

*Leelo Kuvallik*

University of Tartu  
Department of General Linguistics

Estonian in the Changing World

Edited by Haldur Õim

Tartu 1996

Acknowledgement  
We acknowledge the support of Eduard Treu's Minnesfond, Sweden,  
in publishing this volume

## Contents

Preface .....	5
Tiit Hennoste	
Tartu University Corpus of Written Estonian: A Survey of the Structure of Texts and Principles of Selection. ....	7
Mare Koit, Tiit Roosmaa	
Overview of the Uses of the Corpus of Estonian Literary Language: Current Possibilities .....	33
Heiki-Jaan Kaalep	
ESTMORF: A Morphological Analyzer for Estonian .....	43
Mare Koit	
Implementing a Dialogue Model on the Computer .....	99
Mare Koit, Tiit Roosmaa, Haldur Õim	
Teaching Computational Linguistics: One Vision. ....	115
Leelo Keevallik	
Maintenance of Structured Variability .....	123
Renate Pajusalu	
Regulative Utterances in Estonian Literary Dialogues and Radio Interviews .....	133
Silvi Tenjes	
Gestures in Dialogue .....	163
Haldur Õim	
The need for a Theory of Folk Theories in Cognitive Semantics: A Review and a Discussion .....	193
Haldur Õim	
Naive Theories and Communicative Competence: Reasoning in Communication .....	211

during the last decade. The goal of CL efforts is not just rise the efficiency of different kinds of language processing but through this to transcend language and cultural borders and secure the linguistic and cultural individuality of different nations.

### References

- Guide to the Theoretical Linguistics Programme 1994. Budapest University (ELTE). Budapest.
- Language Industries Atlas 1994. P. M. Hearn and D. F. Button (eds.). IOS Press.
- Language & Technology in Europe 2000 1994. Awareness Campaign. Seminar 10-11 November, 1994. Riga.
- Lingvistibulletiini 1994. Nro 1 (17.2.1994). Vastaava toimittaja F. Karlsson, Yleisen kielitieteen laitos. Helsingin yliopisto
- Magisterske Studium 1995. Studiini plan oboru INFORMATIKA.
- Studienführer 1991. Linguistische Datenverarbeitung/ Computerlinguistik für die deutschen Universitäten. Ermittlet und bearbeitet von M. Lutz-Hensel. GLDV Saarbrücken.
- Survey of Computational Linguistics Courses 1994. Second Edition. Compiled by B. J. Dorr. Association for Computational Linguistics.
- Towards a European Language Infrastructure 1992. Report by A. Danzin and the Strategic Planning Study Group for the Commission of the European Communities (DG XIII). 31 March 1992. Doc. No. 5210/92

## Maintenance of Structured Variability

Leelo Keevallik

University of Uppsala/University of Tartu

The central question of historical linguistics, why does a language change take place, has its obvious counterpart, why it sometimes does not? Why is a language or parts of it maintained under certain circumstances and not in others? What are the factors favouring maintenance? Language change and maintenance are of course merely two sides of the same coin, it is just that the latter has not been an equally attractive research subject for historical linguists. Challenging these traditional preferences, W. Labov has recently stated that "The long-term stability of many components of the linguistic system is even more striking than the rapid transformation of others" (1994: 42). This shift of interest has also been clear in the works of J. Milroy who argues that language change always takes place against the background of language maintenance and therefore the ignoring of stability is not to be justified. In his opinion, the reason behind such favouring is the possibility to construct sophisticated theories of linguistic change within the framework of the language itself, while the maintenance cannot be explained without reference to its social context.

If we aim to answer the questions about how language states can remain stable and how speech communities resist change, we have almost no alternative but to take account of social factors and speaker-based reasons (Milroy 1992:10-11). The same is hinted by Labov who promises to clarify long-term stability within the social matrix of language behaviour in

---

The author is indebted to the Swedish Institute whose scholarship made the peaceful writing of this article possible, and to Bengt Norberg for valuable advice.

his forthcoming book (Volume 2: Social Factors) in the volumes on "Principles of Linguistic Change" (1994: 42). One of the well-known puzzles in this area is the fact that non-standard dialects remain divergent from the mainstream norms of language despite the low status usually assigned to them and regardless of the wide exposure to the standard language via mass media. If it were just for the language system, it would be rather mystical, why would such "dysfunctional" variability be preserved altogether, while it is much more practical for everybody to speak one homogeneous standard language. But a human language is never perfectly stable or uniform and always involves geographical and/or social variability. Variability is not only a result of constant changes in the language system, and consequently a reflection of the different developmental stages, but divergent varieties can persist for generations and even for centuries, usually backed up by social needs and aims of the speakers and by specific prestige phenomena.

Furthermore, the concept of language variability does not merely cover the co-existence of several varieties, but involves also the existence of linguistic options inside a single variety and in every idiolect. Depending on the context, an American can say either *going* or *goin*, either pronounce the [r] in *car* or not, a Swede may say either *flickor* or *flicker* 'girls', and an Estonian may occasionally pronounce *läind* instead of *läinud* 'gone'. Options in the idiolect mirror variation on a wider scale in the language community.

Such freedom of choice crucially raises the question of language norms. In popular opinion, just the form(s) approved of in the standard variety are considered to be correct, i.e. in accordance with the norms of language. Other varieties are perceived as deviant, irregular and ungrammatical. A linguistic variant that deviates from the standard norm is thought to be wrong, no matter how often such a "mistake" is made. This kind of, mainly non-professional, approach to language unfortunately skips the consensus norms - norms that are agreed upon inside the speech community, without a straightforward verbal formulation which differentiates them from overtly prescriptive norms. These norms arise from the actual usage, and are usually acquired by inferring from this usage. They can even include the parameter of likelihood that guides the relative usage frequency of one or another linguistic variant in a certain context. Consensus norms are obviously more free to change than the prescriptive standard norms and the dynamics of this change has been the main field of sociolinguistic inquiry. On the other hand, a stable consensus norm is the core

of language maintenance. Variation governed by a stable consensus norm in the whole speech community is not a reflection of change in progress but pertains over a longer period.

Besides the social factors there may also be institutional factors - the official establishment of language norms - that enforce maintenance of variability, and purely linguistic ones as well. Both social and linguistic variables may form the system of variable constraints and thus be a part of a consensus norm. In the forthcoming, the structural, functional, and institutional support for maintaining variability will be exemplified with the help of a case of long-term variation in Estonian, the suffix of the active past participle.

The varying usage of the active past participle in Estonian dates at least from the 17th century, when in the Northern dialectal area the suffix *-nud* assumedly started to shorten into *-nd* (Raun & Saareste 1965: 66). The assumption is of course based on the early written records of Estonian produced mostly by foreigners and mainly limited to clerical genres. There is no way to find out what the actual usage was among common people, it undoubtedly varied by the end of the 17th century. The suffix was phonetically apt to shorten because it always appeared in a nonstressed noninitial position and because the vowel *u* stood between two homorganic sounds. The same is valid for the modern language. Furthermore, the shorter variant does not coincide with any other verbal suffix in the language and causes no misunderstandings. It is shorter, thus more economical. But *-nd* has its drawbacks - it is not as universal as *-nud* with regard to the preceding stem. *-nud* can appear after any segment whereas *-nd* has constraints. The phonotactic rules of Estonian do not allow long consonant clusters, especially not in word-final positions. According to a substantial statistical survey on standard Estonian syllable structure (47 250 words) there are no clusters of more than two qualitatively different consonants at the end of the word other than in monosyllables where their combinations are also limited and do not include *C+n+d* (Hint 1988:41, 46, 51-53). In colloquial Estonian the phonotactics may be different. For example, as far as the past participle is concerned, the cluster *lnd* may appear finally after multisyllabic stems. Monosyllabic form like *lasknd* 'let' and *andnd* 'given' are in fact sometimes produced but they are not comfortable for the native speaker. Therefore, if the verb stem happens to end in a consonant or a consonant cluster, then the suffix can undergo further shortening and lose the *n* (e.g. *lauinud*, *lauind*, *laud* 'sung'; *kakeinud*, *kakeind*, *kakeld* 'fought'; *naernud*, *naernd*, *naerd* 'laughed'; *tõusnud*, *tõusnd*, *tõust* 'risen'). In frequent forms

the last lateral of the stem may disappear instead (*olnud, olnd, ond* 'been'; *tulnud, tulnd, tund* 'come'). After an alveolar stop the suffix can lose both the *u* and *n* segments (*andnud, andnd, and* 'given'; *tahtnud, tahtnd, taht* 'wanted'; *võtnud, võtnd, võtt* 'taken'). In the cases of *andnd, tahtnd, and võtnd* it is impossible to pronounce the form otherwise than by making the last *n* syllabic which does not happen anywhere else in the language. The resulting forms are exceptional, non-natural, and clearly more difficult from the articulatory point of view than the standard forms - another argument against the articulatory ease that Kroch thought to be characteristic of non-prestige dialects (1978:20). Nevertheless, besides these difficult forms, much easier ones tend to develop out of them. If the stem ends in a single alveolar lenis *d*, the following forms may actually appear: *püüdnuud, püüdnud, püüüd* 'tried'; *leidnuud, leidnd, leind* 'found', most probably because of the assimilation of the first *d*. In stem-final consonant clusters some other segment may drop if the final consonant is not an alveolar stop (e.g. *lasknuud, lasknd, lasnd, last* 'let'; *jooksnuud, jooksnd, joost* 'run'). Again, there would be too many consonants in the end of the word and *n* would have to become syllabic. In short, although *-nd* is more economical and comfortable in most of the cases, there are structural constraints that disfavour its usage.

Regardless of the complexity of shortening process from the viewpoint of the language system, there is a clear tendency to make the past participle shorter in oral usage. Every single verb has a shorter form of the participle in addition to the current standard variant with *-nud*, which is also the only prescribed form in the written usage. The standard variant *-nud* is historically primary but the *-nd* forms have prevailed in many geographical dialects at the end of the 19th century, with *-nud* just occasionally popping up. Both definitely belong to the language competence of the native speakers. It may well be that many native Estonian children acquire the shorter variant first and the full suffix later.

The native speakers of Estonian are certainly not a homogeneous group. There are no clearly separable social varieties in the language especially because of the fluidity of the social structure in recent years, but substantially divergent geographical dialects were spoken until the beginning of the century and regional differences still exist. While the position of the standard language has been strengthening since the middle of the 19th century, there has been a continuously growing group of Estonians who are said to "speak the written language", i.e. the variety closest to the written standard as opposed to the regional dialects. This variety has been

chosen for the study of the active past participle variation. It should also be mentioned here that this is the variety where the standard past participle forms are used most frequently, possibly with the exception of the East-Estonian dialect area. Studies on other varieties are very likely to show a different amount of non-standard forms.

The informants for the present study were chosen from among highly educated speakers, or if younger, from families with highly educated parents. In order to achieve a homogeneous group of subjects, the preferable condition was the social immobility of the informant. It was possible to carry out this principle to a certain extent only as the oldest generation now is basically the first generation of Estonians who had easier access to the universities. They mostly came from the countryside farms, settled down in the towns, and supported the educational aspirations of their children. Thus the oldest group of educated Estonians could not be anything but upwardly mobile. Candid recordings of the chosen people were carried out in the years 1992-1994 in the two higher educational centers of Estonia, Tallinn and Tartu. Only informants who lived permanently in one or the other were picked. 57 persons were recorded altogether, both male (29) and female (28), all acquainted to the author. No recording situation was organized artificially, so the observer's paradox was reduced to the minimum. The informal situations were mostly conversations between friends in home settings, at cozy parties or picnics, the formal situations included diploma discussions at the department of Estonian in Tartu, radio interviews and anniversary meetings. The total length of the recordings was about 50 hours and the number of past participles in them was 3229. For analyzing the data the computer program Varbrul (1987, copyright S. Pintzuk) was used which basically calculates the significance of the correlation between the dependent variable and other variables, thereby helping to define variable constraints.

The total ratio of the non-standard endings was 32%. It is not surprising that educated speakers of standard Estonian produce *-nud* suffixes in the majority (68%) of the cases as this is the 'correct' variant. Rather, it is interesting whether there is any systematic correlation between the behaviour of the variable and other linguistic or social parameters.

As expected, a preceding voiceless consonant appeared to be a powerful inhibitor of shortening. In other words, the ending *-nd* is used significantly less often after a voiceless consonant than after a voiced sound, be it a vowel or a consonant (6% shorter forms after a voiceless consonant, 66% after a voiced consonant,  $p < 0.05$ , and 65% after a vowel).

Besides the preceding segment, there appeared to be another structural constraint, namely the number of syllables in the stem. It has been reported that the natural prosodic pattern of Estonian words consists of interchanging stressed and unstressed syllables, i.e. binary feet (Eek & Help 1986). With the development of the three-quantity system and many vowel losses in the history of Estonian the natural rhythm pattern was lost. In the contemporary language there are some morphological changes that indicate a return to the original pattern (Eek & Help 1986: 34-35). Also, the past participle ending acts as a means for achieving binary feet. There is a significant tendency ( $p < 0.01$ ) to produce words with an even number of syllables (Keevallik 1994: 102). So, for example, when the stem is monosyllabic or trisyllabic, it is more likely to be followed by the longer ending than are the bisyllabic stems or stems with four syllables. *-nud* forms an extra syllable whereas *-nd* does not (see above for rare exceptions). The aim to achieve binary feet rhythm patterns of words triggers the suffix to some extent.

The active past participle is an interesting case from the functional point of view as it fulfills a large array of syntactic functions. It can be the main constituent of a predicate in the indicative present and past perfect (*on käinud 'has gone', oli käinud 'had gone'*), in the imperfect negation (*ei käinud 'did not go'*), and in present tense as an adjective (*surnud mees 'a dead man'*). It can appear in the past of the conditional (*oleks käinud 'would have gone'*) and the modus obliquus (*olevat käinud 'has allegedly gone'*). The analysis of the data showed there to be considerable functional conditioning of the suffix.

First, the imperfect negation was significantly shorter than the other grammatical tenses (present perfect - 25% shortening, past perfect 32%, present 25%, imperfect - 49%; level of significance 0.000).

Second, the modus obliquus forms tended to have nonstandard suffixes at the level of significance of 0.003. They shortened in 44% of the cases as opposed to 31% in the indicative and 37% in the conditional mood. The paradigm of modus obliquus in Estonian is used to report what somebody else has told about some event or action in the past. The speaker does not take full responsibility for what he/she is saying. Past participles in the modus obliquus imply that the speaker is "telling a story" and are thus characteristic of a particular style, which may be one of the reasons for choosing the colloquial suffix. In contrast to the written standard, the auxiliary *olevat* can be omitted in speech, and very often is omitted, especially in folk-stories and fairy tales, even when in the written form. Besides, the indi-

cative mood and the modus obliquus can merge, so that the finer distinction is lost and both are expressed with the help of the simple past perfect (*oli käinud 'had (allegedly) gone'*) or, more rarely, with the present perfect (*on käinud 'has (allegedly) gone'*). On the other hand, there is a spreading option in the written standard, introduced by grammarians "from above", which has not reached oral usage yet (*olevat käinud = käinuvat*) (Uuspõld 1983: 65-67). The modus obliquus is the area where colloquial and written usage go considerably apart. In the oral data, the relatively smaller frequency of the standard suffix in this mood may reflect essentially the same colloquial/written opposition between the varying usage of *-nud* and *-nd*.

Besides, the past participles appear as auxiliaries (*ei olnud käinud 'had not gone'*), copulas (*on olnud hea 'has been good'*), and as members of complex verb phrases where they act mainly as modals (*ei saanud minna 'could not go'*). It turned out that there was some semantic conditioning of variability. When the form was only an auxiliary it tended to shorten considerably more than the main verbs (50% of the cases as opposed to 32% of shortening of main verbs). The same was true for the modals (49% of shortening as opposed to 31% of all the rest, level of significance  $p = 0.000$ ). Copulas behaved like main verbs, shortening in 31% of the cases. There was no meaningful difference between predicates, adjectives, and predicatives which are the rest of the functions that the past participle can fulfill. There were too few participles in subject positions to draw any conclusions about them. But as far as the semantic relevance of the form is concerned, it clearly contributes to the choice of the variant.

On the other hand, not only language internal functions should be taken into account but functionality on a wider scale as the functions of language are to a great extent social and pragmatic. It is not only in the configuration of the system that variability may be functional but also in the speaker-use language. Variability can be functional on the social level as a marker of the speaker's background, age or occupation, but it can also be an indicator of style. It is the absence of variation that would be dysfunctional (Weinreich, Labov, Herzog 1968). Labov in his New York City study (1966) has first shown that speakers normally attach great importance to this kind of variation and assign strong social values to what are essentially arbitrary differences. Variation within the structural parts of the language (such as phonology and morphology) is used by speakers to mark varying social roles. The active past participle ending in Estonian serves as a style marker among the speakers of the standard. *-nud* is used almost exclusively (93%) in formal situations whereas a more colloquial style

allows variation (only 60% of standard forms, level of significance  $p=0.000$ ). This result certainly reflects the impact of the written language norms on oral usage. Besides, in formal situations the informants tended to speak slower which is an additional factor favouring precise articulation and full forms. When at the quick speech rate the percentage of shorter forms was 47, then at the normal rate it was 28, and at the slow rate merely 8 (level of significance  $p=0.000$ ). The speech rate correlated with the factor of formality.

Last but not least, the institutional support is of importance when speaking about this particular case of variation. Official standardization aims at maintaining communicative efficiency, preciseness, and non-ambiguity. In such a perspective, variability in the standard language is redundant and disturbing. At the present moment, the only accepted variant in the written standard (*nud*) has secured its position even in the direct speech of characters in fictional prose (Kaugver, Unt, Traat). This kind of usage clearly does not mirror the actual situation. Therefore, time by time proposals are put forward to allow *nd* among other mostly morphological and semantic innovations that would bring the written standard closer to the speech usage (e.g. Kaplinski 1994). It is noteworthy, though, that nobody ever has demanded the exclusion of *nud* from the written language as such an appeal would obviously violate the reality in the opposite way.

But *nd* has not always been so completely rejected. It was accepted in the written variety first from 1693 till the 1870s. During this period it was widely used. In the 1870s the awareness of linguistic matters rose and a more normative approach to the mother tongue won popularity, bringing about aspirations for homogeneity and the rejection of variability. Naturally, the less universal variant *nd* had to retreat. Unfortunately there is no way to determine whether or how the consensus norms changed. For the second time *nd* was approved of by the influential linguist Aavik, first in 1916. It was on his initiative that *nd* was elevated as a more stylish and a better sounding variant in some phonological contexts (after a vowel, in long forms). Aavik's ideas were most eagerly praised among the younger generation of writers and poets who followed them in their works. Novelists felt free to use it at least in direct speech. *nd* disappeared again by the end of the 1930s under the pressure of normativism. Nowadays it is not even mentioned in grammars, which is definitely surprising with regard to the results of the present study. The evaluation of *nd* has therefore changed quite radically as it already did a couple of times in the history. Opinion polls show that the connotations of *nd* are negative at the

moment. Teachers of Estonian regard it as a mistake, a dialectal feature, something that "hurts the ear when heard from the radio or TV" (Keevallik 1994: 42-47). This is no wonder as *nd* has been deliberately excluded from the "correct" usage. It is very likely, however, that by doing so the standard has blocked a language change that under different circumstances might have led to the expulsion of *nud*.

Whatever the reasons, both variants are viable in contemporary Estonian and show no tendency of disappearing: there are no meaningful usage differences between different age-groups of educated speakers (Keevallik 1994:128-130). Besides, the comparison of the modern spoken data with the written usage from the middle of the 18th till the middle of the 19th century presents a good reason to believe that it has not changed too much. For this brief study 18 prose texts (newspaper articles, popular stories, handbooks) by 10 authors from the period were considered. When the present percentage of shortening is 32 then in these texts it was 25 (1155 out of 4671 forms), a divergence that may be ascribed to the different style and medium. Writers then almost never used *nd* after a consonant (in 3 cases out of 1139 forms with consonant stems (0%) as opposed to 307 out of 1155 (27%) at the present moment) which may indicate the spread of the shorter suffix variant from positions after vowels to positions after voiced consonants. Besides, the early writers used considerably less shorter suffixes after monosyllabic stems (7% as opposed to 33% at the present moment) and considerably more after bi- and trisyllabic stems (63% and 61%, respectively, as opposed to 34% and 23% at the present moment). That may well be a conscious choice by the writers not to use too short forms or to achieve certain rhythm patterns but it can also show the direction of historical development. Anyway, both suffix variants were available and used interchangeably without categorical linguistic constraints just like nowadays, proving that the variation has been maintained through centuries.

In the case of the active past participle in Estonian, the primary reasons for relatively stable variation are structural and institutional. The otherwise more comfortable shorter suffix *nd* has structural constraints on its use, the universal longer suffix *nud* is the norm in the grammars. Due to this the variation has become sensitive to context and style. The non-standard suffix is considered colloquial, the distribution of the two options varies according to the attention paid to speech (formality of the situation, speech tempo). The variability is securely embedded both in the language system and in pragmatics. It has been maintained for at least 3 centuries while the consensus on its evaluation has changed several times. The language sys-

tem does not necessarily aim at a state where one grammatical meaning is expressed by a single form.

### References

- Eck, A. & Help, T. 1986. Rütminihked eesti keele kujunemisloos. Tallinn, ENSV TA.
- Hint, M. 1988. Eesti ilukirjanduskeele statistiline fonotaktika. Silbistruktuuri elemendid. Tallinn, TPedI.
- Kaplinski, J. 1994. Minu eesti keel. - Sirp, nr 2, 14. jaanuar, lk. 5-6.
- Keevallik, L. 1994. <nud> varieerumine tänapäeva eesti kõnekeeles. Tartu. Magistritöö.
- Kroch, A. S. 1978. Toward a theory of social dialect variation. - *Language in Society*, Vol. 7, No. 1, pp. 17-36.
- Labov, W. 1966. *The Social Stratification of English in New York City*. Washington DC: Center for Applied Linguistics.
- Labov, W. 1994. *Principles of Linguistic Change. Volume 1: Internal Factors*. Blackwell. *Language in Society* 20.
- Milroy, J. 1992. *Linguistic Variation and Change*. Blackwell. *Language in Society* 19.
- Raun, A. & Saareste, A. 1965. *Introduction to Estonian Linguistics*. Wiesbaden.
- Uuspõld, E. 1983. Tänapäeva eesti kirjakeele grammatilise vormistiku mõningaist arenguist. In: *Kirjakeel 1983*. Tallinn: Valgus, pp. 62-70.
- Weinreich, U., Labov, W. And Herzog, M. 1968. Empirical foundations for a theory of language change. In: Lehmann, W. P. & Maliel, Y. (eds.), *Directions for Historical Linguistics*, Austin: University of Texas Press, pp. 95-189.

## Regulative utterances in Estonian literary dialogues and radio interviews

Renate Pajusalu  
University of Tartu/University of Turku

The present paper deals with one type of dialogue utterances - regulative utterances. By regulative utterances we mean utterances whose utmost purpose is not to convey new information about the topic mentioned but to influence the proceeding of the dialogue. Here belong various utterances indicating misunderstanding, explanations about one's earlier utterances, requests to explain something in more detail, instructions to the partner about his following utterances, etc. Regulative utterances occur to a different degree in different dialogue types. The paper treats regulative utterances occurring in literary dialogues and in radio interviews.

### 1. Introduction

One of the cornerstones underlying the classic treatment of dialogues is the principle of co-operation. The maxims of communication described in Grice (1975) have been elaborated and supplemented but their basic idea has remained the same: the participants of a dialogue together strive to a common final result and follow certain principles in that process. The ultimate goal is usually fulfilling a communicative task, commonly transferring some information from one partner to the other. One of the most laconic communication models claims communication to be a permanent interaction of two contrary principles, those of the R-principle and Q-principle. The R-principle is the principle of speaker economy and can be summarized as