A “Portrait” Approach to Multichannel Discourse

Andrej A. Kibrik & Olga V. Fedorova

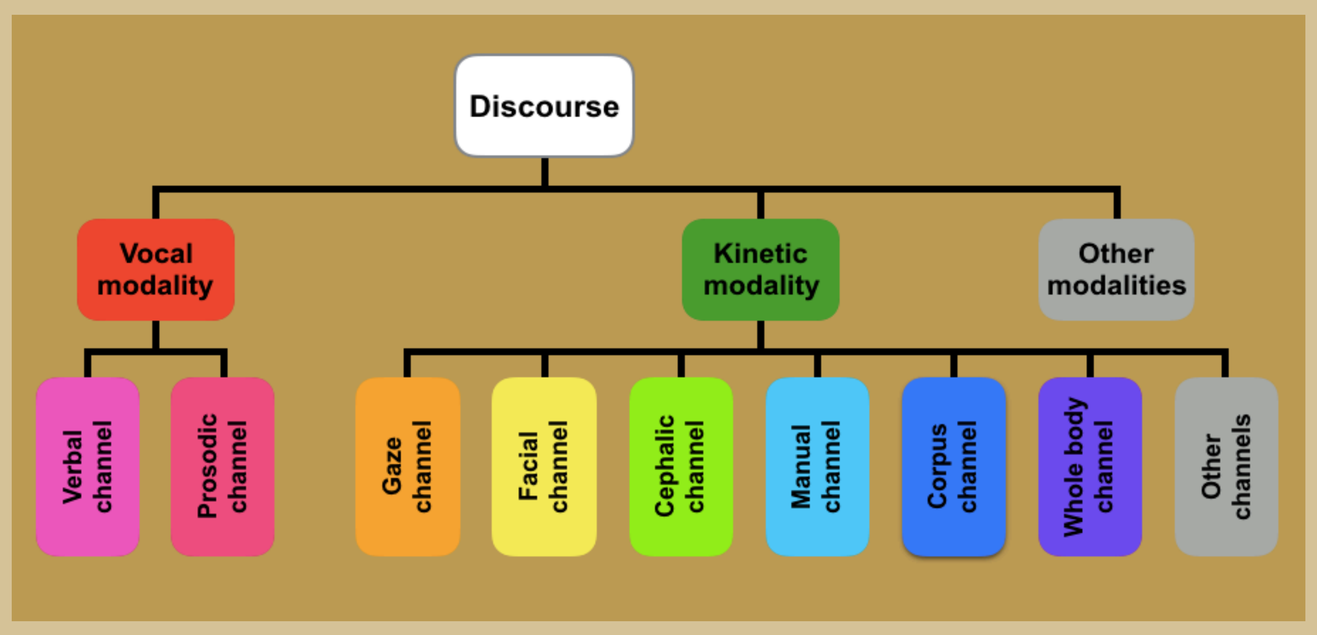
ФЕДЕРАЛЬНОЕ АГЕНТСТВО НАУЧНЫХ ОРГАНИЗАЦИЙ РОССИИ
ИНСТИТУТ ЯЗЫКОЗНАНИЯ
РОССИЙСКОЙ АКАДЕМИИ НАУК

MOSCOW STATE UNIVERSITY LOMONOSOV

RSF Russian Science Foundation
#14-18-03819

http://multidiscourse.ru

Language as is: Russian Multimodal Discourse



Technical solutions

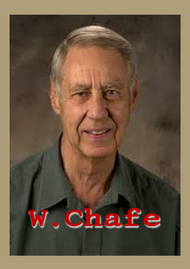
Zoom H6 Recorder
96 kHz / 24 bit

Tobii Glasses II
50 Hz,1920x1080

Jai Go-5000M
100 fps,1392x1000

Russian Pear Chats and Stories

Design



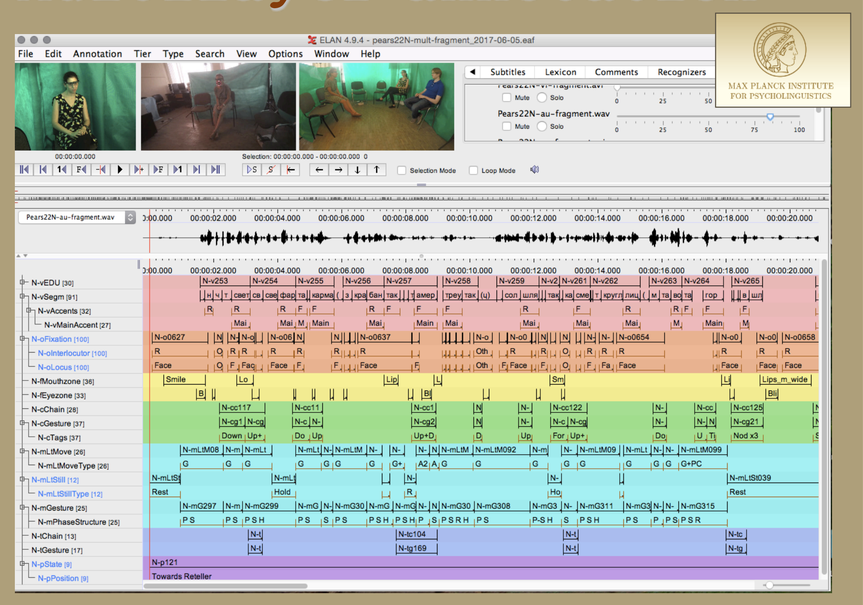
Pear Film (Chafe ed. 1980)

- The Narrator and the Commentator watch the film
- First Telling
- The Narrator tells the Reteller about the film
- Conversation
- The Commentator adds details, all three discuss the film
- Retelling
  The Reteller tells about the film to the Listener, who has just joined the group
- Second Retelling (written)
  The Listener writes down the contents of the film

Resource

- 24 sessions recorded in the summer of 2015
- 96 participants
  - 18 to 36 yrs old
  - 36 men & 60 women
  - 42 persons with higher education & 54 students
- 9 hours
  - 12 to 38 minutes
- About 100 K words
- 6 Tb of data

Multilayer annotation



Prosodic Portrait

A complete account of a spoken discourse presupposes a speaker’s Prosodic Portrait, i.e. a range of her prosodic characteristics. Compiling such a portrait is stage zero in the work on a transcript, preceding vocal annotation as such.
In particular, two kinds of the falling intonation must be distinguished: a final falling and a non-final falling. In order to posit punctuation marks in a transcript belonging to a certain speaker, one needs to explore and describe the speaker’s prosodic system.

Minimal F0 value, Hz	91
Maximal F0 vaue, Hz	214
Standard level of EDU onsets, Hz	140±5
Target level of final fallings, Hz	105±3
Target level of non-final fallings, Hz	115±3
Target level of rises in the canonical comma intonation, Hz	140~185
Standard level of falling on post-accent syllables, typical of the canonical comma intonation, Hz	120±5
Target level of rises in the “three dots” intonation, Hz	118~132

Oculomotor Portrait

Oculomotor Portrait is created after the procedure of semi-automatic annotation and serves at the stage of data analysis.
Application developer is Ivan Zherdev.

Total duration	1170.667
Total count	2249
Mean duration & std	0.52 & 0.66
Min, Quantile 25, 50, 75, Max	0.06, 0.16, 0.26, 0.58, 10.5
R: count & ratio	1033 & 0.46
R: duration & ratio	848.55 & 0.72
R: mean & std	0.82 & 0.85
R: Min, Quantile 25, 50, 75, Max	0.06, 0.24, 0.5, 1.16, 10.5
C: count & ratio	129 & 0.057
C: duration & ratio	43.8 & 0.036
C: mean & std	0.33 & 0.28
C: Min, Quantile 25, 50, 75, Max	0.06, 0.16, 0.24, 0.4, 1.32
L: count & ratio	16 & 0.007
L: duration & ratio	3.28 & 0.003
L: mean duration & std	0.2 & 0.12
L: Min, Quantile 25, 50, 75, Max	0.06, 0.12, 0.19, 0.26, 0.44
Surroundings: count & ratio	1071 & 0.48
Surroundings: duration & ratio	275.02 & 0.23
Surroundings: mean & std	0.26 & 0.21
Surroundings: Min, Quantile 25, 50, 75, Max	0.06, 0.13, 0.2, 0.3, 2.2

Gesticulation Portrait

Gesticulation Portrait fulfils a double function. It is necessary both at the stage of manual gesture annotation and at the stage of gesture analysis.

Inclination to stillness	low
Inclination to adaptors	high
Amplitude of manual gesture	low
Velocity of manual gesture	high
Two-handed vs. one-handed	0.32 vs. 0.68
One-handed gestures in accordance with handedness: right vs. left	0.68 vs. 0.32
Total count	481
Total gesticulation duration	414.53
Mean duration & std	0.86
Min, Quantile 25, 50, 75, Max	0.09, 0.42, 0.64, 1, 8.7
Conversation: count & ratio	210 & 0.44
Conversation: duration & ratio	198.72 & 0.48
Conversation: Min, Quantile 25, 50, 75, Max	0.16, 0.45, 0.68, 1.15, 8.73
Retelling: count & ratio	271 & 0.56
Retelling: duration & ratio	215.81 & 0.52
Retelling: Min, Quantile 25, 50, 75, Max	0.09, 0.38, 0.62, 0.96, 3.73

Poster presented at 11th edition of the Language Resources and Evaluation Conference, 7-12 May 2018, Miyazaki (Japan)

olga.fedorova@msu.ru