On Becoming an Independent User

Maisa Martin, Sanna Mustonen, Nina Reiman
and Marja Seilonen
University of Jyväskylä / University of Jyväskylä / University of
Jyväskylä / University of Jyväskylä, University of Eastern Finland

The chapter presents tentative results of the project *Linguistic Basis of the Common European Framework for L2 English and L2 Finnish (Cefling)* for three structures of Finnish: the use of local cases, and transitive and passive constructions. The data consist of 669 texts written by adult learners of Finnish as a second language, rated on a functional CEFR scale at levels A1 – C2. The chapter also presents the DEMfad Model used for the analysis and some tentative results which show that often the frequency of use of a structure increases significantly from level A2 to B1 while the leap in accuracy follows it, with the greatest growth between levels B1 and B2. In addition, some aspects of linguistic complexity and the use of constructions as opposed to rules as the starting point of the analysis are discussed.

1. Introduction

The focus of this book is on the relationships between the communicatively defined levels (as in the Common *European Framework of Reference for Languages*, *CEFR*, Council of Europe, 2001) and the linguistic domains the language users control when they have been assessed to be on one of these levels. The learner is assumed to progress through six stages: Basic User (Breakthrough & Waystage), Independent User (Threshold & Vantage), and Proficient User (Effective Operational Proficiency & Mastery). Two assumptions underlie this research: (1) Language proficiency can be described as progressive, and stages along the progression can be established. (2) The range and quality of linguistic items used at a given level, in comparable tasks, bear at least some similarity across learners in the sense that some growth patterns can be shown. The first assumption is taken for granted in this chapter (see Alanen, Huhta, & Tarnanen, this volume), the second one is under our scrutiny.

The research reported in this chapter is based on a project called *The linguistic basis of the Common European Framework levels: Combining second language acquisition and language testing research*, also known as Cefling

EUROSLA MONOGRAPHS SERIES 1
Communicative proficiency and linguistic development, 57-80

(https://www.jyu.fi/ cefling). Like the SLATE network, the Cefling project attempts to bring together the knowledge acquired in the areas of Second Language Acquisition (SLA) and Language Testing. Cefling focuses on writing only, although the methods developed for the analysis are equally applicable to speaking. The target languages of the project are English and Finnish but only the latter is discussed in this chapter. The studies reported here are piloting many of the problematic issues involved, and most of the results are thus only tentative.

There are two main aims for this article: to show how the development of certain linguistic structures can be followed across CEFR levels, and to find evidence for potential co-development in different domains, as well as for interesting diversions of the development from the linear progression. Three structures of Finnish are targeted here as examples of the development in Finnish and as different realizations of the DEMfad Model (below): The use of locative cases, and the transitive and passive constructions.

The emphasis on communication and the importance of a comprehensive view of language, underlying the CEFR, form a part of the conceptual background of the project. As to how language knowledge develops, the broad underlying framework of the project is a usage-based and cognitively oriented view of language learning: acquisition takes place by encountering a growing number of instances of the second language (L2) from which regularities are extracted by use of the general cognitive mechanisms. The domains to be studied are not defined as rules or items but as constructions. Constructions are here loosely defined as units of language which contain a form and a meaning, both of which can vary within some limits (e.g. Goldberg, 2003).

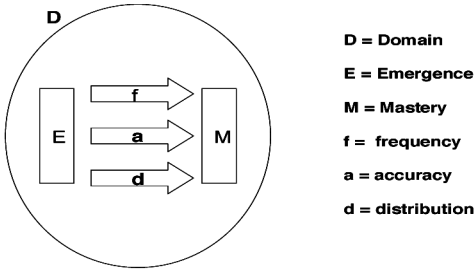
Construction Grammar (see e.g. Fillmore & Kay, 1996; Goldberg, 1995, 2003) has been previously employed to describe several structures of Finnish, e.g. infinitive constructions and the local and dative cases (see e.g. Leino, J., 2003, 2008; Leino P. et al., 2001; Visapää, 2008). However, the construction approach as the linguistic basis of studying second language development has not been previously used for Finnish, and only minimally for other languages (see e.g. Eskildsen, 2008).

1.1. The DEMfad Model

Language proficiency is commonly described as developing in three dimensions: complexity, accuracy, and fluency (the CAF triad). Many measures have been used to track the development of these dimensions (for an overview, see e.g. Housen & Kuiken, 2009; Wolfe-Quintero, Inagaki, & Kim, 1998). In this study, too, three dimensions are tracked, in some ways but not completely comparable to the CAF triad. For the analysis of our data the dimensions are combined in the DEMfad Model, which is intended to make the tracking of the

development comparable across the levels, domains and languages. The structure of the model is shown below.

Figure 1. The DEMfad Model (Franceschina, Alanen, Huhta, & Martin, 2006)



The **domain**s in this model are the areas of developing language skills, such as a construction or a set of constructions or the use of certain linguistic devices in service of semantic functions, a set of vocabulary etc. Thus the definition of a domain here is theory-independent (instead of a construction, a domain could equally well be defined as the application of a rule in the data, if the underlying linguistic framework were rule-based). Obviously some uniform guidelines are needed, particularly to avoid overlaps when the co-development of domains is examined. In this chapter the three domains (local case use, transitive constructions, passive constructions) are considered parallel but separate, with no consideration of potential co-development, so the overlap issue is not discussed here. It will need clarification in future theoretical work on the model as the number of empirical studies based on it grows.

The **emergence** is defined as the first occurrence of some indication of the presence of a domain, e.g. the use of the locative cases (Study 1) emerges when a noun is used with an ending interpretable as a locative suffix. The passive use (Study 3) can similarly be recognized by the presence of a morphological cue. The transitive construction (Study 2) has emerged if a noun or pronoun, verb, and another noun or pronoun follow each other in some order in some context where they can be interpreted to express subject, verb and object (Finnish is an SVO language), regardless of their formal properties.

The construction-based approach solves one of the problematic issues of acquisition criteria (see e.g. Pallotti, 2007): whether an item is actually acquired in a general sense or memorized as a chunk is not of interest in this approach. The first appearance of a recognizable construction, however chunk-like or however far from the target form, is the starting point of the development in the

domain. This is because in the construction-based view the memorized chunk is the basis of acquisition. The development is seen as gradual expansion: the variety of lexicon which can be used in the construction, and the semantic and formal variation within the construction, will grow.

Mastery in the DEMfad model is loosely defined as approximately idiomatic (target-like) use of the domain from the standpoint of frequency and distribution. For accuracy a tentative level of 80%[1] correctness has been established, but this can easily be varied if necessary in the future. For the other two parameters, frequency and distribution, no such mastery level has been set, as the control data from native speakers are being collected and assessed at the time of writing of this chapter, to provide future applications of the DEMfad Model a better point of comparison.

Frequency here is related to the concept of fluency. Fluency is notoriously difficult to define (see e.g. Wolfe-Quintero, Inagaki, & Kim, 1998), in writing as well as in speaking. As the writing tasks were performed in test situations and under time pressure, the number of words written can be seen as one indication of fluency, even if we are keenly aware that there are many other factors involved. Thus the number of words per text is used as an overall measure of fluency. Some of the other aspects of fluency, such as idiomaticity, are discussed as a part of the qualitative analysis of each domain. In a given domain, frequency of occurrence is calculated per 1000 words of running text. Even if the domains – be they constructions or sets of vocabulary – are not comparable in their likelihood of occurrence, and the task effects are great in this area, the patterns of changes in frequency give some indication of the development even across domains.

While the quantitative measures of fluency can be defined without reference to target language, **accuracy** does not exist without a target. The expressions of a given domain are compared to how a native speaker of the same age and education might formulate the same notion, i.e. the grounds of comparison do not always equal the written norm of the target language, particularly in the tasks requiring informal register. Most of the errors, however, are fairly easy to detect as ungrammatical inflections or usages, spelling errors, etc. Obviously, the native writer might commit many of the same errors when not paying attention. Again, the control data from similar groups of native writers will help in making both quantitative and qualitative comparisons between L1 and L2 users.

---

1 The 80% cut-off point is not based on any particular study but chosen simply to give some space for individual variation and random errors committed by both native and non-native language users. For the difficulty of drawing the line, see e.g. Abrahamsson & Hyltenstam (2009).

In the studies reported here no classification of errors is necessary, as the error types are not in focus, just the overall accuracy. In each domain errors are defined by what is required in the domain; for instance in a study of noun inflection the correct form of a case ending would be required for accuracy, while in the domain of the use of local cases (as in Study 2), accuracy is defined by the choice of the case, not by the spelling of the case ending, as long as it is recognizable.

Distribution is the parameter of the DEMfad Model most in need of further work. In the sub-studies of the moment several approaches are experimented with the purpose of finding potential aspects of distribution. The term *distribution* was chosen over the term *complexity* for several reasons. Most importantly, we were not satisfied with the measures of complexity commonly used in the studies of L2 development. The extent of subordination, for instance, does not seem a very useful measure in a language like Finnish, where – apart from relative clauses – subordinate clauses are not syntactically (e.g. word order) or morphologically (e.g. tempus or modus marking) different from main clauses. The L2 learner, who, unlike the L1 learner, already has the mental capacity required for subordination in general, only needs to acquire the necessary conjunctions. Furthermore, lengthy sentences with numerous co- and subordinations are not considered good style even in academic Finnish.

The underlying construction-based framework also sets new demands for tracking the development of linguistic complexity. A construction can grow by the number of different lexical items which can be used within it. It can also grow by the extent to which it can be semantically or syntactically varied, without taking on additional words or morphemes, which are commonly used to measure complexity. The use of a construction may also grow in complexity in a more traditional sense (i.e. in length) as constructions are combined in various ways, as when inserting a construction inside another one (e.g. a necessity construction in a transitive one: *hän ostaa auton* 'he buys a car' > *hänen täytyy ostaa auto* ´he must buy a car').

The term *distribution* at this stage must then be understood as a cover term for several types of phenomena which can be tracked in learner development. Some issues resemble *complexity* in some of the senses in which it has been used in previous SLA research. Some issues come close to *variability*. Many issues discussed under *distribution* do not easily render themselves to quantification. In this chapter we present examples of the types of issues which arise and some ways of dealing with them. In principle we are calling for a more multidimensional view of complexity, as does e.g. Norris and Ortega (2009). In developing the DEMfad Model we also attempt to differentiate and clarify important constructs of second language development, in a way responding to the plea of Pallotti (2009). In this chapter, however, the

problematic issues are brought up more by examples than by theoretical argument, which of course will also be necessary for the future development of the Model.

1.2. Finnish as L2

Unlike English, there is little previous information about the structural development of L2 Finnish. Individual MA level studies exist but as they are based on many different types of data and on diverse theoretical approaches, the results are not comparable between the studies. Nor is the proficiency of the learners rigorously determined; if learners at different levels are compared, the levels are defined by the courses the learners are taking or by the number of years of study etc. For this reason there are no established acquisition orders to be used as a starting point.

It is not possible to know a priori which structures might turn out to yield interesting information about the development of structures across the communicative CEFR levels, so the choices of structures have been based partly on their importance and frequency in written Finnish and partly on where researchers expect to see development based on their teaching- and assessment-based experience. In addition to the domains discussed in this chapter we have the results of some MA theses on the verb *to be* (Kynsijärvi, 2007), on some infinitive constructions (Paavola, 2008), on noun phrases (Ukkola, 2009), and on negation (Martin, 2008). Several other domains are being studied at the moment.

Below the data and methods are described, followed by the sections on the three domains chosen for this chapter. In each section we first present the main characteristics of the domain in question: what is there to be learnt? Then the results of each domain are presented. The overall results are discussed in the last section.

2. Data and methods

The Cefling Project as a whole uses two sets of data for each language: Writing samples from adults taking the National Proficiency Certificate exams and from young learners (12–16, grades 7–9), collected specifically for this project. Similar tasks are used in all sets of data. There are more than 20 different first languages (L1) among the learners of Finnish as a second language (L2). All participants live in Finland and the young learners go to school with Finnish as the language of instruction. Each writing sample has been independently rated to be at a given CEFR level (for the detailed discussion of task design, piloting, and

rating procedures see Alanen et al., this volume). The CEFR scale used for this purpose is purely functional, i.e. the communicative proficiency level has been assessed without attention to the adherence to linguistic norms or the capacity to use certain structures or vocabulary.

The writing tasks for both adults and young learners include three types of texts: an informal message (to a friend etc.), a formal message (a complaint or request to some institution), and an argumentative text (expressing an opinion). In addition, the young learners have written a narrative text.

The distribution of the data across the CEFR levels and the word counts are presented in Table 1. The data set used here includes only the scripts with high inter-rater reliability (see Alanen et al., this volume, for details). In the studies presented here, only the adult data are used.

Table 1. The total number of texts (all tasks) and words in the adult data for L2 Finnish

CEFR Level	Words	Texts	Words/text
A1	4 974	113	44,0
A2	5 702	103	55,4
B1	10 861	126	86,2
B2	9 080	108	84,1
C1	11 550	117	98,7
C2	10 852	102	106,4
Total/Average	**53 019**	**669**	**79,3**

The figures for individual tasks or genres are not given here as only the total results are discussed below. We have aimed at a roughly equal number of texts for each task/genre. The task effects vary between the domains, e.g. passive constructions are more frequent in argumentative texts, as one might expect. Apart from calculating frequencies and percentages, no statistical analyses have been conducted; the amount of data is given here simply to provide background for the reader. The fairly short average length of the texts, about 80 words, is influenced by the tasks: two of the three tasks presented to the adults were messages, in which case a lengthy test performance was not necessarily an advantage. The argumentative texts thus account for more of the growth of the number of words across the levels than do the other texts.

3. Study 1 - The development of the local case phrases in L2 Finnish: from concrete to abstract use

3.1. The Finnish local cases

There are fifteen cases in the Finnish language, eight of which form a subsystem called local cases. Alternative ways of categorizing the cases exist, but this is the approach taken in this study. Functionally, the local case system works like prepositions in Indo-European languages: it is based on the oppositions of directionality and quality. Directionality indicates the difference between expressions of TO (in Finnish the illative, allative, and translative cases), IN/ON/AT (inessive, adessive, and essive cases), and FROM (elative and ablative cases) while quality refers to the nature of the relationships expressed: internal (in English roughly in, into, from/out of X), external (on, onto, off/from X) or general (being/becoming X). Thus either static *being/existence* (on/in/at) or dynamic *direction of the movement* (from on/in/at; to on/in/at) is expressed by one of the cases. (For the classification and the terminology, see Huumo & Ojutkangas, 2006; Jackendoff, 1983.)

In general, and as far as the form, meaning and function of the local cases are concerned, the case system can be understood through locality or spatiality. The spatial domain (in a different meaning here from the DEMfad Model) is the primary one, and the other domains – e.g. action, circumstances, internal states, roles, time and possession – are analogical to the spatial one. According to a number of cognitive theories (see Johnson, 1987; Lakoff, 1987; Lakoff & Johnson, 1980), these expressions are considered to be metaphorical extensions of the spatial relationships and hence more abstract.

The semantics of the local case phrases (static vs. dynamic; direction of the movement; quality of a place) is not particularly transparent to the learner. The system is not thoroughly logical or watertight, either: in non-spatial relationships the use of the cases is more idiomatic, and the movement and direction are often fictive, so that the learner needs to "see the world in a Finnish way". Also, in the spatial domain some verbs require complements in local cases which indicate unexpected meanings, and accordingly, outline the situation differently from the learner's L1 (e.g. the verb *löytää* 'to find' takes the FROM-case in Finnish).

3.2. Research Questions

The study seeks answers to the following research questions: 1) How do the learners use the local case phrases in concrete and in metaphoric domains? 2) How do the static and dynamic uses differ from each other? 3) How do the domains differ from each other?

As the local cases are high in type frequency and have a wide spectrum of meanings and functions, the need for self expression makes them emerge in learner language from early on. However, the frequency, accuracy, and distribution develop from one level to another at a varying rate.

One hypothesis is that the spatial and static expressions are learnt first, as they are cognitively and linguistically simpler than the metaphoric and dynamic ones. This is the view of many cognitive theories (see e.g. Langacker, 1991; Jackendoff, 1983 on *locality hypothesis*) and also Finnish as L2 researchers (see Lauranto, 1997). As to the target-like uses of the cases, however, the dynamic and metaphoric uses of the local cases may be even more frequent than the concrete ones (for the frequencies of the Finnish cases see e.g. ISK, 2004, p. 1179).[2] Thus, in light of the usage-based view of language learning, whereby language is learnt in and through language use (see e.g. Tomasello, 2003), the predictable and consequential hypothesis is that the most common uses – the metaphorical and dynamic ones – of the local cases would be learnt first.

Therefore the overall aim of the study is to examine the basic assumptions about the learning order of local cases and to suggest which cognitive theory seems to be better at explaining the emergence of the local case system in L2 Finnish. The learning order was tested by using the L2 data from Cefling, and the concrete and metaphorical uses were compared across CEFR levels.

3.3. Some Results

Below, the frequency (per 1000 words) and accuracy of the static local cases are presented. The parameter of distribution is here built in the research design: the concrete and metaphoric uses are assumed to develop differently across the CEFR levels.

It turned out that the spatial expressions are most typical of level A1, after which their frequency in the data slightly decreases till level C2 (see the Figures 1 and 2). As the language skills develop, the learner is using fewer concrete expressions, and their frequency becomes more target-like.

The static expressions in the spatial field (*Spat-AT*)[3] are mastered as early as at level A1, even if the form of the case may still falter. In terms of accuracy, these phrases are mastered by level B1 (see Figure 1). The *Spat-TO*-phrases are

2 ISK = *Iso suomen kielioppi* 'The big Finnish grammar', the authoritative descriptive grammar of Finnish.

3 In the abbreviations below *Spat* refers to the spatial, concrete uses of local cases, *Circ* to the abstract, circumstantial uses. AT indicates static uses, TO and FROM dynamic uses.

Figure 1. The frequency and accuracy of spat-AT phrases

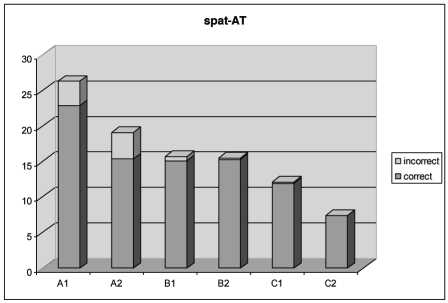
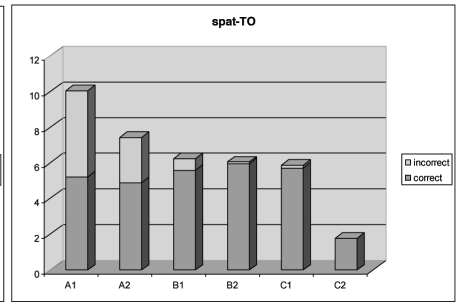


Figure 2. The frequency and accuracy of spat-TO phrases



more complex, both cognitively and morphologically, and hence the number of non-target-like uses is rather high at A1 and A2 levels (see Figure 2).
Meanwhile, the metaphoric uses of the local cases appear to emerge in the reverse order to the concrete uses. The metaphoric uses, *circ-AT* -phrases (*circ* indicating here internal states, action, circumstances, and roles) increase till level B2 (see Figure 3 and Example 1 from level B2).

1) *heillä kaikki on kunnossa*
They-ADESS everything be- PRES-3SG order-INESS
'Everything is in order with them.'

circ-TO -phrases increase till level C2 (see Figure 4). Again, the L2 Finnish learners produce a great number of incorrect TO-expressions, which, in fact, are not mastered until levels B2-C1 (see Figure 4).

Figure 3. The frequency and accuracy of metaphoric, static phrases

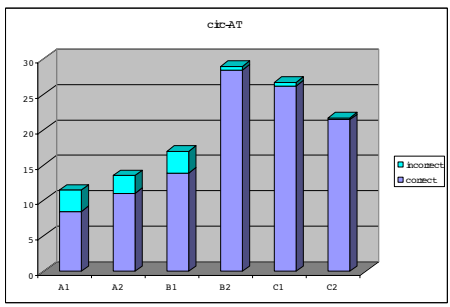
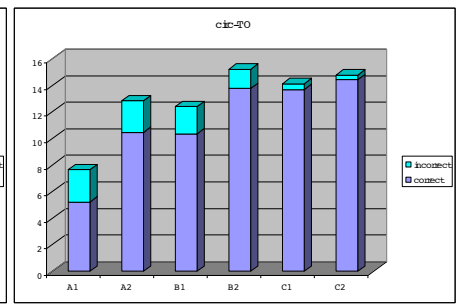


Figure 4. The frequency and accuracy of metaphoric, dynamic phrases



Thus, the developmental path suggested by the data fits the path predicted by one of the cognitively oriented theories and the locality hypothesis: the static, spatial expressions are mastered by level B1 at the latest, whereas it takes more time to learn the metaphoric uses of the same morphosyntactic features. The use of the local cases, particularly in metaphorical domains and in dynamic functions (TO and in particular FROM, which was not presented here but which is notable by its absence till C levels), is a challenge to learners across all CEFR levels.

It should be pointed out that the number of errors or the acquisitional order alone do not fully reveal the logic of the learning process. In addition, a more detailed qualitative analysis is needed to trace the path. The full description of the distributional growth – semantic accuracy, variation of stem words or verbs used in the phrases, etc. – which would provide a more complete picture of the developmental stages, is in progress but remains beyond the scope of this chapter. Characterizing the general tendencies of how the cases are used in concrete and in abstract domains at different CEFR skill levels and in different tasks may also shed light on the sociolinguistic features of the learner language.

4. Study 2 - The development of the transitive construction in L2 Finnish: The Finnish transitive construction

The Finnish transitive construction is an abstracted pattern used in constructing transitive clauses in actual language use. The prototypical construction is the [S [VO]] type: the subject is in the nominative case, but in certain sentence types it may also be in the genitive and partitive cases. In addition, the object takes three different major cases in Finnish: the cases of a total (bounded or resultative event) object are nominative and genitive (example 1), whereas the case of an unbounded (or irresultative, including a negative aspect), indefinite, or partial object is partitive (example 2). Personal pronouns and personal interrogative pronouns have a special accusative form ending in *-t*. Broadly speaking, the case depends on the meaning of the object, or the whole event/action (boundedness/unboundedness). Moreover, in the case of the total object, the choice between the nominative and genitive cases depends on the verb form: for instance, necessive and passive structures require a total object in the nominative case (example 3).

1) *Liisa kirjoitti esseen.* total, positive, resultative/finished action
Liisa write-PST-3SG essay-SG-**GEN**
'Liisa wrote an essay.'

2) *Liisa kirjoittaa esseet**ä**.* unbounded, irresultative action
Liisa write-PRES-3SG essay-SG-**PARTIT**
'Liisa is writing an essay.'

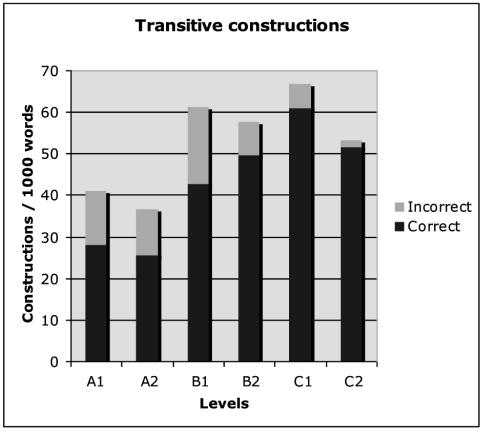
3) *Liisan täytyy kirjoittaa esse**e**.* total, positive, resultative action, **necessive**
Liisa-GEN must-PRES-3SG write-INF1 essay-SG-**NOM**
'Liisa has to write an essay.'

4.1. The development of frequency and accuracy of transitive constructions

The aim of this study is to examine how the frequency and accuracy of Finnish transitive constructions develops in learner language, and, for distribution, what kinds of uses of the transitive construction are typical at each CEFR proficiency level. The transitive construction is a category that is frequent and common in any language. It is semantically open and therefore high in type frequency. In terms of syntactic structure, the selection of the object case in the Finnish transitive construction causes problems, even for advanced learners. In other respects, the prototypical construction is fairly regular and word order is relatively free. Hence the accuracy of the construction is examined using case selection as a criterion of syntactic correctness. The selection of the object cases has been of interest in the area of L2 Finnish, whereas the construction as a broad syntactic unit has not been studied earlier.

Figure 5 presents the frequency of the transitive construction per 1000 words including both correct and incorrect uses across CEFR proficiency levels.

Figure 5. The frequency of the transitive construction / 1000 words across CEFR levels.



As can be seen from Figure 5, the construction emerges in learner language early on and is present at all levels of proficiency. At level A, however, the construction is not as frequent as in the following stages. There is some evidence that learners at level A tend to omit the object in otherwise possible transitive constructions: they use transitive verbs intransitively more often than more advanced learners:

4) *Minä sain uuden valokuvian ja **näytän sinulle**.* (A2)
I get-PAST-3SG new-SG-GEN photograph-PL-PARTIT-GEN and show-PRES-1SG you-ALL
'I got (*a) new photographs and **show you**.'

Regarding accuracy, there is a clear decrease in incorrect object cases between B1 and B2 (see Figure 5). Comparing frequency and accuracy, it can be seen that the apparent growth in the quantity of the construction takes place between A2 and B1, whereas accuracy does not increase until B2. According to the definition of the DEMfad model (mastery: 80% target like occurrences), the use of the object cases is mastered at level B2, in which 86% of the occurrences are target-like. By contrast, at level B1, accuracy is around 70%; almost one third of the uses are incorrect. It can be assumed that an increase in linguistic means may cause problems in accuracy at B1.

4.2. Qualitative change across CEFR levels

The increase of accuracy may indicate development to a certain extent, but it does not explain how the structure and its use change qualitatively. In other words, the quantitative approach can only reveal general tendencies of linguistic development: it does not allow us to draw any far-reaching conclusions about the learning process of a category as broad as a syntactic construction. Therefore, the concept of distribution will be employed here to understand qualitative changes, and will here refer to the types of clausal environments in which a given construction is used. A further interesting question is what types of syntactic variants of the transitive construction are typical at each of the levels. This question has been approached from the point of view of Construction Grammar (e.g. Fillmore & Kay, 1996; Goldberg, 1995), according to which constructions constitute a continuum on the basis of their productivity and abstractness, including idioms with fixed lexical content, idioms that are partially filled, constructions with some filled material, and fully general linguistic patterns (e.g. Goldberg 2003, pp. 219–220).

The findings of a tentative qualitative analysis show that the prototypical construction is the most typical variant at level A. However, it can to some extent be applied to different clausal contexts: for example, it is used in certain subordinate positions, and combined with simple infinitive constructions. The

ditransitive variant (e.g. *give somebody something*) is also used early on, which can be expected due to its semantically concrete nature and syntactically regular form in Finnish.

It is somewhat surprising that even though the use of the transitive construction increases between A2 and B1, the quality does not change that much. The prototypical and ditransitive variants are still frequent, but at the B1 level the transitive constructions are mastered also in subordinate clauses. Infinitive constructions used simultaneously with the transitive construction are more varied, and the necessity construction, for example, becomes frequent:

5) *jos mökillä on vieraat,* ***minun täytyy laitaa ruokaa*** *paljon ja tiskata paljon.* (B1)
If cottage-ALL be-PRES-3SG guest-PL-NOM I-GEN must make-INF1 food-SG-PARTIT a lot and do-INF1 (the dishes) a lot.
'If there are guests at the cottage, **I must make a lot of food** and do the dishes a lot.'

One apparent qualitative change, however, that differentiates B1 from B2 is that at B2 the construction is used in the passive voice much more often.

Lexically more constrained and thus semantically more specific and coherent (see Barðdal, in press) variants of the transitive construction do not emerge until level C. At level B, these kinds of idiomatic transitive expressions are used only occasionally. These partially filled or lexically fully fixed variants can also be understood as distinct constructions (Goldberg, 1995; Leino, 2009), but at the same time, they are still instances of the versatile, general and more abstract [S[VO]] pattern. From the viewpoint of learner language, it seems more fruitful to examine these occurrences as representatives of the prototypical construction, since this perspective reveals how the uses – and constraints – of the construction begin to diverge from those of the general one. The examples below illustrate the use of these less open types:

6) *He eivät ole* ***saaneet siirrettyä*** *opetusmateriaaliaansa uuteen järjestelmään.* (C2)
They no-PRES-3PL be-STEM-NEG **get**-PART-PL transfer-**PASS-PAST PART-PARTIT** teaching material-SG-PARTIT-POSS new-SG-ILL system-SG-ILL
'They haven't been able to transfer their teaching material into the new system.'

Example 6) above expresses resultative action. In the *saada tehtyä* ('get done') variant the auxiliary is always *saada* 'to get', whereas the main verb can vary, provided it is a passive past participle and in the partitive case.

7) *Kansa* ***teki*** *poliitikoi****sta*** *pellejä.* (C2)
People-SG-NOM **make**-PST-3SG politician-PL-**ELAT** clown-PL-PARTIT
'People made politicians clowns.'

Example 7) is an instance of the transitive construction that expresses a change of state. The fixed elements are the verb *tehdä* ('to make') and the elative case *-sta*. In addition to these more constrained transitive expressions, other typical advanced uses at level C are non-finite clauses: they are hardly used at lower levels.

In sum, the clausal environments in which the transitive construction is used become more diverse as language proficiency develops. An increase in linguistic means makes it possible to express meanings more specifically (level C). Furthermore, these results also support the findings concerning the use of passive (Study 3) and local expressions (Study 1).

5. Study 3 - The Finnish passive

In L2 Finnish the use of generic expressions, such as the passive, can be a good indicator of proficiency. Genericity signifies actions or events that are described without specifying the agent, i.e., the person in action or the person observing. By means of generic expressions it is therefore possible for language users to reach a more abstract level in their communication, as opposed to using personal expressions. The aim of this study is to examine the development of the uses of passive in learner language.

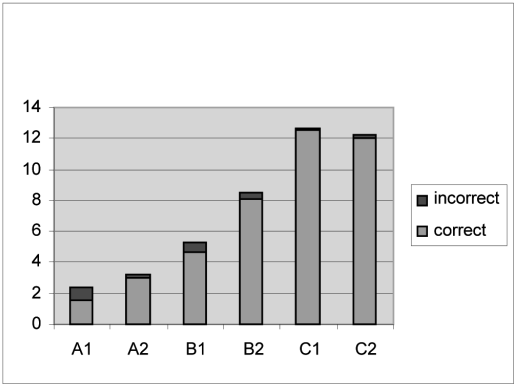
The use of all generic expressions is very common in Finnish, and it is a characteristic of both spoken and written language (see Hakulinen, Karlsson, & Vilkuna, 1980). Besides the passive, other generic forms include e.g. the zero person and the *sinä* 'you'-passive, all of which can be used to create open reference. Furthermore, passive forms are commonly used to replace 1st person plural forms. In this study, however, genericity in L2 Finnish is examined by focusing on only one of the generic expressions: the impersonal passive construction. It is investigated how this construction emerges, varies and develops in the L2 Finnish writing tasks.

The term *impersonal* is used here to convey that the agent has no specific referent. The use of the structure in question, with a passive form as the finite verb, has not been studied before in the context of L2 Finnish. The Finnish passive has the special characteristic that it always refers to a human agent: *Jos ovi avataan.* 'If the door is opened (by somebody)' (see e.g., Shore, 1986). Therefore the Finnish passive is currently often seen as a part of the personal system in Finnish (e.g., Helasvuo, 2006). The passive is a structure lacking an overt subject although the agent can be specified through the context (e.g., Laitinen, 2006). In a passive sentence the agent remains implicit and the passive describes events in less detail than an active sentence (ISK, 2004, pp. 1254–1256, 1284).

5.1. The use of passive in L2 Finnish

The passive construction as a generic expression is present at all proficiency levels from A1 to C2, and the use increases up until C1 level. A more distinguishable increase in the use of this impersonal expression occurs between levels B1 and B2, and levels B2 and C1.

Figure 6. Frequency and accuracy of the passive construction (/1000 words)

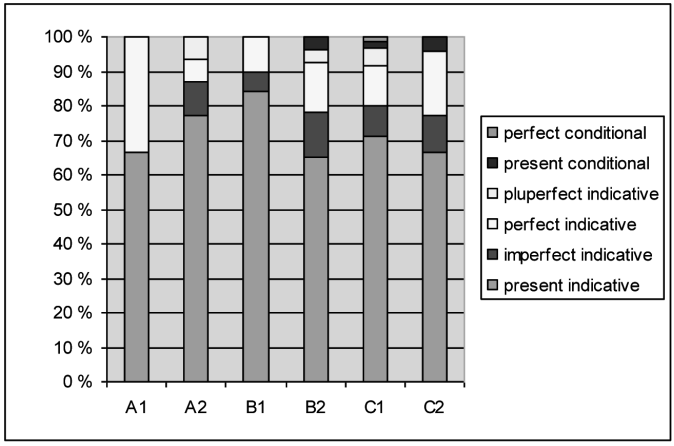


Non-target passive expressions can be found somewhat more frequently at levels A1 and B1, however, at level C the accuracy of the generic passives is almost 100 %: there are only three inaccurate passive forms in the texts. In terms of the DEMfad-model, mastery for accuracy is already reached at level A2 (over 90 % of the forms accurate). In L2 Finnish the passive occurs mostly in the present indicative tense (Figure 7). The morphological form of the passive present might be relatively easy to acquire for a learner of Finnish because in most verb types it is very similar to the dictionary form of the verb, the 1st infinitive (*tehdä* 'to do', *tehdään* 'is done'). The form is also familiar not only from its use as the 1st person plural imperative (*tehdään* 'let's do') but also from the spoken language, where, completed with a personal pronoun, it is widely used in the sense of the 1st person plural indicative (*me tehdään* 'we do').

The similarity between the 1st infinitive and the colloquial use of the passive form in active meaning is arguably also the source of non-target use of the passive, which is most common at levels A1 and B1 (Figure 6). There are examples of the inaccurate use of passives of this kind: the 1st infinitive form is used as a passive or vice versa, for example *myydä* 'to sell' and pro *myydään* 'to be sold'; *mahdollisuus puhutaan suomen kieliä* pro *puhua* 'possibility to speak Finnish'. Moreover, the auxiliary verb is used in the active and the main verb in

the passive in compound passive forms (*olin sanottu* - pro *minulle oli sanottu* 'I was told') and an active construction is blended with a passive one (*minä lähetetään* pro *minä lähetän* 'I will send'; *verkkosivuilla luetaan* pro *verkkosivuilla lukee/ verkkosivuilta voidaan lukea* 'one can read on the websites'.

Figure 7. Tense and mode variation within the passive construction occurrences



While the use of the passive in L2 Finnish increases gradually from one level to another, not only does the range of the verbs grow but also the variation in tenses and modes (Figure 7). However, the present tense is the most frequently used across all levels. At levels A1 – A2 the verbs used in the passive are basic types such as *puhua, mennä* 'speak, go'. As for tenses, at A1 only the present and perfect tenses occur. Figure 7 shows that also the passive imperfect and pluperfect tenses emerge in the texts soon, already at level A2. Conversely, the conditional form, for one, does not appear until level B2. All in all, at level B verb selection varies with synonyms (*sanotaan, mainitaan, on kehuttu* 'is said', 'is mentioned', 'has been praised'). The full repertory of passive forms, along with the past conditional, can be found at level C1 (Figure 7). Compared to B2, the expressions are more abstract and idioms are used at level C (*kulttuuri on valjastettu kaupallisten intressien vetäjäksi* (C2) 'the culture has been harnessed to lead commercial interests'; *on nostettu pöydälle aihe, josta kannattaa todella keskustella* (C1) 'a subject has been raised which is truly worthy of discussion'. The distribution and internal variance of the passive construction thus clearly develops across the levels.

6. Discussion

In addition to providing new information on the development of certain aspects of L2 Finnish writing, the purpose of drawing together results from somewhat disparate domains is theoretical and methodological. Below the implications of the results are discussed in relation to the CEFR and some constructs of the CAF triad.

The three domains of developing Finnish, the use of cases and transitive and passive constructions, studied in this chapter were each examined for frequency, accuracy, and distribution across the CEFR levels. The results of such a varied set of structures – and slightly different theoretical approaches – are obviously variable. As can be expected, the occurrence and accuracy of structures generally increases across the levels. The growth of the passive use in Study 3 is a good example of a fairly steady development, peaking at level C.

When the structure itself is extremely frequent and can hardly be avoided in any type of communicative task, as the local cases in Study 1, the total number of occurrences remains fairly stable while the proportions of the cognitively different uses of the structure have a converse relation: the concrete uses of local cases decrease and the metaphorical ones increase. A similar steady total frequency was found in Kynsijärvi's (2007) study of the verb *to be*, which was mainly used in present and past personal forms (i.e. single word forms) at level A. The auxiliary use of *to be* grew at level B, and other constructions, similarly restricted in meaning and form as the examples given in Study 2, were normally found only at level C.

The transitive construction in Study 1 is a good example of a structure whose frequency grows most between levels A2 and B1, while the accuracy "leap" is between B1 and B2. The development of noun phrases (Ukkola, 2009) was found to have a similar pattern but the frequency increased even more from level B to level C (both differences statistically very significant). For negation structures (Martin, 2008) the total frequency remained stable across the levels, as one might expect, while the accuracy growth from A2 to B1 was statistically significant when compared to the other steps of the CEFR scale. As the studies reported in this chapter are still at the stage of looking for the best indicators of development, no statistical testing has been done here. Nevertheless, it is interesting if leaps of development of any two domains should occur at the same place.

The rating scales used for proficiency were purely communicative, i.e. contained no reference to the language structures or vocabulary, and the raters were strictly instructed to pay attention to communicative proficiency only. They were also experienced and trained in this type of rating (see Alanen et al., this

volume). Yet it is possible – and unavoidable – that the linguistic features of the texts subconsciously influenced the rating and thus explain the results. Even if this is the case, however, there is no a priori reason to expect a similar pattern of development across any two domains, particularly as the domains studied in the project are quite different in nature.

Assuming that the leap from level A to level B, found in some studies in the frequency and/or accuracy of use of linguistic structures, is due to actual developmental factors, why is it where it is? Are the CEFR levels simply not equidistant in such a way that the communicative proficiency suddenly takes a bigger step between A2 and B1 than between other levels? This would explain some of our results: communicative and structural skills grow step-by-step. An alternate explanation is the one displayed in the title of this chapter: the concept of Independent User. It is possible that to become an independent user one needs a fairly large repertoire, and a degree of control, of target language structures, while the Basic User can function non-independently, with the help of interlocutors or, when writing, rely on the willingness of the readers to decipher the intended meaning from fragmentary expressions, without much grasp of the grammatical features of the language.

As to the threesome of complexity, accuracy, and fluency, the latter two seem to respond to the quantitative measures in this study, as in numerous studies before this one. Yet the theoretical question of construct definition remains: to what extent do these measures cover the construct of accuracy, let alone fluency? Finding an answer to this theoretical question remains elusive.

The third parameter, the complexity of the CAF triad, has not been subjected to any type of quantification here as the authors do not find any of the existing measures sufficiently refined. A potential measure which has been considered for future work is IPSyn (Scarborough, 1990). It has been applied to Finnish L1 development in Nieminen (2007). The results, however, were not entirely conclusive, because in the structural development of Finnish (whether L1 or L2) the morphological and syntactic issues are intricately entangled, which constitutes a challenge to all theories, models, and methods of the study of syntactic development. To clarify these issues Nieminen (2007) offers another approach, Utterance Analysis. It offers many insights into the co-development of morphology and syntax, but is too detailed and work-intensive to offer a solution to a large-scale study. In the future we plan to look at new ways of solving complexity issues, such as the application of Dynamic Systems Theory (see e.g. de Bot, Lowie, & Verspoor, 2007). In any case, the results of the Cefling project so far make it clear that the way linguistic structures are used changes across the CEFR levels. Whether this is called complexity or distribution, better qualitative and quantitative ways of accounting for this growth are required.

The construction-based view of language brings up two interesting issues. One is the question of emergence. All structures we have studied to date are present at level A1. Obviously all structures are not there in a single piece of writing but at the group level there is no doubt that structures emerge earlier than is often thought to be the case. There are examples of verb chains (Paavola, 2008), transitive constructions (Study 2 here), and passive (Study 3). Subordination is very common (Martin, 2009). In curricula and elementary textbooks these issues are considered difficult and presented late, if at all. In research it has been customary to write off these occurrences as unanalyzed chunks (see e.g. Pienemann, 1998), at best serving as input for later grammatical learning. In our approach the chunk is a construction which has been learnt, with the potential of expansion and variation provided by future encounters with similar occurrences.

Another question is raised by the construction-and-distribution approach of the Cefling studies: In addition to the three well-known CAF dimensions, is there another one, something which might be called abstractness? Unlike the rule-based approaches, the construction-based framework seems to bring out something about the growth of not only complexity of structures but also the growth from concrete to abstract uses of the structures. In each of the three domains discussed in this chapter there are signs of some "fourth dimension" which also seems to develop across the levels. The abstractness (metaphorical use) is clearly present in the use of local cases (Study 1), as it is built in the theoretical framework of the study, but also transitive and passive constructions seem to extend not only in the number of verbs but also in the quality and variety of verbs. The contexts where the construction is used become more abstract. Transitive constructions are used with abstract subjects, verbs, and objects. The many types of generic expressions of Finnish, including the passive, differentiate and indicate more refined details and implications. In all domains of this chapter (local case use, transitive and passive constructions) expressions become more idiomatic as constructions with limited possibilities of variation and non-literal meaning are added.

The very tentative notion abstractness, however, requires careful definition to avoid circularity. After all, some reference to more abstract uses of language is often made in the assessment criteria of communicative development. What is required is to separate the abstract uses of constructions from the abstraction level of the topic and ideas. Like distribution and variability, abstractness could also be included as one face of complexity in the future search for its more multidimensional definition.

References

Abrahamsson, N., & Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning, 59*(2), 249–306.

Barðdal, J. (in press). Predicting the productivity of argument structure constructions. *Berkeley Linguistics Society* 32.

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press. Retrieved from http://www.coe.int/t/dg4/linguistic/Source/Framework_EN.pdf

De Bot, K., Lowie, W., & Verspoor, M. (2007). A Dynamic Systems Theory approach to second language acquisition. *Bilingualism: Language and Cognition, 10* (1), 7–21.

Eskildsen, S. W. (2008). *Constructing a second language inventory – the accumulation of linguistic resources in L2 English* (Doctoral dissertation). University of Southern Denmark.

Fillmore, C. J., & Kay, P. (1996). *Construction grammar.* [Lecture notes]. Center for the Study of Language and Information, Stanford, CA.

Franceschina, F., Alanen, R., Huhta, A., & Martin, M. (2006, December). *A progress report on the Cefling project.* Paper presented at SLATE Workshop, Amsterdam.

Goldberg, A. E. (1995). *Constructions. A construction grammar approach to argument structure. Cognitive theory of language and culture*. Chicago, IL: The University of Chicago Press.

Goldberg, A. E. (2003). Constructions: A new theoretical approach to language. *Trends in Cognitive Science*, *7*, 219–224.

Hakulinen, A., Karlsson, F., & Vilkuna, M. (1980). *Suomen tekstilauseiden piirteitä: kvantitatiivinen tutkimus*. Publications/Department of General Linguistics. No 6. University of Helsinki.

Helasvuo, M.-L. (2006). Passive personal or impersonal? A Finnish perspective. In M.-L. Helasvuo & L. Campbell (Eds.), *Grammar from the human perspective: Case, space, and person in Finnish* (pp. 233–255). Amsterdam: Benjamins.

Housen, A., & Kuiken, F. (Eds.). (2009). Complexity, accuracy and fluency (CAF) in second language acquisition research [Special issue]. *Applied Linguistics*, *30*(4).

Huumo, T., & Ojutkangas, K. (2006). An introduction to Finnish spatial relations: Local cases and adpositions. In M.-L. Helasvuo & L. Campbell (Eds.), *Grammar from the human perspective: Case, space and person in Finnish* (pp. 11–20). Amsterdam: Benjamins.

ISK = Hakulinen, A., Vilkuna, M., Korhonen R., Koivisto, V., Heinonen, T., & Alho, I. (2004). *Iso suomen kielioppi*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Jackendoff, R. (1983). *Semantics and cognition*. Cambridge, MA: MIT Press.

Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination, and reason*. Chicago, IL: The University of Chicago Press.

Kynsijärvi, T. (2007). *Se johtuu siitä, että minulla oli muistinmenetys. Olla-verbirakenteiden kehkeytyminen oppijankielessä* (Unpublished master's thesis). Department of Languages, University of Jyväskylä.

Laitinen, L. (2006). Zero person in Finnish. A grammatical resource for construing human reference. In M.-L. Helasvuo & L. Campbell (Eds.), *Grammar from the human perspective: case, space, and person in Finnish* (pp. 209–231). Amsterdam: Benjamins.

Lakoff, G. (1987). *Women, fire and dangerous things*. What categories reveal about the mind. Chicago, IL: Chicago University Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago, IL: The University of Chicago Press.

Langacker, R.W. (1991). *Foundations of cognitive grammar: Descriptive application*. Stanford, CA: Stanford University Press.

Lauranto, Y. (1997). *Ensi askeleita paikallissijojen käyttöön. Espanjankielisten suomen oppijoiden sisä- ja ulkopaikallissijat konseptuaalisen semantiikan näkökulmasta*. Department of Finnish, University of Helsinki.

Leino, J. (2003). *Antaa sen muuttua. Suomen kielen permissiivirakenne ja sen kehitys* (Doctoral dissertation). Helsinki: Suomalaisen Kirjallisuuden Seura.

Leino, J. (Ed.). (2008). *Constructional approaches to language* (Vol. 5). Amsterdam: Benjamins.

Leino, J. (2009). *Results, cases, and constructions. Argument structure constructions in English and Finnish*. Manuscript in preparation.

Leino, P., Herlin, I., Honkanen, S., Kotilainen, L., Leino, J., & Vilkkumaa, M. (2001). *Roolit ja rakenteet. Antaminen, allatiivi ja suomen datiivin arvoitus*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Martin, M. (2008, September). *Negation as a marker of proficiency level in L2 Finnish*. Poster presented at the 18th EUROSLA Conference, Aix-en-Provence, France.

Martin, M. (2009, November). *Lauseet ja sanat kielitaidon kehityksen osoittimina: onko määrä laatua?* Paper presented at the symposium of AFinLA, the Finnish association of applied Linguistics in Tampere, Finland.

Nieminen, L. (2007). *A complex case: a morphosyntactic approach to complexity in early child language* (Doctoral dissertation). Jyväskylä Studies in Humanities 72. University of Jyväskylä.

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, *30*, 555–578.

Paavola, V. (2008). *Haluatko menna muunkansa kalastaman? Verbiketjujen kehkeytyminen suomi toisena kielenä -oppijoiden kielessä* (Unpublished master's thesis). Department of Languages, University of Jyväskylä.

Pallotti, G. (2007). An operational definition of the emergence criterion. *Applied Linguistics*, *28*, 361–382.

Pallotti, G. (2009). CAF: defining, refining and differentiating constructs. *Applied Linguistics, 30*, 590–601.

Pienemann, M. 1998. *Language processing and second language development. Processability theory.* Studies in bilingualism 15. Amsterdam/Philadelphia: Benjamins.

Scarborough, H.S. (1990). Index of productive syntax. *Applied Psycholinguistics, 11*(1), 1–22.

Shore, S. (1986). *Onko suomessa passiivia?* Suomi 133. Helsinki: Suomalaisen Kirjallisuuden Seura.

Tomasello, M. (2003). *Constructing a language. A usage-based theory of language acquisition.* Cambridge, MA: Harvard University Press.

Ukkola, A. (2009). *Taas yksi nollatutkimus – substantiivilausekkeiden määritteet S2-oppijoiden kielessä* (Unpublished master's thesis). Department of Languages, University of Jyväskylä.

Visapää, L. (2008). *Infinitiivi ja sen infiniittisyys. Tutkimus suomen kielen itsenäisistä A-infinitiivikonstruktioista* (Doctoral dissertation). Helsinki: Suomalaisen Kirjallisuuden Seura.

Wolfe-Quintero, K., Inagaki, S., & Kim, H.-E. (1998). *Second language development in writing: measures of fluency, accuracy, and complexity* (Technical report No. 17). Second Language Teaching & Curriculum Center: University of Hawai'i, Mānoa.