF/general		INF	*tulema-ssA* come-INE
	locative	INF	*tulema-stA* come-ELA
		INF	*tulema-sen* come-ILL

As time passed, the repertory of non-finites became re-structured so that some of the functions adopted distinct morphological forms (shown by the arrows in Table 4). The operative principle was a tendency toward the one-form-one-meaning principle, which constituted more symmetric subsystems within the non-finite paradigm (see also Ylikoski 2011: Chapter 3). For example, MA forms became

---

13. Early Proto-Finnic is considered to be the diachronic stage when the (Proto-)Sami and Balto-Finnic languages were not yet disengaged (ca. 1500–1000 BCE). Late Proto-Finnic, in turn, represents the historical stage during which the Balto-Finnic languages gradually evolved into individual daughter languages (ca. 1000–0 BCE), breaking off the historical links with Sami. (e.g. Korhonen 1981: 27; Ylikoski 2011: 237–239; Heikkilä 2014.)

symmetric with regard to the internal local cases (illative 'to', inessive 'in', and elative 'from').[14]

Even though it is not possible to depict the entire development of the Finnish non-finite paradigm, it is plausible to infer that the infinitive lative case was part of the bigger picture. Its historical fate (and subsequent loss) through semantic bleaching was the single most crucial factor in the grammaticalization process of the CoC. Thus, to examine the historical development of the CoC as part of the paradigmatic changes in the system of non-finite case markings seems to be more fruitful than reasoning with paraphrases as in Jarva & Kytölä (2007: 269). Synchronically, the CoC might be a bit unusual compared to other non-finite constructions in Finnish, but its historical peculiarities cannot be derived in a straightforward manner.

The proto-CoC was one of those verbal chains whose A-infinitive with a terminal *k-lative was accompanied by a finite verb. Without going further into the details of other non-finite structures – additional background information can be found, for example, in Saukkonen (1965, 1966) and Leino (2005) – it can be argued

---

14. As can be seen in Table 4, the process of restructuring continued and new non-finite forms appeared during the historical development of Finnish as an individual language. The MA form in particular acquired new case markers and the existing non-finites became somewhat idiosyncratic in their paradigms and even partly dependent on a dialect in certain constructions. Moreover, some non-finite forms are already periphrastic or construction-specific, such as the -mAssA form (the MA form's inessive) used in a progressive meaning with the verb *olla* 'to be, exist'; e.g. *hän on juoksemassa* 's/he is running'. As for the present research, there is neither reason nor space to investigate different (synchronic) infinitive constructions further, and even less for considering the historical development of these constructions. Research on the latter topic is also very scarce (however, cf. Leino 2005: 89–120).

Table 5. The morphological paradigm of (active) non-finite forms (except participles) in Finnish from Early Proto-Finnic to modern Finnish (Lehtinen 2007; Ylikoski 2003)

Early Proto-Finnic (PF)	Late Proto-Finnic (LPF)			(modern) Finnish		
tuleta-k come-LAT/TRA	INF	*tulða-(k)*	come-ø	INF	*tulla*	come-ø
	CONV	*tulðak-kse*(+PX)	come-TRA	CONV	*tullakse*(+PX)	come-TRA
	CONV	*tulðe-ssa*(+PX)	come-INE	CONV	*tullessa*(+PX)	come-INE
	CONV	*tulðe-n*	come-INS	CONV	*tullen*	come-INS
tulema-na come- INE/ general locative	INF	*tulema-ssa*	come-INE	INF	*tulemassa*	come-INE
	INF	*tulema-sta*	come-ELA	INF	*tulemasta*	come-ELA
	INF	*tulema-sen*	come-ILL	INF	*tulemaan*	come-ILL
				CONV	*tulema-lla*	come-ADE
				CONV	*tulema-tta*	come-ABE
				(INF	*tulema-n*	come-INS)

that the essential factor behind the historical change was a global re-structuring of non-finite case distribution during the MPF and LPF eras (see Table 4). With respect to the proto-CoC, this means that when the MA-infinitive's illative – illative being a crucial case form for building a symmetric subsystem within the local cases of the Balto-Finnic non-finite system – emerged as a new infinitival form, it began to take over the functional field previously occupied by the *k-lative.[15] This, in turn, led to a process in which the *k-lative began gradually to bleach, first semantically in becoming unproductive and then eroding structurally as well. As for the proto-CoC, the battle between the functionally overlapping infinitival forms gave rise to an ambiguity formalized in Figure 3.
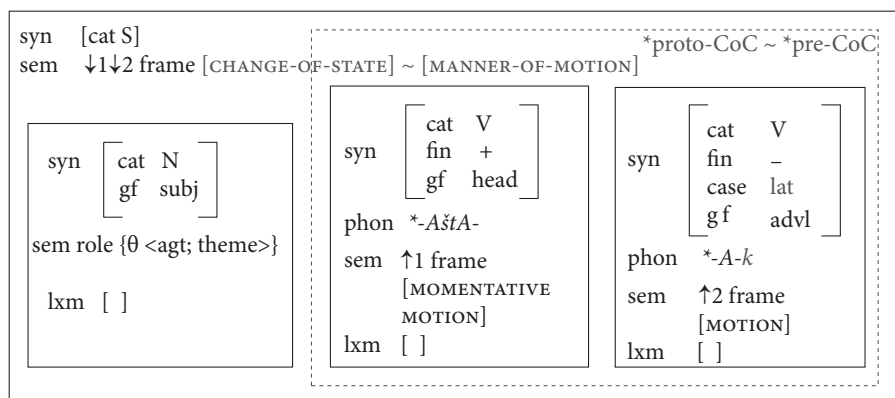
syn	[cat S]		
sem	↓1↓2 frame [CHANGE-OF-STATE] ~ [MANNER-OF-MOTION]		*proto-CoC ~ *pre-CoC

syn [cat N / gf subj]
sem role {θ <agt; theme>}
lxm [ ]

syn [cat V / fin + / gf head]
phon *-AštA-
sem ↑1 frame [MOMENTATIVE MOTION]
lxm [ ]

syn [cat V / fin – / case lat / g f advl]
phon *-A-k
sem ↑2 frame [MOTION]
lxm [ ]

Figure 3. Semantic ambiguity as input for grammatical reanalysis

As the *k-lative was gradually pushed away by the MA-infinitive illative, it ended up in the continuum of piecemeal unproductiveness. This is marked by the gray font in Figure 3. There was a period in LPF when the construction could be interpreted in two different ways, either as the proto-CoC (see section 4.2.1) or as something which one may call the pre-CoC. For instance, example (10a), *koira pyrästi joostak, could be interpreted as 'a dog abruptly started to run' (the proto-CoC) or as 'a dog sprinted' (the pre-CoC). As already argued, the linear ordering of verbs in the proto-CoC had an isomorphic (i.e., motivated) relation to the perceived events in the real world. Its semantic frame was CHANGE-OF-STATE. The proto-CoC was thus iconic. The pre-CoC, however, was not even though the order of elements

---

15. Internal local cases in the MA-infinitive are prototypical in spatial use: *hän meni uima-an* 's/he went to swim' (ILL), *hän oli uima-ssa* 's/he was swimming' (INE), *hän tuli uima-sta* 's/he came from swimming' (ELA). There is a clear semantic symmetry. By comparison, there is just one form in the MA-infinitive external case that does not have a spatial, locative meaning but, rather, an instrumental meaning: *hän meni uima-lla saareen* 's/he went to the island by swimming' (ADE).

remained unchanged. Rather, the pre-CoC with the motion-*cum*-manner interpretation became semantically more compressed than the proto-CoC. This means that the momentative colorative verb eventually began to modify MANNER qualities of the whole process represented by the A-infinitive, not just the transition between two qualitatively different stages. Its frame could thus also be interpreted as MANNER-OF-MOTION, as indicated in Figure 3.

At first, the semantic ambiguity of the proto- and pre-CoC existed coincidentally without having any apparent impact on the linear surface structure COL + INF. Presumably, their interpretation depended on the context. Still, it is somewhat mysterious as to what precisely triggered the interpretation during the ambivalent period and how that happened. One possibility, although not very pronounced, could simply be pragmatic: interpretation was made solely according to the ongoing discourse situation by inference. However, regarding grammaticalization, the more important aspect to consider here is whether there was any semantic ambiguity at all, as this is the true fuel for *reanalysis* (Bybee 2010: 136–150, 199). Moreover, it is important to note that the reanalysis itself was a gradual historical process: a rivalry between different case markings in non-finite verb forms led to constructional and functional ambiguity. This can be seen as the principle of economy in a language system: first, to make a functional distinction between the two forms and second, to establish that same distinction. As for the grammaticalization of the CoC from a reanalysis point of view, this was the picture of the change. Regarding grammaticalization principles in general, it is not just synchronic *structural* variation that stands for a favorable basis of grammatical change, but also semantic ambiguity, which had evolutionary effects as well.

Reanalysis took place when functional overlapping developed into functional separation. In other words, the distinction emerged when the semantic ambiguity evolved into a real semantic difference. Accordingly, the MA-infinitive illative completely adopted the very same functional slot that had previously been governed by the infinitive with the **k*-lative. Grammatical precedents for expressions such as those in the examples (10a–c) can be found in Finnish and other Balto-Finnic languages, as the examples (11a–c) clearly show.

(11) a. *Nuotio   leim-aht-i   **pala-ma-an**   (made-up)
          campfire flame-MOM-PRT.3SG burn-MAINF-ILL
          'the campfire burst into flames'

      b. **nagra-ma-h**   *muh-aht-i   mužikka še*   (Karelian; KKSK)
          laugh-MAINF-ILL COL-MOM-PRT.3SG man   that
          'that man burst out laughing'

c. *nõksata-b*   ***seis-ma***[16]   (Estonian; Mihkla 1964: 92)
    COL-PRS.3SG stand-MAINF
    's/he stands up quickly'

Although the proto-CoC had been a structurally somewhat loose and iconic verb chain, the pre-CoC became a more compact package after the aforementioned reorganization of the case system. This is demonstrated formally in Figure 4.
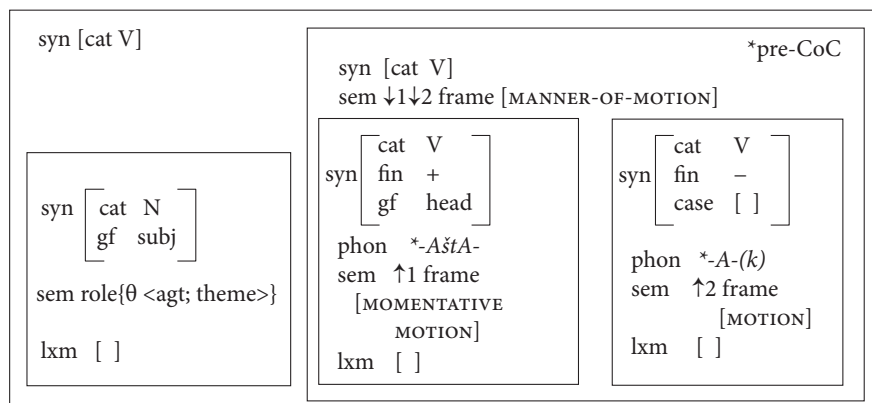

Figure 4. The pre-CoC as a new grammatical construction

At the level of actual instances or constructs, a structural difference between the proto-CoC and the pre-CoC is not noticeable at first, simply because the *k-lative's structural erosion was slower than its actual becoming a functionally unproductive case. In Figure 4, this situation is shown by putting the *k-lative within empty square brackets. Now the A-infinitive became a caseless verb form and syntactically unanalyzable (syn [case [ ]]) as well. As the reanalysis finally took place, semantic ambiguity was lost and the verbal chain was interpreted as a construction *per definitionem*, i.e. as a structural-functional Gestalt. This is indicated by the thicker black line in the larger box representing the pre-CoC as a newly conventionalized grammatical construction. As has been proposed, this change was gradual.

In sum, the grammatical development from the proto-CoC COL + INF_LAT to the pre-CoC COL + INF was the consequence of a general rearrangement in the Balto-Finnic case system. The emergence of the MA-infinitive illative and the fading of *k-lative led to a reanalysis; in other words, the previous meaning

---

16. In Estonian, the infinitive forms are similar to those in Finnish, i.e. A-, E-, and MA-infinitives, but the case distribution differs slightly. The illative cannot be attached to the MA-infinitive form even though in example (11c), *seisma* is functionally equivalent to Finnish *seisomaan* in the parallel example *hän ponnahtaa seisomaan* 's/he stands up quickly'. I would like to express my gratitude to Peeter Teenets for having clarified some of the Estonian infinitive usages.

CHANGE-OF-STATE was gradually re-interpreted as MANNER-OF-MOTION. The unproductive *k-lative faded phonetically and became the morphophonemic x-feature of the A-infinitive (see footnote 9), which is not realized at all in written standard Finnish. Thereafter, in order to develop into the current CoC, the pre-CoC underwent semantic extensions by analogy and by the change in word order.

4.2.3   Analogy and semantic extensions

According to my data, MOTION is a prototypical frame for the CoC but there are other frames as well (4b–c, 5b, 6a, 8b, 9a). The current CoC can be applied to any basic and commonplace actions such as eating, drinking, sleeping, sitting, standing, speaking, and so on and even to non-animate actions. The essential question is how these semantic frames reached the pre-CoC. Besides structural reanalysis, another crucial force of grammaticalization – *analogy* – played an active role in the further development of the pre-CoC (cf. Hopper & Traugott 2003: Chapter 3.5). A historical equivalent to the synchronic prototype and the source of the analogy can be called a host class (Himmelmann 2004: 32). This is schematically represented in Figure 4 as the pre-CoC designated manner of motion. It was a constructional mold for similar verbal chains in semantic fields other than motion.

Next, I will argue that the gradual development occurred as follows: as described above, the proto-CoC preferred momentative (and inchoative) colorative verbs, which focused on the culmination of two separate stages (Figure 2). After the *k-lative became unproductive, the semantic relationship between the colorative verb and the infinitive became more compact and the semantics of this syntagm gradually began to designate MANNER-OF-MOTION (Figure 4). The outcome of the reanalysis enabled verbs other than momentative – inchoative motion verbs to act as colorative verbs in the pre-CoC, as shown in (12a–b).

(12)   *sija-t   ne käyv-vä **jöynättel**   uppokesät tuola*   (LA)
        pig-NOM.PL they walk-INF COL.PST.3SG summers there
        'pigs wandered-COL there during summers'
        *hevose-t   käy-rä **lont-i**   oma-a   aika-a-nsa*   (LA)
        horse-NOM.PL walk-INF COL-PST.3SG own-PAR time-PAR-PX.
        'horses slouched and took their time when they moved'

This was the first analogical extension of the pre-CoC to motion verbs in general and it strengthened the MOTION frame as characteristic and prototypical of the CoC. It also established the MOTION frame as a host class for further analogical expansions. If we consider Figure 4 again, it can be seen that the colorative verbs' phonological value (phon *-AštA-) and semantic frame (MOMENTATIVE MOTION) were no longer valid.

The second semantic extension by analogy was the development that also enabled frames other than pure motion to be applied in the pre-CoC. In this respect, there are two competing explanations. Saukkonen (1965: 58) argued that, as a productive case, the *k-lative had two original meanings: concrete spatial direction (*lokaali* 'local') and abstract direction or purpose (*finaali* 'final') 'in order to do something' (e.g. *syödäk 'in order to eat'). According to Saukkonen (1965: 239), these functions could be equally old. Traces of the 'final' (i.e., PURPOSIVE) function can be found in older written Finnish as well as in current Finnish. Example (13) represents a Finnish permissive construction, which is supposed to have had ambiguous semantics (Leino 2005: 106–112).

(13) *Kalle anta-a   omena-n Ville-n   syö-dä*   (made-up)
      Kalle give-PRS.3SG apple-GEN Ville-GEN/DAT eat-INF
      'Kalle gives an apple to Ville to eat'   (PURPOSIVE) ~
      'Kalle lets Ville eat an apple'   (PERMISSIVE)

Now, if we suppose that (a) examples such as (14a–b) existed between the proto- and pre-CoC era along with the COL + INF type motion variants and (b) that the *k-lative was still somewhat productive, we have to infer that the A-infinitive's *k-lative case could not reasonably be interpreted as local, but either as purposive or as the newly established pre-CoC. The first interpretation would require the PURPOSIVE *k-lative to have gone through a similar process of gradual reanalysis as spatial proto-CoCs, while the second interpretation would require further innovations to have been made by analogy with the pre-CoC (Figure 4). I prefer the latter explanation.

(14) a. *sepä-t   nalakuttel-j takko-on   ne*   (LA)
          smith-NOM.PL COL-PST.3SG hammer-INF them
          ?'blacksmiths used.to.COL in order to hammer them'   (PURPOSIVE) ~
          'blacksmiths used to hammer-COL them'   (CoC interpretation)
      b. *si-llä   napsauttan-neet siihel lyy-äs   sitten*   (LA)
          it-ADE COL-PCP.PL   there hit-INF then
          ?'they had.COL-ed there in order to hit it [with an axe]'   (PURPOSIVE) ~
          'they had hit-COL there with it [an axe]'   (CoC interpretation)

There are several suspicious issues that lead one to suppose that probably the PURPOSIVE meaning of the *k-lative did not have a role in the grammaticalization of the CoC. First, as Saukkonen (1966: 139) put it, it is not always clear whether the *k-lative should be interpreted as local or as purposive. This is not a remarkable problem as soon as a usage situation and contextual features disambiguate interpretations effectively. But, second, if the local meaning of the *k-lative was not available, as in examples (14a–b), this does not guarantee that the purposive

interpretation would not be more or less unnatural. Saukkonen's (1966: 136) examples, such as *älä pölpötä puhua* 'do not prattle in order to talk' and *hän taapertaa mennä* 's/he toddles in order to go', in my opinion, represent the kinds of forced interpretations (preceded by a question mark in translations) similar to those in examples (14a–b).

Third and most important, the purposive meaning of the *k*-lative requires an extra meaning, namely, intentionality. The original spatial (local) meaning of the proto-CoC designated the MOTION frame, so that the colorative verb represented the inchoative aspect of the motion and the A-infinitive represented the actual motion type in question (10a–c). The function of the *k*-lative was to explicate the transition from one state to another, which was represented by the A-infinitive verb. Both in the spatial proto-CoC and in the spatial pre-CoC, the colorative motion and the motion designated by the A-infinitive were contained in the MOTION frame in general. The main result of the reanalysis, then, was the fading of the inchoative aspect and the fact that the colorative motion came to be contained in the MOTION frame of the A-infinitive in particular (cf. Figure 3 and Figure 4). This kind of containment in the same general frame is lacking in the intentional (purposive) meaning of the *k*-lative. Also the semantic hyponymy relation of the colorative and the A-infinitive does not favor the *k*-lative's purposive function. Usually in intentional contexts, one action in one frame is conducted in order to develop chances for some other action to occur in the other frame, as in (13). These are the main arguments for why interpreting examples such as (14a–b) as PURPOSIVE seems unnatural.

For these reasons, I do not follow Saukkonen (1966: 134–141), who maintained that the purposive meaning of the *k*-lative must have had a central role in the grammaticalization of the CoC. Instead, I prefer the idea that as soon as the pre-CoC emerged, it extended analogically to frames other than MOTION as well, and prior to this model, non-motion CoCs did not appear. This does not mean that the purposive function of the *k*-lative did not exist at all but, rather, that it was more apparent in the other verb syntagms than in the proto-CoC (13). Moreover, examples such as (15), where the verb syntagms could possibly be analyzed in all three ways, depending on the context between the proto- and pre-CoC periods, were surely marginal.

(15) *hän   rigenda-p   rat-a*   (Vepsian; Saukkonen 1966: 138)
      s/he.NOM COL-PRS.3SG work-INF
      's/he works quickly'   (CoC) ~
      's/he rushes [to a certain place] to work'   (local) ~
      's/he rushes [as an on-going action] in order to work'   (purposive)

In general, the MA-infinitive illative (e.g., *tulemaan*) replaced the local function of the *k*-lative earlier than the A-infinitive translative (e.g., *tullakseen*) replaced the final function (see Table 5 in footnote 14). So in relative chronology, the A-infinitive translative is younger than the MA-infinitive illative as a morphological form (see Itkonen 1983: 363).

4.2.4   *Crystallization of the CoC and a note on syntax → pragmatics*

Figure 5 shows the end result of the grammaticalization process, namely, the current CoC. As described in the preceding sections, phases in the grammaticalization cycle syntax → morphology → morphophonemics (→ zero) represent mainly major structural changes in the Finnish infinitive and case system in general and the reanalysis process of the CoC in particular. The most important change has been the *k*-lative's gradual functional and structural fading (syn [case [ ]]), which led to the semantic reorientation of the relationship between the A-infinitive and the colorative verb. This can be seen in the thick black line of the box, which represents the established constructional status of this verb syntagm. The other two structural changes are the extended word order variation from COL + INF to INF + COL and the fact that there cannot be anything between the verbs in INF + COL variant; e.g. **mennä niin rauhallisesti körötteli*, but *körötteli niin rauhallisesti mennä* 'COL-ed so calmly (to go)'.

The later phase that the grammaticalization process went through after the reanalysis was extension by analogy. This has had its effects on the structural organization of the CoC by relaxing semantic constraints on both verb frames. The A-infinitive can basically be any verb whatsoever, not just a motion verb, and the colorative verb does not have to be only momentative. In fact, in my data there are even a few examples in which the colorative part of the CoC is not actually colorative in its lexical semantics, as in example (16).

(16)   *navetta.tie-tä   pitkin tul-lal   loikk-i   susi*   (DMA)
       cowhouse.path-PAR along come-INF leap-PRT.3SG wolf
       'along the path to/from the cowhouse a wolf came leaping'

The verb *loikkia* 'to leap' in (16) is not particularly colorative but it occupies the same structural position as true colorative verbs in the CoC. The infinitive *tulla* 'to come' designates motion in general and deictic direction, while *loikkia* has a manner of motion meaning in its lexical semantics. In the CoC, a more or less neutral finite motion verb becomes 'colored' as the whole construction emphasizes a manner of the action.
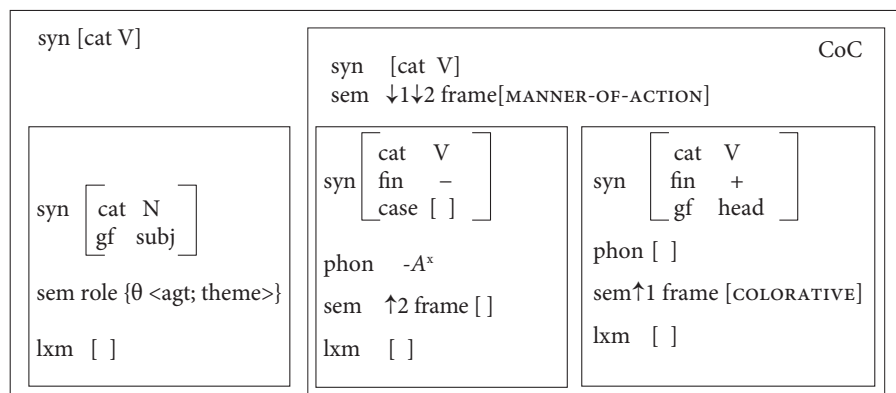
syn [cat V]   CoC

    syn   [cat V]
    sem   ↓1↓2 frame[MANNER-OF-ACTION]

    syn [cat   N
          gf   subj]   syn [cat   V   syn [cat   V
                                       fin   –   fin   +
    sem role {θ <agt; theme>}   case [ ]]   gf   head]

    lxm [ ]   phon   -Aˣ   phon [ ]

                                 sem   ↑2 frame [ ] sem ↑1 frame [COLORATIVE]

                                 lxm   [ ]   lxm   [ ]

Figure 5. Colorative Construction

The semantic focusing on the manner of the action and the structural orienting toward a tight INF + COL word order variant (see Table 3) led to a crystallized form of the CoC as described in Figure 5. Even though the description of the grammaticalization of the CoC in this paper has not been usage-based in the sense of involving text linguistics methods and statistical analysis, it does offer some corollaries at the usage (i.e. pragmatics) level. First, a clear semantic difference between the word order variants is hard to express but in the INF + COL variant, the neutral A-infinitive verb categorizes the action in general, which is then modified semantically by the colorative verb. Thus, the A-infinitive names the explanatory frame for the descriptive colorative verb to modify. Precisely in that sense, the INF + COL order is pragmatically motivated: the A-infinitive offers an interpretive background for the colorative verb. Second, the INF + COL order makes a structural distinction between the CoC and the other infinitive expressions in Finnish and foregrounds its special character in that respect. Typically, in Finnish word order, the modifiers come after the finite verb. Third, the semantics of the CoC (MANNER-OF-ACTION) is more general than that of the proto-CoC (CHANGE-OF-STATE) or the pre-CoC (MANNER-OF-MOTION). In that respect, the CoC is pragmatically more usable than its historical predecessors, i.e., it can cover a broader range of states-of-affairs in the world than the older versions of the construction.

5.   Summary and discussion

In this paper, I have found it fruitful to take reconstructed proto-constructions as a starting point in trying to outline the gradual and episodic grammaticalization cycle of the Finnish CoC. Figure 6 gives a short recapitulation of the incremental process of the diachronic micro-steps (Traugott & Trousdale 2010: 20). As

a grammatical construction, the CoC is lexeme-group specific, that is, it builds almost exclusively on onomatopoetic and descriptive colorative verbs in Finnish, and the A-infinitive serves as a sort of nametag explaining the overall action. The following summary is based directly on Figure 6.

First, the proto-CoC was a source structure for grammaticalization. In the Middle Proto-Finnic era, phrases such as **koira pyrästi joostak* 'a dog abruptly started to run' and similar analogs in (10) described a sudden CHANGE-OF-STATE. The proto-CoC was quite a lenient verbal chain for two reasons: syntactically, there could be other words between the colorative and the A-infinitive and semantically, the verbs designated distinct events. An iconic word order COL + INF$_{\text{LAT}}$ sustained a sequential interpretation. Pragmatically, it was usable only for limited events, such as when someone or something was observed to begin moving abruptly.
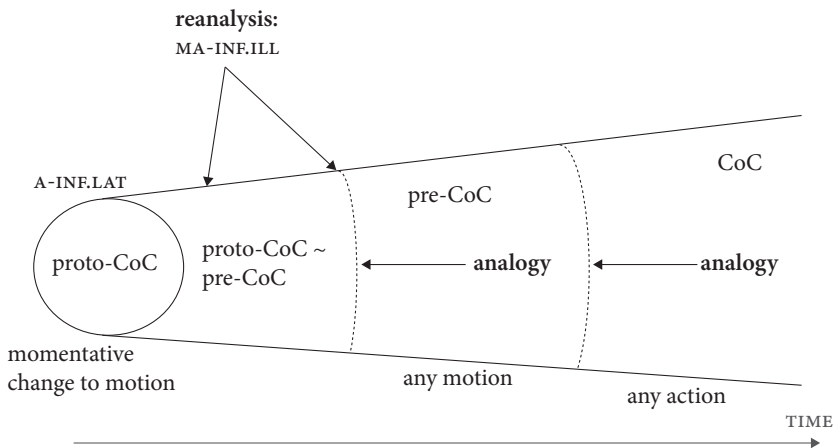

Figure 6. Summary for the grammaticalization of the CoC

Second, the point at which grammaticalization sets in goes back to the emerging co-existence of the **k*-lative and the MA-infinitive illative, respectively. The latter acted as an 'external pushing force', gradually taking over the spatial function of the **k*-lative, leading to the reanalysis of syntagmatic characteristics of the proto-CoC. The reanalysis itself, having structural surface markers or changes at the beginning (cf. Traugott & Trousdale 2010: 33), led to an ambivalent situation in which both CHANGE-OF-STATE (the proto-CoC) and MANNER-OF-MOTION (the pre-CoC) interpretations were possible. Supposedly, these were pragmatically induced (Bybee 2010: 199). The pre-CoC reached a MANNER interpretation due to the reanalysis; in effect, at this point, a new paradigmatic unit emerged. This essentially meant that the pre-CoC was compressed to semantically simultaneous verbal actions, as both verbs in the construction described the very same frame, i.e. a type of motion (with the A-infinitive) and a manner of motion (with the colorative verb).

Third, the MANNER-OF-MOTION interpretation enabled the colorative verb to be any motion verb. This set off a strong host class, which then acted as an analogical source for all the other action frames. The analogy likewise acted as an 'internal pulling force', which absorbed new expressions from novel semantic domains. After the pre-CoC was established, analogy worked as a general mechanism to extend the usage of the CoC to non-motion frames as well. This can be seen as paradigmatic change. I find it likely that the purposive meaning of the *k-lative did not have any role in the grammaticalization of the CoC. Regardless of the precise periods of analogical expansion, the result was a fresh grammatical construction whose semantic character was that of MANNER-OF-ACTION. Structurally, it consisted of a colorative finite verb with an infinitive. A strong tendency in favor of the fixed word order of INF + COL, particularly in Finnish, pinpointed its distinction from other infinitival expressions and motivated the linearity in a new way: an infinitive designated an overall frame and the colorative verb elaborated on it by depicting some special 'colorative' aspects of the manner of the action.

There are two crucial aspects to point out from the perspective of grammaticalization. First, the process as a whole was gradual in the sense of being a series of micro-changes (Traugott & Trousdale 2010: 23–26). Within the constructional framework, this was demonstrated by single-feature changes (see Figures 2–5). The gradualness means that the grammaticalization process was non-reducible and inseparable from the co-effects of both reanalysis and analogy (Traugott & Trousdale 2010: 38–39). Second, the genesis of the CoC is a paradigm example of constructionalization; in other words, both the form and the meaning of a construction changed into a different form and meaning. The gradual change can logically be formalized as follows (where f = form and m = meaning): $f_1/m_1 > f_1 \sim f_2/m_1 \sim m_2 > f_2/m_2 > f_n/m_n$. The reanalysis from the proto-CoC to the pre-CoC went through an ambiguous phase in which a gradual structural change led to a meaning change ($f_1/m_1 \gg f_2/m_2$) and then analogy spread the new expression to non-motion frames as well ($> f_n/m_n$). The relative logical order of changes was reanalysis and then analogy, but the actual constructionalization process kept these tightly in tandem.

This paper thus offers a detailed description of the constructionalization of the Finnish CoC. With the evidence from dialectal data and other closely related Balto-Finnic languages combined with CxG formalization, it is possible to give a plausible view of the gradual historical process involving reanalysis, semantic ambiguity, and analogy. This strengthens the pivotal idea of constructionalization as one type of grammatical change in general. However, it remains for future empirical research to solve such problems as the possible cross-constructionality and micro-constructional variation between Balto-Finnic languages and perhaps other genetically related Finno-Ugric languages more generally. Moreover, potential

universal aspects of ideophonic expressions await a thorough typological study. The lexicalization of ideophones, co-verbs, converbs, and serial verb constructions offers an important line of inquiry for these kinds of studies.

Abbreviations and symbols

ABE	abessive	PASS	passive
ADE	adessive	PAST	past tense (preterite)
ADVL	adverbial	PCP	participle
ALL	allative	PF	Proto-Finnic
cat	category	phon	phonological (information in a construction)
CoC	colorative construction	PL	plural
COL	colorative verb	PRS	present tense
COND	conditional	PTV	partitive case
CONV	converb	PX	possessive suffix
DAT	dative	SG	singular
ELA	elative	sem	semantic (information in a construction)
FIN	finiteness, finite verb	subj	subject
GEN	genitive	syn	syntactic (information in a construction)
gf	grammatical function	TAM	tempus + aspect + modus (markers)
ILL	illative	TRA	translative
INE	inessive	θ-role	semantic role (agent, patient, theme, etc.)
INF	(Balto-Finnic, Finnish) (A-)infinitive	1, 2, 3	index numbering
INS	instructive	↓↑	source and goal of semantic information (indicating semantic integrations in and between frames)
LAT	lative	*	ungrammatical or reconstructed form/structure
lxm	lexical material in a construction	+	a positive feature in an attribute – value matrix, e.g., syn [fin +] = syntactically a finite verb form
MAINF	(Balto-Finnic, Finnish) MA-infinitive	−	a negative feature in an attribute – value matrix, e.g., syn [fin −] = syntactically a non-finite verb form
NOM	nominative	#	morpheme boundary
PAR	partitive		

Data archives

DMA = *Digitaalinen muoto-opin arkisto* (Digital Morphology Archive). Collected by the Morphology Archive of the Department of Finnish, Finno-Ugrian, and Scandinavian Studies at the University of Helsinki and the CSC – IT Center for Science. Accessed 12 June 2016. Available at https://korp.csc.fi

KKSK = *Karjalan kielen sanakirja* (Dictionary of Karelian). Online database for the list of lexical entries. Helsinki: Institute for the Languages of Finland. Accessed 12 June 2016. Updated periodically. Available at http://kaino.kotus.fi/sanat/kkss/

LA = *Lauseopin arkisto* (Syntax Archives). Institute for the Languages of Finland and the School of Language and Translation Studies at the University of Turku. Turku: Syntax Archives. Accessed 12 June 2016. Available at https://korp.csc.fi

SMS = *Suomen murteiden sanakirja* (Dictionary of Finnish Dialects), vols. 1–8. Helsinki: Institute for the Languages of Finland. Accessed 12 Jun 2016. Available at http://kaino.kotus.fi/sms/

References

Anttila, R. (1973). Internal reconstruction and Finno-Ugric (Finnish). In Th. A. Sebeok (Ed.), *Current trends in linguistics*, Vol. 11: *Diachronic, areal and typological linguistics* (pp. 317–353). The Hague: Mouton.

Anttila, R. (1977). Toward a semiotic analysis of expressive vocabulary. *Semiosis 5, Internationale Zeitschrift für Semiotik und ihre Anwendungen, Heft* 1, 27–40.

Barlow, M., & Kemmer, S. (Eds.). (2000). *Usage-based models of language*. CSLI Publications. Stanford, California: Center for the Study of Language and Information.

Bergs, A., & Diewald, G. (Eds.). (2008). *Constructions and language change* [Trends in Linguistics. Studies and Monographs 194]. Berlin: Mouton de Gruyter. doi: 10.1515/9783110211757

Bergsland, K. (1994). *Sydsamisk grammatikk* [South Sami Grammar]. Second edition. Karasjok: Davvi Girji.

Bybee, J. (2010). *Language, usage and cognition*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511750526

Diewald, G. (2007). Konstruktionen in der diachronen Sprachwissenschaft. In K. Fischer & A. Stefanowitsch (Eds.), *Konstruktionsgrammatik: Von der Anwendung zur Theorie* (pp. 79–103). Updated 2nd edition, Stauffenburg Linguistik Band 40. Germany: Stauffenburg Verlag.

Erhard Voeltz, F. K., & Kilian-Hatz, C. (2001). Introduction. In F. K. Erhard Voeltz & C. Kilian-Hatz (Eds.), *Ideophones* [Typological Studies in Language 44] (pp. 1–8). Amsterdam & Philadelphia: John Benjamins Publishing Company. doi: 10.1075/tsl.44.02voe

Fillmore, C., & Kay, P. (1995). *Construction Grammar*. CSLI Lecture Notes. Stanford, California: Center for the Study of Language and Information.

Fried, M., & Boas, H. (Eds.). (2006). *Grammatical constructions: back to the roots* [Constructional Approaches to Language 4]. Amsterdam & Philadelphia: John Benjamins Publishing Company.

Fried, M. (2009). Construction Grammar as a tool for diachronic analysis. *Constructions and Frames*, 1(2), 262–291. doi: 10.1075/cf.1.2.04fri

Fried, M., & Östman, J. -O. (2004). Construction Grammar. A thumbnail sketch. In M. Fried & J. -O. Östman (Eds.), *Construction Grammar in a cross-language perspective* [Constructional Approaches to Language 2] (pp. 11–86). Amsterdam: John Benjamins Publishing Company. doi: 10.1075/cal.2.02fri

Givón, T. (1979). *On understanding grammar*. New York: Academic Press Inc.

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago & London: The University of Chicago Press.

Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

Haiman, J. (1980). The iconicity of grammar: Isomorphism and motivation. *Language*, 56(3), 515–540. doi: 10.2307/414448

Hakulinen, L. (1961). *The structure and development of the Finnish language* [Indiana University publications, Uralic and Altaic Series, Vol. 3]. Translated by J. Atkinson. Bloomington, IN: Indiana University Press.

Haspelmath, M. (2010). Comparative concepts and descriptive categories in crosslinguistic studies. *Language*, 86(3), 663–687. doi: 10.1353/lan.2010.0021

Heikkilä, M. K. (2014). *Bidrag till Fennoskandiens språkliga förhistoria i tid och rum* [dissertation]. Helsinki: Helsingfors universitet.

Heikkinen, V., & Voutilainen, E. (2009). Koloratiivirakenne *Hirventappopaikka*-romaanissa [The colorative construction in the novel *Hirventappopaikka*]. In V. Heikkinen (Ed.), *Kielen piirteet ja tekstilajit: Vaikuttavia valintoja tekstistä toiseen* [Features and genres of the language: Expressive choices from a text to another] (pp. 135–169). Tietolipas 229. Helsinki: Suomalaisen Kirjallisuuden Seura.

Heine, B., Claudi, U., & Hünnemeyer, F. (1991). *Grammaticalization: A conceptual framework*. Chicago: The University of Chicago Press.

Herlin, I., Leino, P., & Visapää, L. (2005). Kas siinä pulma [That is the question]. In I. Herlin & L. Visapää (Eds.), *Elävä kielioppi: Suomen infiniittisten rakenteiden dynamiikkaa* [Living grammar: The dynamic nature of Finnish infinitival structures] (pp. 9–38). Suomalaisen Kirjallisuuden Seuran Toimituksia 1021. Helsinki: Suomalaisen Kirjallisuuden Seura.

Himmelmann, N. P. (2004). Lexicalization and grammaticalization: Opposite or orthogonal? In W. Bisang, N. P. Himmelmann, & B. Wiemer (Eds.), *What makes grammaticalization: A look from its fringes and its components* [Trends in linguistics. Studies and monographs 158] (pp. 21–42). Berlin: Mouton de Gruyter.

Hopper, P. J., & Traugott, E. Closs (2003). *Grammaticalization*. Cambridge Textbooks in Linguistics. Second edition. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139165525

Ikola, O., Palomäki, U., & Koitto, A.-K. (1989). *Suomen murteiden lauseoppia ja tekstikielioppia* [Syntax and text linguistics of Finnish dialects]. Suomalaisen Kirjallisuuden Seuran Toimituksia 511. Helsinki: Suomalaisen Kirjallisuuden Seura.

ISK = Hakulinen, A., Vilkuna, M., Korhonen, R., Koivisto, V., Heinonen, T. R. & Alho, I. (2004). *Iso suomen kielioppi* [The extensive Finnish grammar]. Helsinki: Suomalaisen Kirjallisuuden Seura. Available at http://scripta.kotus.fi/visk [Accessed 18 Jan. 2016].

Itkonen, T. (1964). *Proto-Finnic final consonants: Their history in the Finnic languages with particular reference to the Finnish dialects I: 1. Introduction: The history of -k in Finnish*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Itkonen, T. (1983). Välikatsaus suomen kielen juuriin [Retrospective view to the roots of Finnish]. *Virittäjä*, 87, 349–386.

Jarva, V. & Kytölä, S. (2007). The Finnish colorative construction and expressivity. *SKY Journal of Linguistics*, 20, 235–272.

Jomppanen, M. (2011). Pohjoissaamen ja suomen kielen koloratiivikonstruktio vertailussa [Comparison of colorative constructions in Finnish and North Saami]. In A. Kaivapalu, J. Laakso, P. Muikku-Werner, & M.-M. Sepper (Eds.), *Lähivõrdlusi/Lähivertailuja* 21 (pp. 106–131). Tallinn: Eesti Rakenduslingvistika Ühing. Available at http://www.rakenduslingvistika.ee/ul/files/LV2105Jomppanen.pdf [Accessed 28 Sep. 2016].

Karelson, R. (1958). Soome-ugri keelte lausestruktuurist seoses konjunktsioonidega [Finno-Ugric clause structure with conjunctions]. *Eesti Seltsi Aastaraamat* IV, 195–213.

Karlsson, F. (1999). *Finnish: An essential grammar*. New York: Taylor & Francis e-Library.

Kettunen, L. (1943). *Vepsän murteiden lauseopillinen tutkimus* [Syntactic research on Veps dialect]. Suomalais-Ugrilaisen Seuran Toimituksia LXXXVI. Helsinki: Suomalais-Ugrilainen Seura.

Kim, J.-B. (2012). The English binominal NP construction: a construction grammar approach. Plenary presentation at *7th International Conference on Construction Grammar* in Seoul, Korea.

Korhonen, M. (1981). *Johdatus lapin kielen historiaan* [An introduction to the history of Saami language]. Suomalaisen Kirjallisuuden Seuran Toimituksia 370. Helsinki: Suomalais-Ugrilainen Seura.

Leino, J. (2005). Frames, profiles and constructions: Two collaborating CGs meet the Finnish permissive construction. In J.-O. Östman & M. Fried (Eds.), *Construction grammars: Cognitive grounding and theoretical extensions* [Constructional Approaches to Language 3] (pp. 89–120). Amsterdam & Philadelphia: John Benjamins Publishing Company. doi: 10.1075/cal.3.05lei

Lehmann, C. ([1982] 2002). *Thoughts on grammaticalization*. Arbeitspapiere des Seminars für Sprachwissenschaft der Universität Erfurt, Nr. 9. Second, revised edition. Erfurt: Seminar für Sprachwissenschaft, Philosophische Fakultät.

Lehtinen, T. (2007). *Kielen vuosituhannet. Suomen kielen kehitys kantauralista varhaissuomeen* [Millenia of the language. Evolution of Finnish language from Proto-Uralic to early literal Finnish]. Tietolipas 215. Helsinki: Suomalaisen Kirjallisuuden Seura.

Mihkla, K. (1964). Öeldis [Predicate]. In K. Mihkla (Ed.), *Eesti keele süntaks I: prooviartikleid lihtlause süntaksi alalt* [Syntax of Estonian I: research articles on the syntax of simple sentences] (pp. 66–101). Tallinn: Eesti NSV Teaduste Akadeemia Keele ja Kirjanduse Instituut.

Mikone, E. (2001). Ideophones in the Balto-Finnic languages. In F. K. Erhard Voeltz & C. Kilian-Hatz (Eds.), *Ideophones* [Typological Studies in Language 44.] (pp. 224–233). Amsterdam & Philadelphia: John Benjamins Publishing Company. doi: 10.1075/tsl.44.18mik

Nickel, K. P. (1994). *Samisk grammatikk* [Sami Grammar]. Second, revised edition. Karasjok: Davvi Girji O. S.

Nielsen, K. ([1926] 1979). *Laerebok i lappisk (samisk), I Grammatikk* [Schoolbook on Sami, I Grammar]. Second edition. Oslo: Universitetsforlaget.

Noël, D. (2007). Diachronic construction grammar and grammaticalization theory. *Functions of Language*, 14 (2), 177–202. doi: 10.1075/fol.14.2.04noe

Posti, L. (1980). *Vatjan kielen Kukkosin murteen sanakirja* [The dictionary of Votic dialect of Kukkos]. Lexica Societatis Fenno-Ugricae XIX. Helsinki: Suomalais-Ugrilainen Seura.

Põlma, V. (1967). *Onomapoeetilised verbid eesti kirjakeeles* [Onomatopoetic verbs in Estonian literary language]. Tallinn: Eesti NSV Teaduste Akadeemia Keele ja Kirjanduse Instituut.

Ringe, D. (2008). Internal reconstruction. In B. D. Joseph & R. D. Janda (Eds.), *The handbook of historical linguistics* (pp. 244–261). Third Edition. Oxford: Blackwell Publishing.

Rostila, J. (2004). Lexicalization as a way to grammaticalization. In F. Karlsson (Ed.), *Proceedings of the 20th Scandinavian Conference of Linguistics*. Available at http://www.ling.helsinki.fi/kielitiede/20scl/Rostila.pdf [Accessed 14 Oct. 2016].

Rostila, J. (2006). Storage as a way to grammaticalization. *Constructions* 1/2006. Available at www.constructions-online.de [Accessed 17 Apr. 2016].

Rytkönen, A. (1937). Koloratiivinen konstruktio [The colorative construction]. *Virittäjä*, 41, 95–104.

Saukkonen, P. (1965). *Itämerensuomalaisten kielten tulosijainfinitiivirakenteiden historiaa I: johdanto: adverbiaali infinitiivi* [The history of Balto-Finnic lative infinitival constructions I: Introduction: Adverbial infinitive]. Suomalais-Ugrilaisen Seuran Toimituksia 137. Helsinki: Suomalais-Ugrilainen Seura.

Saukkonen, P. (1966). *Itämerensuomalaisten kielten tulosijainfinitiivirakenteiden historiaa II: adverbaali infinitiivi (täydennys), adnominaali infinitiivi, absoluutti infinitiivi, elliptinen infinitiivi* [The history of Balto-Finnic lative infinitival constructions II: Adverbial infinitive (supplement), adnominal infinitive, absolute infinitivi, elliptic infinitive]. Suomalais-Ugrilaisen Seuran Toimituksia 140. Helsinki: Suomalais-Ugrilainen Seura.

Setälä, E. N. (1880). *Suomen kielen lauseoppi: Oppikirjan koe* [Finnish syntax: A trial for a schoolbook]. Helsinki: K. E. Holm.

Traugott, E. Closs (2008a). Grammatikalisierung, emergente Konstruktionen und der Begriff der "Neuheit". In K. Fischer & A. Stefanowitsch (Eds.), *Konstruktionsgrammatik: Von der Konstruktion zur Grammatik* [Stauffenburg Linguistik Band 47] (pp. 5–32). Germany: Stauffenburg Verlag.

Traugott, E. Closs (2008b). Constructions in grammaticalization. In B. D. Joseph & R. D. Janda (Eds.), *The handbook of historical linguistics* (pp. 624–647). 3rd edition. Oxford: Blackwell Publishing.

Traugott, E. Closs, & Heine, B. (Eds.). (1991). *Approaches to grammaticalization. Volume 1. Focus on theoretical and methodological issues*. Amsterdam: John Benjamins Publishing Company. doi: 10.1075/tsl.19.1

Traugott, E. Closs, & Trousdale, G. (2014). *Constructionalization and constructional changes*. Oxford: Oxford University Press.

Traugott, E. Closs, & Trousdale, G. (2010). Gradience, gradualness and grammaticalization. How do they intersect? In E. Closs Traugott & G. Trousdale (Eds.), *Gradience, gradualness and grammaticalization* [Typological Studies in Language 90] (pp. 19–44). Amsterdam: John Benjamins Publishing Company. doi: 10.1075/tsl.90.04tra

Virtaranta, P. (1986). *Haljärven lyydiläismurteiden muoto-oppia* [Morphology of Ludic dialects in Haljärvi]. Suomalais-Ugrilaisen Seuran Toimituksia 190. Helsinki: Suomalais-Ugrilainen Seura.

Ylikoski, J. (2011). A survey of the origins of directional case suffixes in European Uralic. In S. Kittilä, K. Västi, & J. Ylikoski (Eds.), *Case, animacy and semantic roles* (pp. 235–280). Amsterdam & Philadelphia: John Benjamins Publishing Company. doi: 10.1075/tsl.99.09yli

Ylikoski, J. (2003). Defining non-finites: Action nominals, converbs and infinitives. *SKY Journal of Linguistics*, 16, 185–237.

Author's address

Markus Veli Juhani Hamunen
Department of Finnish, Finno-Ugrian and Scandinavian Studies
University of Helsinki
Fabianinkatu 33 00014
Finland

markus.hamunen@helsinki.fi