





(0.8)

V: ἰ ἰσθμῶν ἰσθμῶν. | DIJ: Ἰσθμῶν ἰσθμῶν |

(5.5)

V: ὁ ἰσθμῶν ἰσθμῶν ἰσθμῶν [ὁ ἰσθμῶν,] ἰσθμῶν? | KYE: Ἰσθμῶν ἰσθμῶν | | VTE: Ὀὐκ ἰσθμῶν ἰσθμῶν ἰσθμῶν |

H: [ἰσθμῶν.] | KYJ: Ἰσθμῶν | | VTJ: Ὀὐκ ἰσθμῶν ἰσθμῶν ἰσθμῶν |

V: ἰ ἰσθμῶν ἰσθμῶν ἰσθμῶν ἰσθμῶν ἰσθμῶν < ἰσθμῶν ἰσθμῶν ἰσθμῶν ἰσθμῶν? (0.5) ἰσθμῶν ἰσθμῶν ἰσθμῶν ἰσθμῶν. > | DIJ: Ἰσθμῶν ἰσθμῶν ἰσθμῶν |

H: [ἰσθμῶν.] | YA: Ἰσθμῶν |

(.)

H: ἰσθμῶν ἰσθμῶν. | RIE: Ἰσθμῶν ἰσθμῶν |

V: ἰσθμῶν ἰσθμῶν ἰσθμῶν? | RIJ: Ἰσθμῶν ἰσθμῶν ἰσθμῶν |

### 2.3. Êíííóíééàòèáíúá ñòðàòáãèè

Ìú èñòíáèì èç ìííýòèý êíííóíééàòèáííé ñòðàòáãèè è êíííòðóèèèáííé ìíááèè àèèííá (Constructive Dialogue Model, CDM) [Jokinen ó-àñòíéèì àèèííá àèý ìííòðíáíéý ñéááòpúáé ðáíéèèè èáé ðááèèèè íà ìðááúáòúòp ðáíéèéó ìàðòíáðà. Êíííóíééàòèáíúá ñòðàòáãèè íáíáúáííì óðíáíá.

Äèý ìðáááéáíéý êíííóíééàòèáíúá ñòðàòáãèè á CDM èñííèüçòpòñý -áòúðá èííòáèñòáèüíúó òàèòíðà:

&ndash; íæèääáííñòü ðáíéèèè;

&ndash; ñáyçü ðáíéèèè ñ òáííé;

&ndash; áíñèèáíóòíñòü òáèéé áíáíðýùááí;

&ndash; èíèèèèèèè áíáíðýùááí èèè ìàðòíáðà.

Áñá èííòáèñòáèüíúá òàèòíðú áéíáðíú, ñííòááòñòááííí, êíííóíééàòèáíúá ñòðàòáãèèè  $2*4=16$ .

Íáéáíéáá óáá-ííé ñòðóèèòóðíé ìðááñòááéáíéý êíííóíééàòèáíúá òáèéé ýáèýáòñý ìááàçéí òáèéé (stock) è ìðéíòèì «ìíñéááíèì áíøáè, ìíñéááíáàòáèüíì ìáíá ìáá áðóáíé. Íí ìáðá áíñòéæáíéý ááðóíèò òáèéé ííé óááèýpòñý; áñá òáèè àèèííá ñ-èòáòpòñý áíñòéáíóòíè,

Êíííóíééàòèáíúá ñòðàòáãèèè ðàçíá-áíú á 20 áñòáñòááííúó è 20 ñèìóèýòèííúó àèèííáð.





ᐃàçíá-áííúé éíᐃíõñ ẽñĩĩẽúçóáòñý á òáíᐃáòè-áñéèò è ìᐃééááíúò ẽññéááíááíéýò (íáíᐃ., [Hennoste et al. 2005]) è á ᐃááíòá íáá á [Fishel 2004], à òàèæá á ïĩñòᐃíáíèè íáᐃáúò àèèíáíáúò ñèñòáí íá ýñòííñèí ýçúéá. Ýòè ñèñòáíú, Òᐃáíñĩᐃòíúé è Òáàòᐃáéúíúé Á ìᐃèíáíáíèáí çíáíèé í ñòᐃóéòóᐃá éíóíᐃíàòèííúò àèèíáíá è ïíáèèéíáíá.

Nĩèñíè èèòáᐃáòóᐃú

Mark Fishel 2005. Dialogue Act Recognition in Estonian Dialogues Using Artificial Neural Networks. In: Proceedings of the Second Baltic Conference on Human Language Technologies, Tallinn, 4&ndash;5 April 2005, 249&ndash;254.

Olga Gerassimenko, Tiit Hennoste, Mare Koit, Andriela Rääbis, Krista Strandson, Maret Valdisoo, Evely Vutt. Annotated Dialogue Corpus as a Language Resource: An Experience of Building the Estonian Dialogue Corpus. The First Baltic Conference &ldquo;Human Language Technologies. The Baltic Perspective&rdquo;. Commission of the Official Language at the Chancellery of the President of Latvia, Riga, 2004, 150&ndash;155.

Tiit Hennoste, Olga Gerassimenko, Riina Kasterpalu, Mare Koit, Andriela Rääbis, Krista Strandson, Maret Valdisoo. Questions in Estonian Information Dialogues: Form and Functions. Text, Speech and Dialogue. 6th International Conference TSD 2005. Springer, 2005, 420&ndash;427.

Gail Jefferson 2004. Glossary of transcript symbols with an introduction. In Lerner, G.H. (Ed). Conversation Analysis: Studies from the first generation. Amsterdam/Philadelphia: John Benjamins, 13&ndash;31.

Kristiina Jokinen. 1995. Rationality in Constructive Dialogue Management; URL: <http://cl.aist-nara.ac.jp/lab/papers/kris/aaai.ps> (used 20.05.2006).

Áèáíáàᐃííñè

ᐃááíòò ïíáááᐃæèáááò Ýñòííñèèé íáó-íúé òíá (ãᐃáíò 5685).

Summary

The Estonian Dialogue Corpus is collected with the aim of developing the dialogue system using the natural language. The spoken dialogues (884 dialogues, 155000 running words) are used to study the rules and norms of the human-human communication; the corpus also includes human-computer dialogues (21, 2500 running words) collected by the Wizard of Oz method used to study the role behaviour of the users and information provider. The presentation considers the means and levels of transcription and annotation dialogues and also the application of the corpus.

Īðēēīæāíēā

1. Ōðāīñēðēīōēīīūā çíàèè

ñīāā èíōííàöèè .

īīēōñīāā èíōííàöèè ,

īīāúāī èíōííàöèè ?

ēīðīòēāÿ īàóçà (īàēñ. 0.2 ñ.) (.)

äèèīā īàóçú ā ñāēóíāàō (2.0)

íäēīæāíēā [text]

ñèèÿíēā íaçāâēñēīūō ääèíö =

ðāñòÿíóòúé çāóê ::



óààðĩĩá ñeĩáĩ

`

ĩðáðááĩĩá ñeĩáĩ

do-

ááĩõ

.hhh

óáúñòðááíeá òáĩĩá

> text <

çàíááeáíeá òáĩĩá

< text >

ñĩĩíeòáeüíúé íòðáçíe

{text}

íáðáçáíð÷eáúé íòðáçíe

{---

## 2. Òeĩĩeĩáeý ðá÷ááúõ àeòíá

I Àeòú, ñĩñòááeýpùeá ñíáæíúá íàðú

### 1.1 Àeòú óĩðááeáíeý àeàeĩĩĩ

#### 1.1.1 Kĩĩíóíeèàöeý

1.1.1.1 Ἐᾶᾶᾶᾶ (RIE RIJ)

1.1.1.2 Ἰᾶᾶᾶ ᾶᾶᾶ

1.1.2 Ἐᾶᾶᾶᾶᾶ ἰᾶᾶᾶᾶ

1.1.2.1 Ἐᾶᾶᾶᾶᾶᾶ, ἰᾶᾶᾶᾶᾶᾶᾶᾶ ἰᾶᾶᾶᾶᾶᾶ

1.1.2.2 Ἰᾶᾶᾶᾶᾶ ἰᾶᾶᾶᾶᾶ

1.1.2.3 Ὄᾶᾶᾶᾶᾶ ᾶᾶᾶᾶᾶ ἰᾶᾶᾶᾶ (VTE VTJ)

1.2 Ἐᾶᾶᾶᾶᾶᾶᾶ ᾶᾶᾶ

1.2.1 Ἄᾶᾶᾶᾶᾶᾶ (DIE DIJ)

1.2.2 Ἀᾶᾶᾶᾶᾶ (KYE KYJ)

KYE: ἰᾶᾶᾶᾶ ᾶᾶᾶᾶ

KYE: ἰᾶᾶᾶᾶ ᾶᾶᾶᾶ, ἰᾶᾶᾶᾶᾶᾶ ᾶᾶᾶᾶᾶᾶᾶ ἰᾶᾶᾶᾶ

KYE: àëüòáđíàòèáíúé áĩđĩñ

KYE: ñĩáöèàëüíúé áĩđĩñ

KYE: èííá

KYJ: äà

KYJ: íáò

KYJ: ñĩĩëàñííá íáò

KYJ: èííé íòááò íà íáúèé áĩđĩñ

KYJ: àëüòáđíàòèèà: íáíá

KYJ: àëüòáđíàòèèà: íáá

KYJ: àëüòáđíàòèèà: òđáòèé áúáíđ

KYJ: ᾶᾶᾶᾶᾶᾶᾶᾶ: ἰᾶᾶᾶᾶᾶᾶ

KYJ: ᾶᾶᾶᾶᾶᾶᾶᾶ: ἔἰᾶ

KYJ: ᾶᾶᾶᾶᾶᾶᾶᾶ ἰᾶᾶᾶᾶ

KYJ: ἰᾶᾶᾶᾶᾶᾶᾶ ἔἰᾶᾶᾶᾶᾶᾶ

KYJ: ἰᾶᾶᾶᾶ

KYJ: ἔἰᾶ

### 1.2.3 ἰᾶᾶᾶᾶ

Ἰ ἰᾶᾶᾶᾶᾶᾶ ᾶᾶᾶᾶ

## 2.1 Ἀᾶᾶᾶ ᾶᾶᾶᾶᾶᾶᾶ ᾶᾶᾶᾶᾶᾶ

### 2.1.1 Ἐᾶᾶᾶᾶᾶᾶᾶᾶ

#### 2.1.1.1 Ἐᾶᾶᾶᾶᾶ (RY)

### 2.1.1.2 Íáðàòíáŷ ñáŷçü

### 2.1.2 Ðàçðåøáíeá ïðíáeái

#### 2.1.2.1 Èñĩðàáeéáíeá

### 2.2 Èíóĩðíàöeíĩíúá àeòú

#### 2.2.1 Íñĩíáíúá àeòú (YA)

#### 2.2.2 Äĩĩeíáíeŷ íñĩíáíúó àeòíá (ĩĩŷñĩíáíeá, óòí=-íáíeá)

Íðeíá=-áíeá: Íĩáðíáíĩ ðàñĩeñáíá áðóíĩá áĩĩðĩñĩá è ìòááòíá &mdash; íàeáíeáá ÷-àñòíóíúó ðá-ááúó àeòíá; ŷñòíĩñeàŷ áááðááeàòóðá,

### Summary

The Estonian Dialogue Corpus is collected with the aim of developing the dialogue system using the natural language. The spoken dialogues (884 dialogues, 155000 running words) are used to study the rules and norms of the human-human communication; the corpus also includes human-computer dialogues (21, 2500 running words) collected by the Wizard of Oz method used to study the role behaviour of the users and information provider. The presentation considers the means and levels of transcription and annotation dialogues and also the application of the corpus.