



Repair-initiating particles and um-s in Estonian spontaneous speech

Tiit Hennoste

University of Helsinki, Finland & University of Tartu, Estonia

Abstract

Particles and um-s used in spontaneous Estonian speech as initiators of different types of repair are analysed. Our model and typology of repair based on conversation analysis is introduced. Three main types of repair and particles used to initiate those are described: prepositioned self-initiated self-repair, postpositioned self-initiated self-repair (addition, substitution, insertion and abandon), and other-initiated self-repair (reformulation, clarification and misunderstanding). In conclusion 6 groups of particles are brought out by the role they play in the initiation of the repair sequence. Data come from Corpus of Spoken Estonian of the University of Tartu, which contains everyday and institutional speech, telephone and face-to-face conversations.

1. Introduction

This article is a part of a larger project, which has two aims: to analyze Estonian conversation and to build a Dialogue System which gives information in Estonian and follows norms and rules of human-human communication. One part of the project is find out how the different communication problems are solved in Estonian spontaneous conversation (see also e.g. [3], [8]). We are interested in markers in actual speech, which are used by humans to interpret speech and which are usable as markers of speech forming processes in Dialogue System.

Problem solving is a process, which has been defined and described differently and in different terms (structure shifts, hesitation, communication strategies, repair etc (see e.g. [1], [2], [13])). We have adopted model and term of Conversation Analysis (CA) as basis for our analysis (see e.g. [9], [15], [16]).

By CA repair organization is a process, which is used to solve different communication problems (grammatical mistakes, incorrect word-selection, changing of the sentence plan, misunderstanding, non-hearing etc).

First, repair is divided into four varieties according to different role of the participants:

- 1) self-initiated self-repair: repair is initiated and carried out by the speaker of the trouble source;
- 2) other-initiated self-repair: repair is initiated by the recipient and carried out by the speaker of the trouble source;
- 3) self-initiated other-repair: repair is initiated by the speaker of the trouble source and carried out by the recipient;
- 4) other-initiated other-repair: repair is initiated and carried out by the recipient.

The most used types in conversation are self-initiated self-repair and other-initiated self-repair. Other types are very rare in native dialogues (but much more used in e.g. L2 conversations).

Second, repair is divided into prepositioned repair (also stalling, hesitation signals etc in different models) and postpositioned repair (also reformulation, retrospective repair, structure shifts etc). The prepositioned repair is used when speaker needs more time than he/she has to solve

his/her text-production problems. And so he/she postpones his/her next part of the text, using hesitation sequence. Postpositioned repair is used when some changes take place in the actual text.

The postpositioned repair could be divided by the positions where the repair can occur. There are two main positions. The first one is within the utterance containing the trouble source where self-initiated self-repair takes place. The second position is in the turn following the trouble source. This is a place where the other-initiated self-repair is mostly initiated.

The repair sequence has its beginning and end in speech. As repair is divagation from the "main line" of the conversation, it is natural, that the beginning of the repair must be marked.

The most used repair initiators are pauses, um-s, particles, repetitions (*ja ja* 'and and'), drawing of the last sound of the word (*mina:* 'I:'), interruption of the word (*põhi-* = *põhimõtteliselt* 'in principle'), intonational breaks. Longer phrases, sentences and nonverbal fillers (gestures, gaze etc) are used very rarely.

Um-s are vocalizations (*er, um* etc), which are used as hesitation pause-fillers in different languages.

Particles (also *D-items, discourse particles/markers, pragmatic particles/markers, inserts* etc) are defined differently in different models. In our model they are items (words or word forms) which have no denotative/lexical meaning and no syntactic relations with other structures in discourse/conversation (see [4]). Their use is defined by their pragmatical function. They may appear in their own in discourse (as *uh huh*) and may be attached prosodically to some larger structure. Some of them are homonyms of the other word classes (mainly adverbs or conjunctive words, e.g. *et* 'that' or *nagu* 'as', but also verb forms, e.g. *kule* 'hear').

In this article we will describe the particles and um-s, which are used to initiate repair and mark the beginning of the repair sequence in Estonian spontaneous speech. We will concentrate on three main types of repair: prepositioned and postpositioned self-initiated self-repair and other-initiated self-repair.

Our data come from Corpus of Spoken Estonian of the University of Tartu [6]. This corpus contains about 1 000 000 tokens. Subcorpus of 130 000 tokens is used in this study, both everyday and institutional speech, telephone and face-to-face conversations.

The CA transcription is used (see Transcription marks).

2. Prepositioned self-initiated self-repair (hesitation sequence)

The most used hesitation markers in our corpus are very short pauses (3200 items), drawings (2400) and um-s. There are about 2600 um-s in analyzed corpus. About 75% of them work as hesitation signals in Estonian spontaneous speech. The others are used to initiate postpositioned self-initiated self-repair (see 4.3).

The most frequent um is *ee* in spoken Estonian (38% of all um-s; example 1), the other more frequent items are *õõ*, *ää* and *mm*.

(1) *ee* (.) *ja kui `kaua* (.) *mi- `mis ajal peab puh noh enne kaitsmist ütleme ee mul on viie`teistkümmes millal ma pean selle bakalaureusetöö `esitama.*¹

'EE (.) *and how many time* (.) *in what time I must bring let's say EE my day is the eleventh when I must bring my BA thesis*²

There is one particle, which main function is to initiate prepositioned self-initiated self-repair in spoken Estonian: *noh* (untranslatable, example 2). This is also the most frequent particle in spoken Estonian (1056 tokens in our corpus).

(2) *C: ee tervist. ega `teie ei oska kogemata öelda Tallinna `busside aegu. (0.5)*
A: ja `milliseid busse. (.)
C: ee (.) *noh millal lähevad `viimased `kiirbussid täna Tallinna. (.)*
A: kell= `öheksa on viimane ekspress.

'C: ee hello. couldn't you tell me the departure times of the buses to Tallinn. (0.5)
A: and which buses. (.)
C: ee (.) *NOH when the last express buses go to Tallinn today. (.)*
A: at nine o'clock is the last express.'

The other particles are used more rarely as hesitation signals. The more used are *nagu* 'as' (example 3); *tähendab/tändab* 'it means'; *jah* 'yes'; *see* 'this'; *ku(r)at* 'devil'; *ütleme* 'let's say' (example 4) etc. The list of hesitational particles is open one, as there are lot of individual particles used by some or one person only.

(3) *C: mhmh* (.) *aga kuidas teil Itaaliaga on. kas n-* (0.5) *kas nagu on `sama olukord.*

'And how is with Italy. Is it NAGU (0.5) *is it the same situation.'*

(4) *jaanuari `kuus nädalaks `ajaks sis ütleme näiteks kui te võtaksite nüüd aparta`mendi sis apartamendi `hinnad `kõiguvad kuskil ütleme nii .hh `neljasaja: neljasaja `viie kümnest ja sealt `üles.*

'in january for a week then let's say for example if you could take an apartment then the prices of the apartments are about let's say four hundred fifty and up.'

Typical hesitation sequence (about 70%) consists of one hesitational unit (pause, drawling etc) in Estonian conversation. The other larger group is a combination of different means (25%), the most typical are:

- repetition of the conjunctive words + particle: *et et noh* 'that that NOH';
 - two different particles: *nagu noh* 'as NOH', *noh ütleme* 'NOH let's say';
 - pause + particle: (...) *noh* '(...) NOH';
 - pause + um: (...) *ee* '(...) EE'.
- The rest 5% are the repetitions of the same means (*noh noh* etc).

3. Postpositioned self-initiated self-repair (self-reformulation)

Self-reformulation is the repair, where speaker changes something in the his/her actual speech.

3.1. Model of self-reformulation

There are different models of the self-reformulation. Our model is based on the models of CA and of Enkvist & Björklund, who use Willem Levelt as basis (see [2], [13], [16]). Our model of the self-reformulation process is as follows:

1. The problem arises and the speaker decides to initiate reformulation.
2. He/she could begin immediately after the problem or with delay.
3. Typically some reformulation initiation marker is used at the beginning of the repair sequence (but it is possible to begin smoothly, without any marker).
4. Then the editing phase could follow sometimes, during which the speaker makes new construction. It is represented by hesitation sequence in the speech.
5. Then the reformulation (=new part of the text) follows.
6. Typically the repair sequence ends smoothly, but it is possible to mark it also explicitly by end marker.
7. There is a possibility to comment the process of the reformulation or the new construct by speaker him/herself or by hearer. Typically there is no any comments.

3.2. Typology of self-reformulation

There are different typologies of the self-reformulation. Our typology is based on CA (see e.g. [15]).

First, we divide reformulations into two groups. The basis is, whether the reformulation takes place in the same utterance (without intonational break) or not. There are two types of repair in the same utterance: substitution ja addition.

Substitution is a reformulation where some part of the utterance (typically word) or grammatical form of some part of the utterance is replaced with the new one (e.g. *vati* GENITIVE > *vatti* PARTITIVE in example 5). Addition is a reformulation, where a new item is added afterwards into utterance (*valget* in example 5). The most typical addition is adjectival attribute in Estonian.

(5) */---/ need olid siis iluduseks siis pandi vati* (0.5) *ee vatti valget vatti nagu oleks lumi sadand, (0.3)*

'/---/ those were as a decoration then was put the cotton [GENITIVE] (0.5) ee cotton [PARTITIVE] white cotton [PARTITIVE] like snow, (0.3)'

There are two additional reformulations which are used after intonational break: insertion and abandon.

Insertion is a reformulation, where speaker interrupts syntactical construction, adds a new one, finishes it and then finishes the interrupted utterance (example 6). Insertion itself is an information which is brought in after or before its "right place" in the text.

¹ C = caller; A= answerer/information officer in telephone conversation examples.

² The word-by-word translations are used in examples. The nontranslatable particles and um-s are written in upper case in translations (NOH).

(6) ja siis mul tuli täna noh s ma mõtsin = et sööks midagi kui teatrist ära lähen = ja (.) siis mul tuli meelde et ma pole (1.0) isegi (.) (muud) **a pühapäeval jah sõin midagi** = aga (.) aga et ma pole sel nädalal **täna on neljapäev** et ma pole sel nädalal absoluutselt mitte ühtegi = m soolast asja söönd.

'and then I remembered today that I want to eat something when I leave theater and (.) then I remembered that I haven't (1.0) even **oh on Sunday yes I eat something** but (.) but that I haven't eaten this week **today is Thursday** that I haven't eaten absolutely nothing solty this week.'

Abandon is a reformulation, when the utterance is interrupted and the new one is introduced (example 7).

(7) /---/ mm ma leian minu jaoks see siiski kontrolli kriteerium on (1.0) ma räägin kontrolli kriteerium on see. **ma olin tegelikult** kui me nüüd läheme tagasi Nikaraaguasse millest ma lootsin=et me tegelikult täna räägime.

'---/ mm I find it's a control criterion for me yet (1.0) I say it is a control criterion. **I was really** if we now go back to Nicaragua what I hoped we will talk about today.'

75% of self-reformulations are substitutions and 20% are additions in our corpus. The insertion and abandon are rare (about 5% together). Abandon is more used in radio interviews and insertion in longer narratives. Of course, different combinations of the types are used in conversation.

3.3. Self-reformulation initiation markers

The different markers are used to initiate different self-reformulation types.

The most used initiator of the substitution is interruption of the last word before the substitution sequence (typically 1-2 syllables are pronounced: *arva-* (example 8).

The substitution-initiating particles are:

- *või / või siis / või seal / või parem / või ühesõnaga* ('or / or then / or there / or better / or in one word'; example 9);
- *ei / mitte* ('no'; example 10);
- *tähendab / tähendab* ('it means'; example 11);
- *noh* (example 8);
- um-s.

The other hesitation signals are also used (pause, drawling; example 8).

(8) no ma nimetaksin siis kohe nimodi kägupealt kahte assja, mida **ma:** > (0.5) **noh** mis on siis selles mõttes üllatus, (0.5) et ma **arva-** arvasin, et: ma lahkun: (0.5) pigem negatiivsete emotsioonidega, (a)ga ma lahkusin väga positiivsetega. >

'then I will mention two things at once, what **I:** (0.5) **NOH** that are surprises in this sense (0.5) that I **thou-** thought that: I will leave (0.5) with negative emotions, but I leaved with very positive ones.'

(9) /---/ ja:: ja=siis (...) mm kirik oli **või** tol korral oli väga külm ilm olnud (.) kirik ka väga külm.

'---/ and:: and=then (...) mm church was **or** it was very cold weather this day and church also very cold.'

(10) .hh no sis ma olen nüüd vahepeal jalutand tast kaugemale ku=ma lähen **ei** kui teil vaja on /---/

'hh then I have now walked far from him if I go **no** if you need /---/'

(11) sest mina (.) **tähendab** ma tahtsin seda sulle rääkida=et (.) mul tuli täna teatris meelde täna ma ei ole ka loomulikult midagi söönd.

'because I [long form of I] (.) **it means** I [short form of I] wanted to tell you that (.) I remembered in theatre today that I haven't eaten nothing today too.'

There are only some markers to initiate addition: interruption of the word, pause and um-s (example 12). No particles are used in our corpus.

(12) veel võiks olla nii et ee (1.0) et ülejäänud **ra-** väheses rasvas /---/

'furthermore it could be so that ee (1.0) that in the rest of **gre-** few grease /---/ '

The marker of abandon is intonational break. Sometimes hesitational markers are used at the beginning of the new utterance (pause (example 13), *noh* and um-s).

(13) aga enne seda kui nad Väljaotsale läksid? (.) ma=tahan= veel=rääkida kuidas Lutsu (.) Palamuse kirikus ristiti (...)

'and before they went to Väljaotsa? (.) I want to tell you how Luts was baptised in Palamuse church'

The marker of the insertion is intonational schift, and sometimes hesitational signals (pause, drawling and um-s) are used in addition.

Some particles and conjunctive words are used at the beginning of the insert sequence, but their role is not to initiate insertion but to show the semantical or/and pragmatical connection between insertion and main utterance. The most used particle is *no* (which is mainly the used to introduce a new (sub)topic in Estonian conversation, example 14; see [4]).

(14) see Oskar Lutsu ema tädi (.) **õ no**=mälestuste raamatus nimetati seda ka Murumunaks see=sis andis /---/.

'this aunt of mother of Oskar Luts (.) **õ NO** in the memoirs she was called also Murumuna she gave /---/'

4. Other-initiated self-repair (next-turn repair)

Other-initiated self-repair is initiated by hearer, who have found some problem in the speakers previous text. We call it next-turn repair, as 90% of those repairs are initiated immediately after the problematic turn in our corpus.

4.1. Typology of next-turn repair initiations

The typology of the other-initiated self-repair is also different in different problem solving models. We have divided those repairs into three groups: clarification, reformulation and misunderstanding (see also [3]).

Clarification is an initiation, by which the hearer repeats exactly or with some variation some utterance, phrase etc of the previous speaker to get confirmation that it was such (did you say that?).

Reformulation (also candidate understanding in CA) is an initiation, by which hearer gives his/her own interpretation (hypothesis, rewording, generalisation etc) to the speakers turn. His/her aim is to get confirmation, that his/her understanding is true (did you think that?)

Misunderstanding is an initiation, by which initiator reports that he/she did not hear or did not understand the previous information, or the information contradicted so much with his/her knowledge and beliefs that it must be checked.

There are two subtypes of misunderstanding:

- 1) the speaker only indicates that there was a problem;
- 2) he/she localises the problem more exactly.

About 50% of repair initiations are clarifications, 25% reformulations and 25% misunderstandings in our corpus.

4.2. Next-turn repair initiation markers

The main initiation means of the next turn repair are question phrases and sentences.

There are five types of questions in our dialogue act typology: wh-question, open yes/no question, closed yes/no question, question that offers answer and alternative question (see [8]).

Open and closed yes/no questions have similar grammatical form but they expect different answers. A closed question expects the answer yes or no (*Are you open in winter? – Yes.*) while an open question expects giving information (e.g. the question *Is there a bus that arrives to Tallinn after 8?* the client intends to learn the departure times of buses).

Closed yes/no questions and questions that offer answer are both questions that expect yes/no answer. Their difference lies in the presuppositions of the user. Asking a question that offers answer the speaker has an explicit opinion, hypothesis, and he/she is expecting a confirmation by the partner. No such presupposition exist in case of a closed yes/no question.

Clarification and reformulation expect yes/no answer (=closed yes/no question or question that offers answer), misunderstanding expects information (=open yes/no question, wh-question or alternative question).

There is a difference between question particles used in clarification and reformulation.

Clarification is mostly (55% of initiations) expressed by interrogative intonation only. The most frequent particle is *jah* 'yes' pronounced with interrogative intonation at the end of the utterance (20%; example 15).

(15) A: *ee=ütleme et=e kella kümnest=ee=h* (.) *neljani on vaba.* (1.0)

C: *õõ siis paneks* (2.0) *kas kell kaks=või või pool kolm=või.*

A: *õõ=hh* (.) *pool= kolm jah?*

C: *jah.* (.)

A: *ee ja kuidas nimi on.*

C: *Saabas.* (.)

A: *Saabas jah.*

C: *jah*

'A: *ee=let's say that=e from ten o'clock=ee=h* (.) *till four is free.* (1.0)

C: *õõ then [we] would take* (2.0) *at two o'clock=or or a half past two or.*

A: *õõ=hh* (.) *half past two=yes?*

C: *yes.* (.)

A: *ee and what is name.*

C: *Saabas.* (.)

A: *Saabas yes.*

C: *yes* '

The second most used particle is *või/vä* (7%, example 16). The other particles are used rarely (*ühesõnaga* 'in one word', *et* 'that', *siis* 'then', *kas* 'question particle of yes/no-question in Standard Estonian').

(16) A: *ega ta üksi- kahekesi ei olnd, nad olid neljaneljakkeisi.*

C: *neljakesi = vä.*

A: *mhmh? /---/*

'A: *she was not alone, there were four persons.*

C: *four-VÄ.*

A: *yes.*'

28% of reformulations are initiated using questions marked by interrogative intonation only. The main particle used here is *et* (22%) at the beginning of the utterance (example 17).

(17) C: *ee milliseid teil pakkuda on.*

A: *\$ e:i pakkuda ei ole meil ammu enam midagi.* \$

C: *[aa hehe]*

A: *[ma= mõtlesin et te] olete üks \$ reisijatest.* \$

C: *aa, et teil on kõik välja*

A: *oi loomu likult.*

'C: *ee what do you offer.*

A: *\$ no: we have nothing for a long time.* \$

C: *[aa hehe]*

A: *[I thought that you] are one of the passengers.*

C: *oh, so you have sold out all*

A: *yes of course.*'

The second most used particle is *jah* 'yes' at the end of the utterance pronounced with interrogative intonation (11%). The other particles are used less than 10%: *nii et* 'so that', *siis/sis* 'then', *kas*, *või/vä*, *tähendab* 'it means', *eks* 'isn't it'.

Misunderstanding is divided into two subgroups, according to the exactness of the localisation of the problem.

1. General initiations are used only to mark, that there was a problem in the preceding turn (example 18). Most of those initiations are formed by general question particles and question words only (*jah* 'really?', *ah* 'what?', *kuidas* 'what?', *mis / mida / millega* etc 'what?').³ Sentences are used rarely (e.g. *mis sa ütlesid* 'what did you say?').

(18) K: *räägi kuda siis on.* (.)

M: *millega.*

K: *ülemusega sõit läks.* (1.5)

M: *no mis ta on, mis seal minna oli.*

K: *ah?*

M: *mis seal minna oli.*

'K: *Tell now how it was.* (.)

M: *What.*

K: *Your trip with boss.*

M: *nothing interestong, normal.*

K: *What.*

M: *Normal.*'

³ *mis* is a pronoun and a question word in Estonian, which is declinable.

2. More exact localisation of the problem is made by repetition of the problematic part of the turn (sentence, phrase, word) using question words that localize the problem (*kus?* 'where?', *kes?* 'who?' etc). This subtype is formulated almost all by wh-questions in our corpus.

The most frequent question word used in misunderstanding is *kuidas* (39%), which is usable with different intonation to indicate non-hearing or surprise ('what?'; example 18, 19) and to specify some element of the utterance ('how?'; example 20).

(18) A: 'Estmar= 'info, 'Leenu=kuuleb tere

C: tere. (0.8) {Leenu.}

A: jah? (0.5)

C: rotilõks. (1.8)

A: **kuidas?**

C: rotilõks. (0.8)

A: jah, rotilõks

C: { 'andke 'kõik.}

A: **kuidas?**

C: { 'kõik kus ma saan 'osta.}

'A: 'Estmar= 'info, 'Leenu=is hearing, hello

C: hello. (0.8) {Leenu.}

A: yes? (0.5)

C: rat trap. (1.8)

A: **what?**

C: rat trap. (0.8)

A: yes, rat trap

C: {give me all}

A: **what?**

C: {all [shops] where I could buy.}'

(19) C: tere=ma palun (.) 'Teksako bensiini jaama. (...)

A: neli kaheksa kaks, üks kolm üks.

C: oi ma ei 'kuule, **kuidas** palun.

A: NELI KAHEKSA kaks, (.) null kolm null.

'C: hello=would you give me (.) 'Teksako oil station. (...)

A: four eight two, one three one.

C: I don't hear, **what** please

A: FOUR EIGHT two, (.) one three one.'

(20) C: õelge (.) kus on Tartus ee 'Kaa sa 'long. (1.5)

A: e= **kuidas**=se sa 'long oli

C: 'Kaa. 'Kaa salong.

C: tell me (.) where Kaa salon is located in Tartu. (1.5)

A: e=**how** this salon was

C: 'Kaa. Kaa salon

5. Discussion

We can divide particles and um-s into two groups by the role they play in the initiation of the repair sequence in spontaneous Estonian speech.

The first group is particles and um-s used to initiate self-initiated-self-repair:

1) hesitation signals: um-s (*ee, õõ, aa, mm* etc), particles *noh, nagu, tähendab, ütleme, ku(r)at, jah* etc;

2) substitution initiators: um-s (*ee, õõ, aa, mm* etc), *noh, või (+siis/seal/parem* etc), *tähendab, ei/mitte* etc.

3) addition initiators: um-s (*ee, õõ, aa, mm* etc)

First, different particles are used differently in spoken Estonian.

um-s are universal self-initiated repair initiators and items which only function is to initiate self-repair in analyzed corpus. They are used in the hesitation sequence, and as initiators of the substitution and addition. They are also used at the beginning of the abandon and insertion sequences, but we have interpreted them as hesitation signals in those positions.

The most frequent repair particle is *noh*. It is unclear, whether *noh* is also universal self-repair initiator or not. There is no *noh* at the beginning of the addition and abandon sequences in analyzed corpus.

noh has also other functions in conversation. It could be a part of the editing phase of reformulation process, where it mostly signals that speaker has found the continuation (see 4.1). And *noh* is used at the beginning of the turn as particle, which signals the pragmatical connections between following and preceding utterances/turns (see [5]).

The other particles are used as particles only in some contexts, and some of them are also homonyms of the other word classes. E.g. *nagu* 'as' is also a softener and adverb; *tähendab/tändab* 'it means' is mostly an initiator of postpositioned repair (see 3.3), but it is also used as the initiator of the formulations, accounts, explanations etc; *jah* 'yes' is mainly response particle (answer to the yes/no question etc) and next-turn repair initiator (see 4.2); *see* 'this' is mainly pronoun and sometimes definite article, *ku(r)at* 'devil' is a swear-word (see also [4], [5], [12]).

The other particles are used typically in different functions at the same time. E.g. *ütleme* 'let's say' initiates example or proposal or works as hedge in Estonian dialogue, but it is almost always a part of hesitation sequence at the same time (see [7], [12]).

Second, there are differences between self-repair varieties.

Particles and um-s were not used to initiate the abandon and insertion in our corpus. Of course, there are sometimes hesitation sequences at the beginning of the abandon and insertion. Particles and conjunctive words at the beginning of the insertion are used to show semantical or pragmatical connection between insertion and main utterance (*no* etc).

There were only um-s and not particles used at the beginning of the addition in analyzed corpus.

Both particles and um-s are used to initiate hesitation and substitution. Some of the particles are used only at the beginning of the substitution (*või+siis/seal/parem* etc, *ei/mitte*), some are used also as hesitation signals (*noh, tähendab* 'it means'). It is possible to interpret the last ones as particles, that have two functions at the beginning of the substitution.

The second group is particles used by the hearer to initiate next turn repair:

1) particles and question words used to initiate clarification and reformulation: *jah* 'yes', *eks* 'isn't it', *et* 'that', *või/vä, siis* 'then', *ühesõnaga* 'in one word', *tähendab* 'it means', *kas, nii et* 'so that';

2) particles and question words used to mark general misunderstanding: *jah* 'really?', *ah* 'what?', *kuidas* 'what?', *mis* etc 'what?';

3) wh-question words used to localize problem more exactly: *kus?* 'where?', *kes?* 'who?' etc.

There are two groups of words used in the initiation of the next-turn repair. The first is question words, which are used mostly in Standard Estonian and/or in "main line" questions (wh-question words, *kas, või/vä*).

kas is a question word of yes/no-question in Standard Estonian. *või/vä* is a yes/no-question word in spoken Estonian, developed from the conjunction *või* 'or' (see [14]).

The second group is question particles (*jah, eks, et, nii et, siis, ühesõnaga, tähendab, ah*).

et ('that') and *nii et* ('so that') are conjunctive words in Standard and spoken Estonian, and also particles ('so') in spoken Estonian. Their main particle function is to mark that the following utterance is a summary or conclusion from the previous text or that it is attributed to some other person than speaker (see [9]).

jah has several different functions. It's main function is response particle (answer to the yes/no question etc; see [4]). In this function it is used mostly separately or at the beginning of the utterance, rarely also in the middle of the utterance. As a question particle it is used at the end of the phrase or sentence in clarification and reformulation, and separately with interrogative intonation to initiate general misunderstanding. And as we have seen, it may be sometimes a hesitation signal.

siis 'then' is mainly an adverb and also particle with different functions in spontaneous Estonian (see [10]).

Eks is a tag question word.

Tähendab 'it means' is also an initiator of postpositioned repair and of the formulations, accounts, explanations etc.

There are some particles which are used in clarification and reformulation (*jah, et, siis, kas*). Some particles are used in reformulation (*eks, tähendab, nii et*), and some in clarification only (*või/vä, ühesõnaga*) in analyzed corpus. It is not clear, whether all those particles are usable in both repair or not.

There is a difference between the formulation of the repair initiating questions and "main line" questions in the dialogue (see [8]). Our analysis has shown that clarification and reformulation are formed by a question that offers answer in almost all cases (except one reformulation formed by wh-question in analyzed corpus), and majority of questions that offer answer (81%) are used for repair initiation.

The "main line" yes/no-questions are mostly closed or open yes/no-questions which are initiated mostly by particles *kas* (55-59%), *või/vä* (9-11%) or *kas+või/vä* (6-7%). The second important difference is that the "main line" questions are mostly formed as full sentences while repair initiations are mostly phrases (or single words). 79% of clarifications are phrases or words. Reformulation and misunderstanding is formed by a full sentence in half of cases.

6. Transcription marks

falling intonation	period.
fall not to low	comma,
raising intonation	question mark?
short interval (max 0.2 sec)	(.)
timed interval	(2.0)
nontimed longer interval	(...)
latching at end of utterance	word=
latching at beginning	=word
drawling	::
stress	` at the beginning of the stressed syllable
cut off	do-
inbreath	.hh
begin of overlap	[
end of overlap]
loud sounds	UPPER CASE
{transcriber's inability to hear what was said}	
smile	hehe

7. References

- [1] Dörnyei, Zoltan & Mary Lee Scott. 1997. Communication Strategies in a Second Language: Definitions and Taxonomies. *Language Learning*, vol. 47, pp. 173–210.
- [2] Enkvist, Nils Erik & Martina Björklund. 1985. *Toward a taxonomy of structure shifts*. MS.
- [3] Gerassimenko, Olga, Tiit Hennoste, Mare Koit & Andriela Rääbis. 2004. Other-initiated Self-repairs in Estonian Information Dialogues: Solving Communication Problems in Cooperation. *Proceedings of the 5th SIGdial Workshop on Discourse and Dialogue*, Cambridge, April 30–May 1, 2004. Ed by M. Strube, C. Sidner. Cambridge, pp. 39–42.
- [4] Hennoste, Tiit. 2000. Sissejuhatus suulisesse eesti keelde IV. Suulise kõne erisõnavara 3. Partiklid [Introduction to Spoken Estonian IV. Particles]. *Akadeemia*, No. 8, pp. 1773–1806.
- [5] Hennoste, Tiit. 2001. Sissejuhatus suulisesse eesti keelde IX. Lausung suulises kõnes 4 [Introduction to Spoken Estonian IX. Utterance in spoken Estonian 4]. *Akadeemia*, No. 1, pp. 179–206.
- [6] Hennoste, Tiit. 2003. Suulise eesti keele uurimine: korpus [Studying Spoken Estonian: Corpus]. *Keel ja Kirjandus*, No. 7, pp. 481–500.
- [7] Hennoste, Tiit. 2004. *Et*-komplementause kesksete põhiverbide funktsioonid eestikeelses vestluses. [Main functions of complement verbs in Estonian conversation]. *Keel ja Kirjandus*, No. 7, pp. 504–523
- [8] Hennoste, Tiit, Olga Gerassimenko, Riina Kasterpalu, Mare Koit, Andriela Rääbis, Krista Strandson & Maret Valdisoo. 2005. Questions in Estonian Information Dialogues: Form and Functions. *Text, Speech and Dialogue. Proceedings of the 8th International Conference, TSD 2005*. ???+Berlin: Springer. In appear.
- [9] Hutchby, Ian & Robin Wooffitt. 1998. *Conversation Analysis. Principles, Practices and Applications*. London: Polity Press.
- [10] Jansons, Airi. 2002. Partikli *siis* funktsioonid suulises kõnes. [The functions of particle *siis* 'then' in Estonian conversation]. *Keel ja Kirjandus*, No. 9, pp. 612–629.
- [11] Keevallik, Leelo. 2000. Keelendid *et* ja *nii et* vestluses [Tokens *et* and *nii et* in Estonian conversation]. *Keel ja Kirjandus*, No. 5, pp. 344–358.
- [12] Keevallik, Leelo. 2003. *From Interaction to Grammar. Estonian Finite Verb Forms in Conversation*. *Studia Uralica Upsaliensia* 34. Uppsala: University of Uppsala.
- [13] Levelt, Willem J. M. 1983. Monitoring and Self-repair in Speech. *Cognition*, vol. 14, pp. 41–104.
- [14] Lindström, Liina. 2001. Grammaticalization of *või/vä* Questions in Estonian. *Papers in Estonian Cognitive Linguistics*. Ed. by Ilona Trigel. Tartu: University of Tartu, pp. 90–118.
- [15] Schegloff, Emanuel. 1979. The Relevance of Repair to Syntax-for-conversation. *Syntax and Semantics, Volume 12: Discourse and Syntax*. Ed. by T. Givon. N.Y.: Academic Press, pp. 261–288.
- [16] Schegloff, Emanuel, Gail Jefferson & Harvey Sacks. 1977. The Preference for Self-correction in the Organization of Repair in Conversation. *Language*, vol. 52(2), pp. 361–382.