

Íé íáíá èç ðàññííòðáííúò ñòáí ðàçíáòèè íá íòáá-àèà òáèyí ííääèèðíááíèy è èññéääíááíèy áñòáñòááííúò äèàèíáíá ñ òí-èè çðáíèy òèííéíáèé [Gerassimenko et al. 2004].

Òèííéíáèy ñíñòíèò èç 126 äèàèíáíáúò àèòíá è íñííáúúääòñy íá ðàññííòðáíèè äèàèíáá èàè ñíòèàèüííáí áçàèííáéñòáèy. Ðá-ááúá íáèíí-íúá àèòú (íáíðèíáð, ííáóäèòáèüííá ðáíèèèè íáðàòíé ñáyçè). Ñ äðóáíé ñòíðííú, àèòú ííäðàçááèyðòñy íá íòííñyùèáñy è íðááí.

Áíñáíúñíò áääääòáòú óñòíúò äèàèíáíá è áääääòáòú íáèí äèàèíá ÁÍÇ ðàçíá-áíú ááóíy áíííòáòíðáíè, ðàçííáèàñèy èíòíðúò óñòðáííú áñíííáàòáèüííé íðíáðáííú.

Íðèíáð 1 (áñòáñòááííúé äèàèíá á íáðááíáá íá ðóññèèé).

((çáííé) | RIE: ÇÁÍÍÉ |

V: `ñíðááí-íáy | RIJ: ÎÒÁÁÒ ÍÀ ÇÁÍÍÉ | | RY: ÈÁÁÍÒÈÒÈÈÁÖÈÈ |

`Kðèñòà | RY: ÈÁÁÍÒÈÒÈÈÁÖÈÈÈ |

çäðááñòáóéòá? | RIE: ÌÐÈÁÁÒÑÒÁÈÁ |

(.)

H: çäðááñòáóéòá, | RIJ: ÌÐÈÁÁÒÑÒÁÈÁ |

=y `òíòáè áú ((íáçááíèá òèðíú)) (.) < àèñ`íáò-áðá > èèè `èàèíé-íèáóáü `íí- ííáð. | DIE: ÄÈÐÁÈÒÈÁ |

(0.8)

V: ì ìèíóòì÷éó. | DIJ: ÍÒÑĐÎ×ĒĀ |

(5.5)

V: òǎěâôîí íá ñěóæèâàíèÿ [òì āñòü,] àà? | KYE: ÍÁÙĒĒ ĀĪĐÎŃ | | VTE: ÓÒÎ×ÍÁÍĒĀ ÓÑĒÍĀĒĒ ÍÒĀĀÒĀ |

H: [ǎǎ.] | KYJ: ĀĀ | | VTJ: ÓÓÎ×ÍÁÍĒĀ ÓÑĒÍĀĒĒ ÍÒĀĀÒĀ |

V: ñ ī íāřāòâāđǎāáířīūì āàířīūì řīāđ < ÷âòúđǎ ÷âòúđǎ řāíū? (0.5) āāâyòü āîñāíū? (.) [řāèí] říëü. > | DIJ: ĒÍÔĪĐÎĀÖĒß |

H: [.hh] | YA: ĒÍĪĀ |

(.)

H: ñīāñēáí. | RIE: ŃĪĀÑĒĒĀĪ |

V: ñīæāèóéñòà? | RIJ: ĪĪÆĀĒÓÉŃÒĀ |

2.3. Êíííóíééàòèáíúá ñòðàòáãèè

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

Äèý ìðáááéáíéý êíííóíééàòèáíúá ñòðàòáãèè á CDM èñííèüçòòñý -áòúðá èííòáèñòáèèúíúò òàèòíðà:

– íæèääáííñòù ðáíéèèè;

– ñáyçü ðáíéèèè ñ òáííé;

– áíñèèáíóòíñòù òáèéé áíáíðýùááí;

– èíèèèèèèè áíáíðýùááí èèè ìàðòíáðà.

Áñá èííòáèñòáèèúíúá òàèòíðú áéíáðíú, ñííòááòñòááííí, êíííóíééàòèáíúá ñòðàòáãèè $2*4=16$.

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

Êíííóíééàòèáíúá ñòðàòáãèè ðàçíá-áíú á 20 áñòáñòááííúá è 20 ñèìóèýòèííúá àèèííáð.

Ίδειαό 2 (ηειόέγέίίίúé áéáíā ā íāđāíāā íā đōññēēé ýçúê).

Đáíēēēā

Ñòđàòāāēý

ìāāçèí

Éēāíò:

Éāē āíāđàòü āí ĩýđíó èç Òāđòó āí ĩēóāíý?

íæēāāāì.-ñāýçí.-āíñòēāí.-āíāíđýù.-íā÷àēí/ēíāō

–

Ἐῖῖῖῖῖῖῖῖ:

Ἀῖῖῖῖῖῖ ἂῖῖῖῖῖῖ ἂ 5 ὀῖῖῖῖ.

Ἀῖῖῖῖῖῖ ἂῖῖῖῖῖῖ ἂ 8 ὀῖῖῖῖ.

ἰῖῖῖῖῖῖῖ.-ῖῖῖῖῖῖῖῖ.-ἰῖῖῖῖῖῖῖῖῖ.-ῖῖῖῖῖῖῖῖῖῖ.-ῖῖῖῖῖῖῖῖῖῖ ὀῖῖῖῖ

Ὀῖῖῖῖῖῖῖῖῖ

Ἐῖῖῖῖῖῖῖῖῖ ἔῖῖ ἂῖῖ ἂῖῖῖῖῖ ῖῖῖῖῖῖῖῖῖ? ...

ἰῖῖῖῖῖῖῖῖῖῖῖ.-ῖῖῖῖῖῖῖῖῖῖῖῖ.-ῖῖῖῖῖῖῖῖῖῖῖῖ.-ῖῖῖῖῖῖῖῖῖῖῖῖ ἂῖῖῖῖῖῖῖῖῖῖῖῖ

Ὀῖῖῖῖῖῖῖῖῖῖ

Ḑāçīā-āīīūē ēīḐīōñ ēñīēūçōāōñý ā òāīḐāòē-āñēēō ē īḐēēāāīūḐ ēññēāāīāāīēyō (īāīḐ., [Hennoste et al. 2005]) ē ā Ḑāāīòā īāā ā [Fishel 2004], ā òāēæā ā īīñḐḐīāīēē īāḐāūḐ āēāēīāīāūḐ ñēñōāī īā ýñōīñēīī ýçūēā. Ýōē ñēñōāīū, ÒḐāīñīḐōīūē ē ÒāāḐḐāēūīūē ā īḐēīāīāīēāī çīāīēē ī ñḐḐōēōḐā ēīḐīḐīāḐēīīūḐ āēāēīāīā ē īīāāēāēīāīā.

Ñīēñīē ēēòāḐāòóḐū

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, 4–5 April 2005, 249–254.

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 “Human Language Technologies. The Baltic Perspective”. Commission of the Official Language at the Chancellery of the President of Latvia, Riga, 2004, 150–155.

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, 420–427.

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, 13–31.

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

ÁēāīāāḐīīñōē

ḐāāīḐō īīāāāḐæēāāāḐ Ýñōīñēēē īāó-īūē Ḑīā (āḐāīḐ 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. Ōðāīñēðēīðēīīūā çíàèè

ñīāā èíòííàöèè .

īīēóñīāā èíòííàöèè ,

īīāúāī èíòííàöèè ?

ēīðīðēāÿ iàóçà (iàēñ. 0.2 ñ.) (.)

äèèià iàóçú â ñāēóíāàõ (2.0)

íäēīæāíēā [text]

ñèèÿíēā íaçāâēñèìúõ ääèìö =

ðāñòÿíóòúé çāóê ::

óääðĩĩá ñéĩáĩ

`

ĩðáðááĩĩĩá ñéĩáĩ

do-

ääĩõ

.hhh

óáúñòðáíéá òáĩĩá

> text <

çáĩáääéáíéá òáĩĩá

< text >

ñĩĩíéòáëüíúé íòðáçíé

{text}

íáðáçáíð÷éáúé íòðáçíé

{---}

2. Òèĩĩéĩáëÿ ðá÷ááúõ àéòĩá

I Àéòú, ñĩñòááëÿpùéá ñĩáæíúá íàðú

1.1 Àéòú óĩðááéáíéÿ àèàéĩĩĩ

1.1.1 Kĩĩíóíéèàòèÿ

1.1.1.1 Δὲδὲῖ (RIE RIJ)

1.1.1.2 Ἰῖαῖα ὀῖῖ

1.1.2 Δὰδὰῖῖ ἰῖῖῖ

1.1.2.1 Ἰῖῖῖῖῖῖῖῖ, ἰῖῖῖῖῖῖῖῖῖ ἰῖῖῖῖῖῖ

1.1.2.2 Ἰῖῖῖῖῖ ἰῖῖῖῖῖ

1.1.2.3 Ὀῖῖῖῖῖ ὀῖῖῖῖῖ ἰῖῖῖῖῖ (VTE VTJ)

1.2 Ἰῖῖῖῖῖῖῖῖ ἰῖῖῖ

1.2.1 Ἰῖῖῖῖῖῖῖ (DIE DIJ)

1.2.2 Ἰῖῖῖῖῖῖ (KYE KYJ)

KYE: ἰῖῖῖῖ ἰῖῖῖῖ

KYE: ἰῖῖῖῖ ἰῖῖῖῖ, ἰῖῖῖῖῖῖῖ ἰῖῖῖῖῖῖῖῖ ἰῖῖῖῖῖῖ

KYE: àëüòáđíàòèáíúé áĩĩđĩñ

KYE: ñĩáöèàëüíúé áĩĩđĩñ

KYE: èííá

KYJ: äà

KYJ: íáò

KYJ: ñĩĩäèñííá íáò

KYJ: èííé íòááò íà íáúèé áĩĩđĩñ

KYJ: àëüòáđíàòèèää: íáíá

KYJ: àëüòáđíàòèèää: íáá

KYJ: àëüòáđíàòèèää: òđáòèé áúáíđ

KYJ: ᾶᾶᾶᾶᾶᾶᾶᾶ: ἰᾶᾶᾶᾶᾶᾶ

KYJ: ᾶᾶᾶᾶᾶᾶᾶᾶ: ἲᾶᾶ

KYJ: ᾶᾶᾶᾶᾶᾶᾶᾶ ἰᾶᾶᾶᾶ

KYJ: ἰᾶᾶᾶᾶᾶᾶᾶ ἲᾶᾶᾶᾶᾶᾶ

KYJ: ἰᾶᾶᾶᾶ

KYJ: ἲᾶᾶ

1.2.3 ἰᾶᾶᾶᾶ

ἰᾶᾶᾶᾶᾶᾶ ᾶᾶᾶᾶ

2.1 Ἀᾶᾶᾶ ᾶᾶᾶᾶᾶᾶᾶ ᾶᾶᾶᾶᾶᾶ

2.1.1 ἲᾶᾶᾶᾶᾶᾶᾶᾶ

2.1.1.1 ᾶᾶᾶᾶᾶᾶ (RY)

2.1.1.2 Īāðāōīāy ñāyçü

2.1.2 Ðaçðåøáíēå ĩðīāēåì

2.1.2.1 Èñīðāāēēāíēå

2.2 Èíōīðīàōēēííūå àèòù

2.2.1 Īñīīāíūå àèòù (YA)

2.2.2 Äīīēīāíēy ĩñīīāíūō àèōīā (īīyñīāíēå, óòī=-íāíēå)

Īðēīā-āíēå: Īīāðīāíī ðāñīēñāīā āðōīīā āīīðīñīā è òāāōīā — íāēāíēåå ÷-āñōíōíūō ðā-āåūō àèōīā; ýñōííñēāy āāāðāēēāðōðā,

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.