

INDEXED GRAMMAR IN FINNISH SYNTAX

JUDIT FARKAS

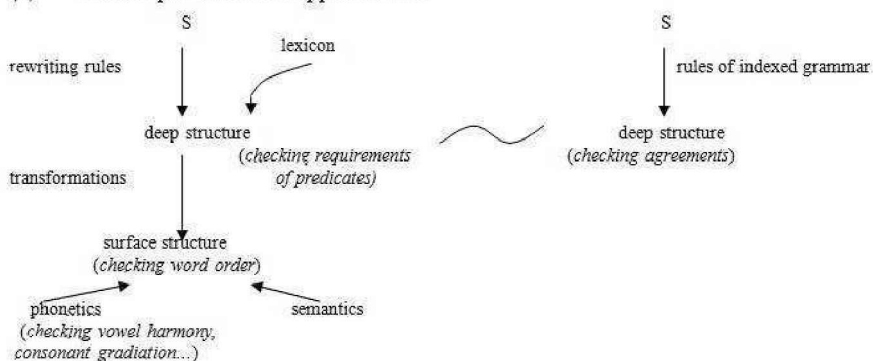
0. Introduction

This paper¹ discusses how an indexed grammar, which belongs to the group of the mildly context sensitive grammars (Partee *et al.* 1990) and provides a straightforward way of treating agreement phenomena, can be applied in the description of the Finnish language.

It is an important general motivation for the work that natural languages have been proven to belong to the group of the mildly context sensitive grammars (Partee *et al.* 1990). As for the particular reasons, using an indexed grammar in the description of Finnish is extremely useful because in this language—especially in the literary style—there is a large variation of agreement phenomena: this kind of grammar emphasizes not only the essential features of certain structures, but many differences can be captured simply and spectacularly between the written and spoken Finnish exactly via reference to certain indices.

The first two parts of the paper demonstrate how a model with an indexed grammar component is structured (see figure (1)). The model applied consists of two major components: a transformational component and an indexed grammar. While the former is often applied in the description of languages in numerous different ways, the latter can be found extremely rarely in Hungarian generative works. Kornai (1989) is a counterexample, which discusses the agreement phenomena of Hungarian noun phrases in this kind of indexed framework.

(1) The components of the applied model



¹ Special thanks are due to Anja Haaparanta, the Finnish lector of the University of Pécs, who reviewed the Finnish example sentences of the dissertation and helped me a lot in this work as a native speaker. I also thank an OTKA project (No. 60595) for supporting this work.

The indexed grammar part of the model is used not only for the effective description of agreement phenomena, appearing often and in various ways in Finnish, but it makes the description more explanatorily adequate by capturing the real nature of this phenomenon, which is *copying*. This indexed part of the grammar, however, can be rephrased into an (index-free) context free grammar because the feature of indexed grammars that the string of indices is allowed to be unbounded has not been exploited (for capturing the agreement phenomena of Finnish, more than five indices need not appear on a certain constituent).

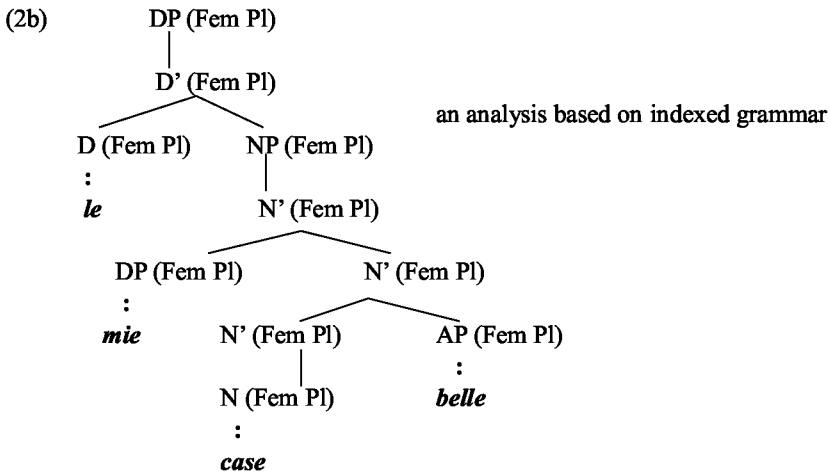
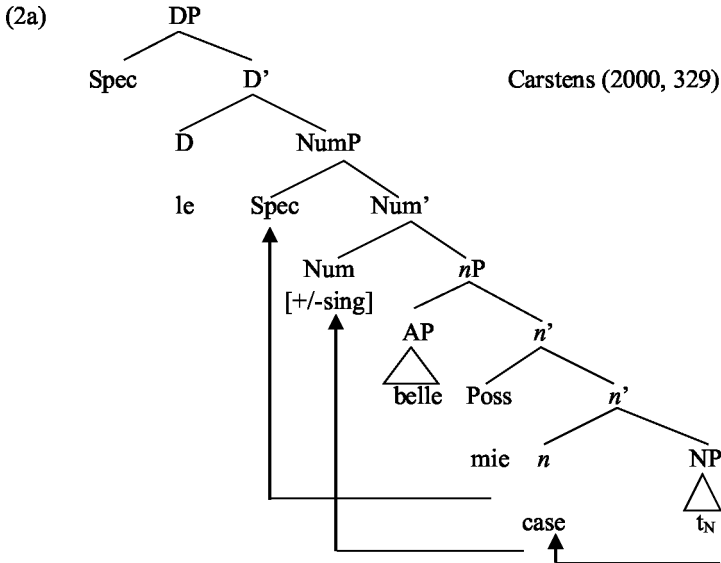
In this way this description captures the essence of agreement phenomena better while at the same time it can also be used for researchers working in a “more traditional” Chomskyan framework.

In part 3 it will be illustrated (by means of an example) how the suggested model works in practice, and in part 4 the most important points will be summarized.

1. The indexed grammar

Indexed grammars can be used primarily in the description of agreement phenomena. Linguists working in different transformational generative frameworks mainly have the same idea concerning the syntactic handling of congruence: they offer a feature-checking solution, which requires an appropriate syntactic structure. It is common in all of these approaches that quite complicated syntactic structures have to be built to get a proper configuration and sometimes such nodes are needed that have poor motivation. As Kornai pointed out, “the transformational treatment of agreement phenomena needs *copying* instead of *moving*, and this kind of transformations lacks an independent motivation” (Kornai 1989: 183). Indexed grammars are excellent exactly in copying: by using them the essence of agreement phenomena can be captured better, and many nodes and transformations needed in feature-checking systems become unnecessary. Example (2) demonstrates two analyses of an Italian sentence; in this way a “classical” analysis (2a) and one using an indexed grammar (2b) can be compared. In the model suggested here the “cheaper” *Merge* is used instead of *Move*, so it is more economical as well.

- (2) *le mie case belle*
 the-FEM.PL my-FEM.PL house-FEM.PL beautiful-FEM.PL
 'my beautiful houses'



Indexed grammars do not belong to the best-known systems. Their characteristics and working features are listed in (3-8) below. In what follows, I give account of the components of these kinds of grammars, the rules of appearing and disappearing of indices in generations.

$G = \langle V_T, V_N, I, S, R \rangle$: indexed grammars (Partee *et al.* 1990: 536–42; Alberti 2006: 251–58) consist of five components.

V_T and V_N are disjoint finite sets, where V_T is the terminal alphabet of grammar G (these are the lexical elements); V_N is the non-terminal alphabet of grammar G (these are the categories: syntactic categories and lexical non-terminal elements²).

I is the set of indices. In this particular model indices play a crucial role in accounting for agreement. S is a distinguished element of V_N : the starting symbol of the grammar.

R is the set of the rules of grammar G .

Indices can appear on non-terminal elements: $A(i)$, for instance, can be realized as $N(Pl, Elat)$, which represents a plural elative noun. Indices can appear or disappear in a derivation according to the following rules:

- (3) An index can appear or disappear exclusively in the left end of the string.
- (4) The length of the string of indices is not restricted.
- (5) Indices descend on non-terminal symbols.
- (6) Terminal symbols never bear indices (indices disappear when symbols become terminal);
- (7) "Pop and copy" (rules referring to disappearing indices): $A(i) \rightarrow \alpha$, where i is an index, A is a non-terminal symbol and α is a mixed string consisting of terminal and non-terminal elements. It can only be applied when i is the first element of the string of indices.
- (8) "Push and copy" (rules referring to appearing indices): $A \rightarrow B(i)$, where A and B are non-terminal symbols and i is an index. On the output of the rule, thus, only a single non-terminal symbol is allowed; which is a slight limitation.

A string over V_T can be generated if there is a derivation from S whose final transitional string contains only terminal symbols (this final string thus contains no non-terminal elements or indices).

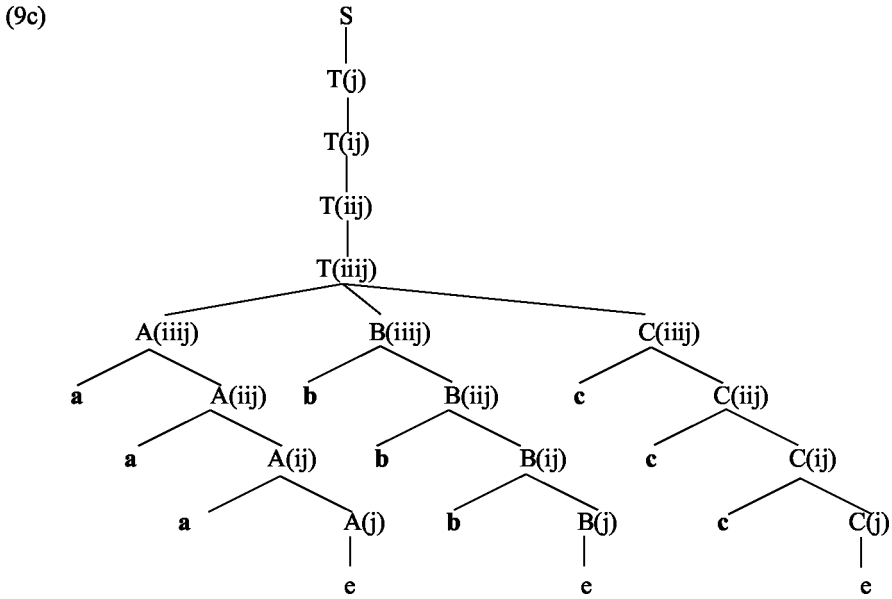
Example (9) illustrates how it functions in practice.

(9a) Language $L = \{x \in \{a, b, c\}^* : x = a^n b^n c^n, \text{ where } n \in \mathbb{N}\}$, generated by indexed grammar $G = (V_T, V_N, I, S, R)$ where:

$V_T = \{a, b, c\}$
 $V_N = \{S, T, A, B, C\}$
 $I = \{i, j\}$
 $R = \{S \rightarrow T(j), T \rightarrow T(i), T \rightarrow ABC;$
 $A(i) \rightarrow aA, A(j) \rightarrow e;$
 $B(i) \rightarrow bB, B(j) \rightarrow e;$
 $C(i) \rightarrow cC, C(j) \rightarrow e \}$

² *Lexical non-terminal*: non-terminal elements in the place of lexical elements, written in capital letters in figures, which, after absorbing indices, become terminal elements by phonetic operations like vowel harmonization and consonant gradation: e.g., *PEKKALLA* \rightarrow *Pekalla*.

(9b) $S \rightarrow T(j) \rightarrow T(ij) \rightarrow T(iij) \rightarrow A(iij)B(iij)C(iij) \rightarrow aA(iij)bB(iij)cC(iij) \rightarrow aaA(ij)bbB(ij)ccC(ij) \rightarrow aaaA(j)bbbB(j)cccC(j) \rightarrow aaabbbccc$



An indexed grammar reaches its highest capacity when generating, say, an $a^nb^nc^n$ type of language (see (9) above), because the string of indices applied in the course of derivation is permitted to be unbounded in the case of this kind of languages.

In the model proposed for the description of Finnish (Farkas 2005a; Farkas 2009), not more than five indices can appear in a certain element, and the order of indices is not optional either. Indices appear in structures according to Baker’s *Mirror Principle* (see Bartos 2000: 654–9)³: in the case of verbs, the order of indices is *mood – tense – number and person (of the subject) or impersonality – negation or progressive aspect – imperative*; while in the case of nouns, only the following order of indices can be found: *number – case – number and person (of the possessor)*, and there can not be more than one from a certain type of index. Restricting the length of strings of indices results in a generative grammar reduced to context free capacity; and this is combined with a transformational component.

³ „*Mirror Principle*: morphologic derivation directly reflects syntactic derivation”, i.e. the order of morphemes is the same as that of the syntactically relevant changes forming words (Bartos 2000: 657).

2. The transformational component

The transformational part of the applied model is based on the standard X'-Theory. It can be regarded as a classical generative one (e.g., Alberti–Medve 2000, 2002), which means that heads can move only to heads while phrases can move to specifiers or to complement positions next to heads. In this model (similar to that of Bartos 2000) syntactic rules pertain to the morpheme level of the language as well. Each productive morphological operation functions in the syntax: not only inflectional morphemes realizing agreements are treated here but productive derivations take place in this component as well. This is because in the indexed component of the model features taking part in agreement phenomena descend as indices and they become particular suffixes only at the end of the derivation, when they “terminalize”, that is, receive a phonetic form, and is connected to a root. For that reason the verb is inserted in the structure as a bare root; and this approach makes it possible to show both the verbal and the nominal features of infinitives and participles. In these kinds of phrases the verb root moves to a head of an infinitive or a participial phrase, where it concatenates with a certain derivational suffix in the same way as it concatenates with the tense or mood suffixes (which have index origins) in finite structures. Head movements, thus, are often used: a root moving higher and higher in the structural hierarchy concatenates with other heads yielding complex heads. In these complex heads the abstract morphemes and the features expressed by indices receive their terminal form in the phonetic form, where (among others) vowel harmonization and consonant gradation take place. For further information on how these abstract morphemes materialize, see Farkas (2009).

Every feature that is relevant more than once in a structure (primarily due to any kind of agreement) is to appear as an index, while a node is to be built according to X'-Theory in the case of elements that are relevant only once.

The proposed model (intended to capture the competence of native speakers of Finnish) accepts a string of words as a grammatical sentence if all the modules accept it (the relation among the components is presented in figure (1) above). The indexed grammar part is to check the deep structure from the point of view of agreement. Indices (in harmony with the nature of this kind of grammars) disappear from the derivation after they have done their checking work. Otherwise, if there remains a non-terminal symbol with an index, the given derivation fails, predicting the input string to be ungrammatical.

Transformations take place in another module. They operate on a deep structure that is like the above discussed deep structure containing indices—with the exception of just the indices themselves.

Transformational rules were needed besides the indexed grammar because the latter is not enough for the complete description of sentences in itself. To a complete analysis, thus, three representations are needed for each sentence. Due to space limitations, only one representation will be given (for each sentence) where the entire information appears together.

3. An example

After demonstrating the main points of the applied model, some relevant and interesting structures will be discussed. Example (10) is a complicated sentence, which contains almost everything relevant. Figure (13a) gives the representation of sentence (10), and figure (13b) illustrates how certain verbal indices terminalize.

- (10) *E-tte-kö ol-leet mielellä-nne halu-nneet anta-a*
 not-PL2-QP be-PTC.PL with_pleasure-PL2 want-PTC.PL give-TAINF
nä-i-tä kah-ta paksu-a lue-tta-va-a kirja-a Pekka-lle?
 this-PL-PART two-PART thick-PART read-PASS-VAPTC-PART book-PART Pekka-ALL
 'Is it true that you did not want to give Pekka these two thick books that should be read with pleasure?'

Let us overview the categories which can be seen in figure (13a).

If an indexed grammar is used in the description of Finnish, only the following six phrases are needed:

- (11) [CP [TP [NegP [IP [VoiceP [VP ...]]]]]]

Two phrases less are needed compared to the models of Holmberg *et al.* (1993) and Manninen (2003a), because indices can take the function of certain phrases. The six phrases required even in this framework have the following functions and interpretations.

In the head of CP the following constituents can appear: the subordinate conjunction; the (negative, auxiliary or main) verb, supplied with a *-ko/-kö* particle in the case of yes/no questions (like in example (10)); the finite verb in imperative sentences or if it is supplied with a *-han/-hän* or *-pa/-pä* particle. The [Spec, CP] can be occupied by an interrogative word or an XP supplied with the above mentioned particles.

I suggest TP instead of F(in)P, suggested by Holmberg *et al.* (1993) and Manninen (2003a) as the name of the phrase in whose specifier the topic of the sentence appears and in whose head the (complex) verb bearing an affix denoting finiteness in some way appears. This phrase is undoubtedly needed for semantic reasons rather than for syntactic ones. If TP is used, the definition of the topic is very simple: the topic is the constituent standing in [Spec, TP]. In this approach the definition of the topic does not depend on whether the sentence contains negation or not. TP and CP create a similar zone of functional projections to TP, QP and FP in the description of Hungarian, and they are separated from phrases responsible for assembling the verb form.

In the head of NegP, the negative verb appears. If the sentence does not contain negation, this phrase is absent from the structure.

IP is the phrase responsible for mood: it is in its head that the index of mood is converted into particular morphemes. The mood is required to be present in the system as an index because in negative imperative sentences the morpheme referring to mood may appear on both the negative verb and the main verb.

VoiceP is the phrase responsible for distinguishing between the active and the (Finnish-type) passive. In contrast to Holmberg *et al.* (1993), in whose opinion PassP is

only required if the sentence contains a passive verb, I suggest that every sentence structure must contain this projection. Up to the head of this PassP, the verb must always move; in this way the verb will precede the subject occupying the [Spec, VP]. This analysis enables to account for sentences in which this distinguished argument is preceded by the verb.

The subject is generated in [Spec, VP] whilst in the head of V the bare verb root is generated⁴, which, in the course of its moving up in the structure, is adjoined to suffixes expressing tense, mood, number, and person, which can be found in the corresponding functional heads.

InfP is the phrase of infinitives. A verb is inserted in the structure as a bare root and then climbs to the head of InfP, where it concatenates with the derivational suffix *-TA*. This way of handling productive derivational operations makes it possible to show both the verbal and the nominal features of these words (Farkas 2005b). The same holds for phrase Partis.

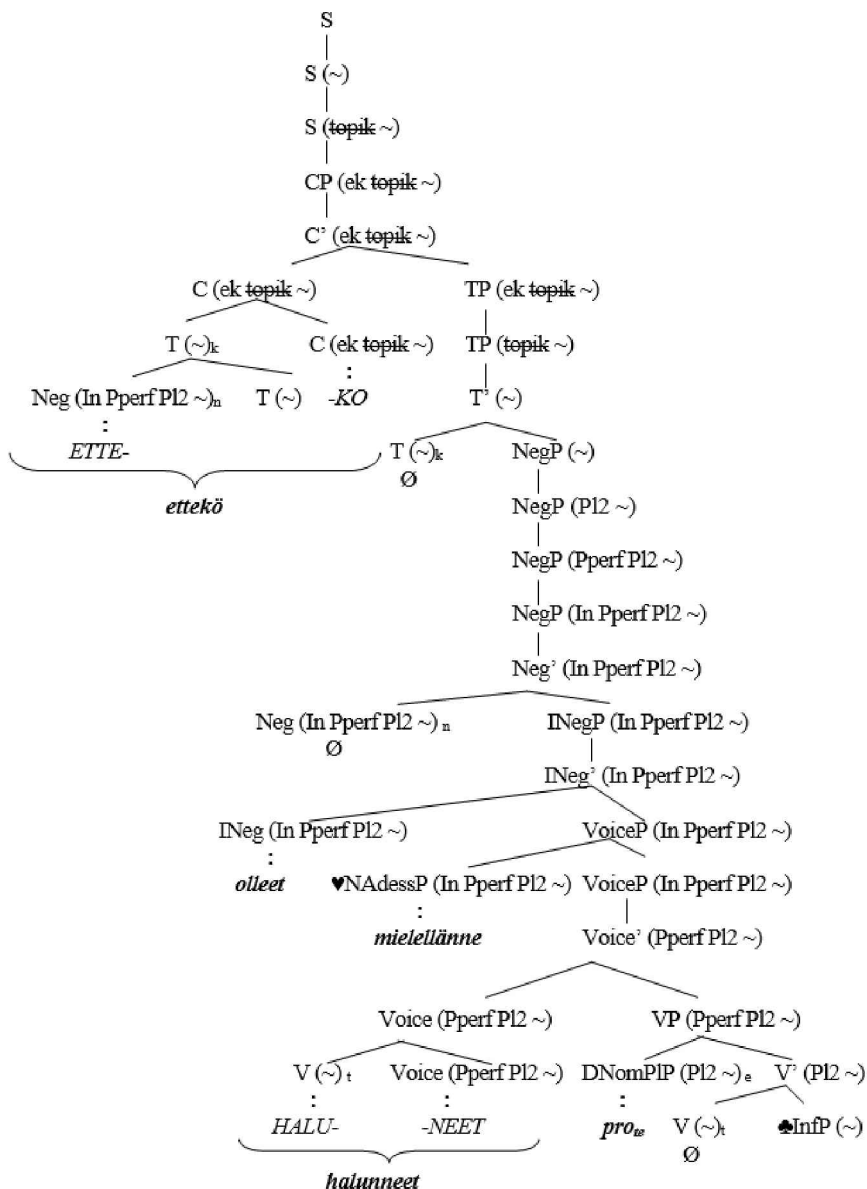
D is the phrase of articles. Literary Finnish does not have an explicit article yet, but this category is useful in distinguishing nouns that behave syntactically in different ways (compare (12a) containing a DP to (12b) with an NP in its structure).

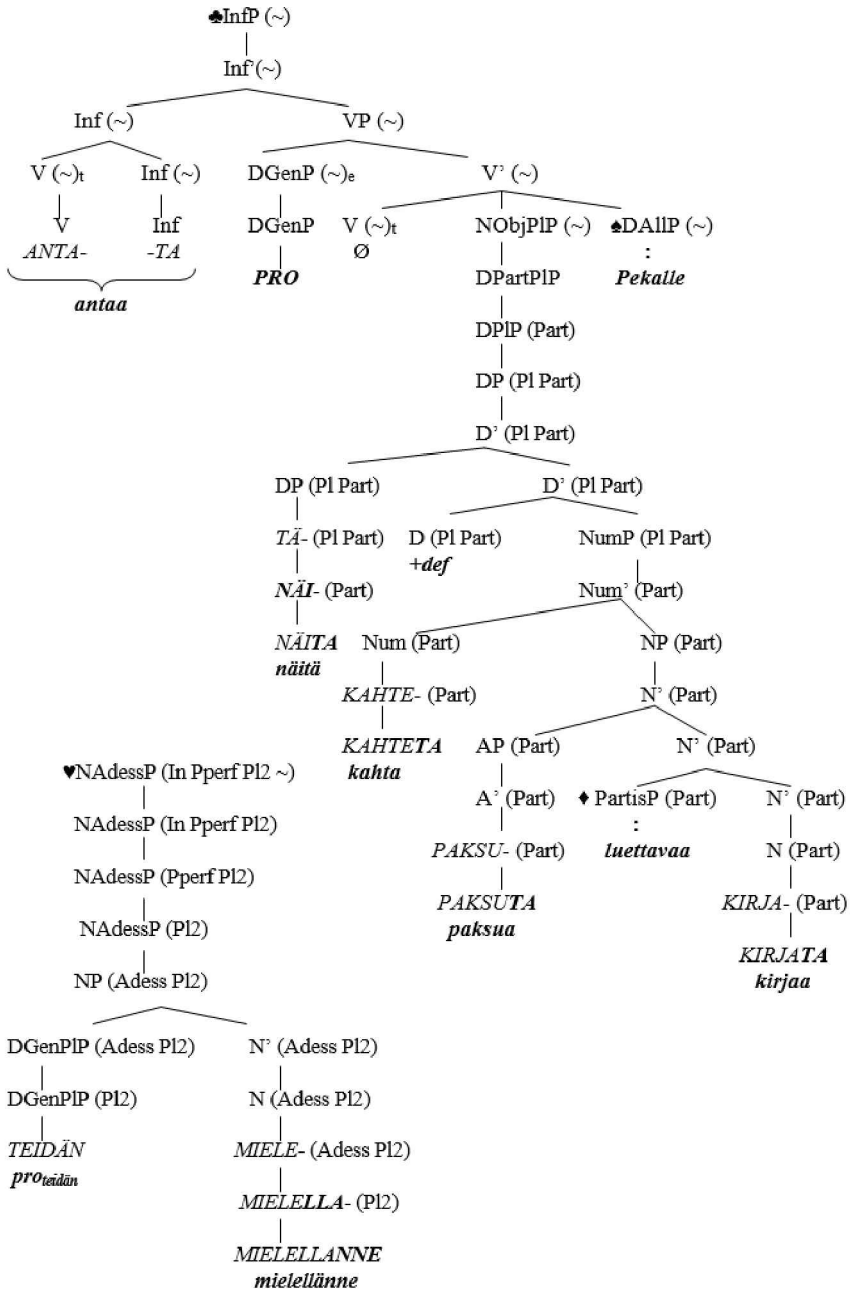
(12a) *Poika on komea.*
 boy is handsome
 'The boy is handsome.'

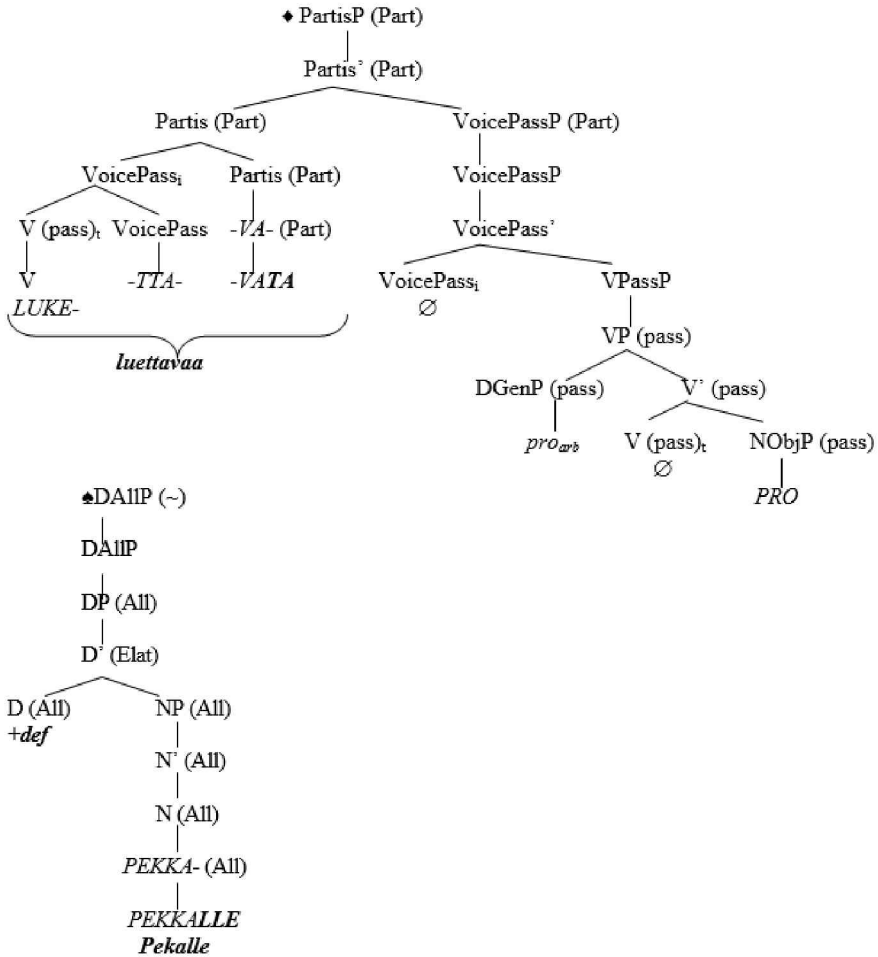
(12b) *Pekka on poika.*
 Peter is boy
 'Peter is a boy.'

⁴ Essentially the same analysis can be found in Holmberg *et al.* (1993) and Vainikka (1989, 1992), among others.

(13a) The syntactic structure of sentence (10)







(13b) Terminalization of certain indices which is not set forth in figure (10a) above

ettekö:

Neg (In Pperf Pl2 ~) → Neg (Pperf Pl2 ~) → E- (Pl2 ~) → ETTE (~) → ETTE } ETTE-KO → *ettekö*
 C (ek ~~topik~~ ~) → -KO (~~topik~~ ~) → -KO (~) → -KO

olleet:

INeg (In Pperf Pl2 ~) → INeg (Pperf Pl2 ~) → OLLUT (Pl2 ~) → OLLEET (~) → OLLEET → *olleet*

halunneet:

V (~) → V → HALU-

Voice (Pperf Pl2 ~) → -NUT (Pl2 ~) → -NEET (~) → -NEET

} HALU-NEET → *halunneet*

Indices used in this sentence:

Index ‘~’ appears in the representation of negative sentences. It is responsible for the partitive case of the object, see (16a). If a sentence contains a non-3rd person imperative verb, index ‘!’ (exclamation mark) is descending in the structure (14), which elicits a Nominativeaccusative case in the objects (for a 3rd person imperative verb, see (15)), just like index ‘arb’ in impersonal constructions (see (17)).

(14) *Luke-kaa tämä kirja!*
 read-IMP.PL2 this-NOM book-NOM
 ’Read this book!’

(14a) NObjP (!) → XNomaccP, where X=N, Num, D

(14b) *MINU-* (Nomacc) → *MINUT* → *minut*(14c) *KIRJA-* (Nomacc) → *KIRJA* → *kirja*

(14d) NObjP (!) → XPartP, where X=N, Num, D

(15) *Luke-koon tämä-n kirja-n!*
 read-IMP.SG3 this-ACC book-ACC
 ’(S)he (must) read this book!’

(15a) NObjP (13) → XAccP, where X=N, Num, D

(15b) NObjP (13) → XPartP, where X=N, Num, D

(16) *He ei-vät ol-leet luke-neet tä-tä kirja-a.*
 they-NOM no-PL3 be-PTC.PL read-PTC.PL this-PART book-PART
 ’They had not read this book.’

(16a) NObjP (~) → XPartP, where X=N, Num, D

(17) *Tämä kirja lue-ta-an.*
 this-NOM book-NOM read-PASS-ARB
 ’This book is read.’

(17a) NObjP (arb): NObjP (arb) → XNomaccP, where X=N, Num, D

(17b) NObjP (arb) → XPartP, where X=N, Num, D

Index ‘*topik*’ (“crossed-out topic”) is also an index that gets into the structure at the level of S, preventing a [Spec, TP] position from coming into existence in certain constructions. It appears in the representation of example (10), given in (13), for in-

stance, because the sentence in (10) contains a *pro* subject, and hence the Theme role can be played by no (explicit) constituent.

Index 'ek' is responsible for *-kO* particle peculiar to yes/no questions.

Index 'In' represents indicative mood, 'Pperf' refers to Pluskvamperfekti (Past Perfect) tense and 'Pl2' shows the number and person of the subject of the sentence.

'Part', 'Adess' and 'All' are case indices, 'Pl' shows the plural number of a noun. The latter index disappears if it meets a certain kind of numeral, like *kaksi* 'two' in example (13a); see the structure of *näitä kahta paksua kirjaa*.

Let us conclude this part with a short note on the fact that the relevant differences between the literary and the spoken Finnish can be captured by using an indexed grammar. One of the most interesting ones (at least from the indexed-grammar point of view) is that in the literary Finnish possessive suffixes can be found if the possessor is a pronoun, while in spoken Finnish there is no agreement between possessor and possession any more. This difference can be captured by leaving out an index, namely, the one telling the number and person of the possessor (compare (19a) to (19b)).

(18) Example of the expression of possessors in literary and spoken Finnish

literary language	spoken language
<i>kirja-ni</i>	<i>mu-n kirja</i>
book-SG1	I-GEN book
'my book'	'my book'

(19a) (*Puhu-n*) *minu-n kirjo-i-sta-ni*.
(speak-Sg1) I-GEN book-PL-ELAT-Sg1
'(I speak about) my book.'

(19b) (*Mä puhu-n*) *mu-n kirjo-i-sta*.
(I speak-Sg1) I-GEN book-PL-ELAT
'(I speak about) my book.'

4. Summary

This paper demonstrated how an indexed grammar can be applied in the description of the Finnish language. This mildly context sensitive grammar is the best way of dealing with agreement phenomena because it emphasizes the real nature of congruence with copying (and not moving) the relevant features, thus it makes the description more adequate. By using this type of grammar, many differences can be captured simply and spectacularly between the written and the spoken Finnish exactly via reference to certain indices. Unlike in written Finnish, for instance, there is no more congruence between possessor and possessee in this variant of the language.

REFERENCES

- Alberti Gábor 2006: *Matematika a természetes nyelvek leírásában II: Grammatikák és automaták*, Budapest, Tinta Könyvkiadó.
- Alberti, Gábor – Anna Medve 2000: Focus Constructions and the “Scope–Inversion Puzzle” in Hungarian, *Approaches to Hungarian 7*, Szeged, SZTE, 93–118.
- Alberti Gábor – Medve Anna 2002: *Generatív grammatikai gyakorlókönyv*, Budapest, Janus/Books.
- Bartos Huba 2000: Az magyar inflexiók jelenségek szintaktikai háttere, in Kiefer Ferenc szerk.: *Strukturális magyar nyelvtan 3: Morfológia*, Budapest, Akadémiai Kiadó, 653–762.
- Carstens, Vicki 2000: Remarks and Replies – Concord in Minimalist Theory, *Linguistic Inquiry* 31/2, 319–55.
- Farkas Judit 2005a: Új módszer az egyeztetési jelenségek kezelésére, in Sinkovics Balázs szerk.: *LingDok 4*, Szegedi Tudományegyetem Nyelvtudományi Doktori Iskola, 47–67.
- Farkas Judit 2005b: A finn igei szerkezet morfoszintaxisa, in Kubínyi Kata – Lauranto, Yrjö szerk.: *Urálisztikai tanulmányok 15*, Budapest, 43–52.
- Farkas Judit 2009: *A finn nyelv (indexelt) generatív szintaxisa*, PhD-disszertáció, Pécs, PTE. Kézirat.
- Holmberg, Anders – Nikanne, Urpo – Oraviita, Irmeli – Reime, Hannu – Trosterud, Trond 1993: The Structure of INFL and the Finite Clause in Finnish, in Anders Holmberg – Urpo Nikanne eds.: *Case and Other Functional Categories in Finnish Syntax*, Berlin – New York, Mouton de Gruyter, 177–206.
- Kornai András 1989: A főnévi csoport egyeztetése. *Általános Nyelvészeti Tanulmányok XVII*, 183–211.
- Manninen, Satu Helena 2003: *Small Phrase Layers. A study of Finnish Manner Adverbials* (Linguistik Aktuell/Linguistics Today, Volume 65), Amsterdam/Philadelphia, John Benjamins Publishing Company.
- Partee, Barbara H. – Alice G. B. ter Meulen – Robert E. Wall 1990: *Mathematical Methods in Linguistics*, Dordrecht, Kluwer Academic Publ.

- Vainikka, Anne 1989: *Deriving Syntactic Representations in Finnish*, Ph.D. dissertation, GLSA, University of Massachusetts, Amherst.
- Vainikka, Anne 1992: Kieliopillisten sijojen syntaktinen kuvaus. *Virittäjä* 1992, 295–322.

INDEXED GRAMMAR IN FINNISH SYNTAX

JUDIT FARKAS

The goal of this article is to demonstrate how an indexed grammar can be applied in the description of the Finnish language. This grammar – just like natural languages – belonging to the group of the mildly context sensitive grammars, is the best way of dealing with agreement phenomena: it emphasizes the real nature of congruence with copying (and not moving) the relevant features, thus it makes the description more adequate. Using an indexed grammar in the description of Finnish is extremely useful because in this language – especially in the literary style – there is a large variation of agreement phenomena: this kind of grammar emphasizes not only the essential features of certain structures, but we can capture many differences simply and spectacularly between the written and spoken Finnish just via reference to certain indices. For example, in spoken Finnish the opposition in *number* has neutralized between Sg3 and Pl3. We can also mention that the possessive suffixes have disappeared: i.e. unlike in written Finnish, there is no more congruence between possessor and possessed in this variant of the language.