An introduction to the multimodal corpus *HuComTech* and its annotation

László Hunyadi
University of Debrecen

Nijmegen, MPI, 19 December 2012
HuComTech:
Human-Computer Technologies
Hungarian Communication Technologies
research modules involved:
research modules involved:

- computational linguistics
research modules involved:

- computational linguistics
- communication theory
research modules involved:

- computational linguistics
- communication theory
- psychology
research modules involved:

- computational linguistics
- communication theory
- psychology
- digital image processing
research modules involved:

- computational linguistics
- communication theory
- psychology
- digital image processing
- engineering (robotics)
Some basic data about the corpus
Purpose of the corpus

to identify elements of human-human communication and structural relations between them that are

- relevant for HCI
- technologically implementable

furthermore, to
- learn the multimodal nature of human communication (both its verbal and nonverbal aspects)
- describe human communication in a multimodal, holistic model
The corpus is intended to represent sufficient data in proper arrangement for purposes of

linguistics
language technology (the teaching and testing of a speech recognition software)
behavioral psychology
robotics
    and more
Corpus:
Corpus:

- appr. 60 hours of video-recordings of 111 speakers aged 18-29, including
Corpus:

- appr. 60 hours of video-recordings of 111 speakers aged 18-29, including

- ≈ 450,000 word tokens
Corpus:

- appr. 60 hours of video-recordings of 111 speakers aged 18-29, including
- \( \approx 450,000 \) word tokens
- 15 read sentences
Corpus:

- appr. 60 hours of video-recordings of 111 speakers aged 18-29, including
- $\approx 450,000$ word tokens
- 15 read sentences
- 10 minute guided dialogues (job interviews)
Corpus:

- appr. 60 hours of video-recordings of 111 speakers aged 18-29, including
- \( \approx 450,000 \) word tokens
- 15 read sentences
- 10 minute guided dialogues (job interviews)
- 15 minute free dialogues
Distribution of subjects by sex

- Male: 54.5%
- Female: 45.5%

Distribution of subjects by age

- # subjects
- Age range: 19 to 30

- 19: 8
- 20: 48
- 21: 48
- 22: 11
- 23: 10
- 24: 7
- 25: 3
- 26: 1
- 27: 2
- 28: 2
- 29: 2
- 30: 2
Annotation

serves the study of multimodality through the study of unimodality and the fusion of aligning markers

Markers to be annotated are determined by a theoretical-technological model of communication
Theoretical considerations
Theoretical considerations

We assume that human communication has a two-way mechanism: speakers and listeners rely on the same mechanism to communicate.
Theoretical considerations

We assume that human communication has a two-way mechanism: speakers and listeners rely on the same mechanism to communicate.

Therefore, in order to properly represent human communication for technology, we need a model to follow this two-way mechanism: to serve both synthesis and analysis.
Theoretical considerations

We assume that human communication has a two-way mechanism: speakers and listeners rely on the same mechanism to communicate.

Therefore, in order to properly represent human communication for technology, we need a model to follow this two-way mechanism: to serve both synthesis and analysis.

We assume that the approach of a generative model proposed for syntax (especially Chomsky 1981) can be useful in building such a two-way model of communication for technology.
The basic structure of the model

Modularity:

representation \quad \rightarrow \quad \text{pragmatic extension}

\quad \downarrow

\text{invariant structure} \quad \rightarrow \quad \text{functional extension}

\quad \downarrow

\quad \rightarrow \quad \text{basic structure}

\quad \rightarrow \quad \text{operational component}
The basic structure of the model

The pragmatic extension:

- Surface structure
- Pragmatic primitives
- Individual variations
- Type scenario
- Specific ontology
- Operations
Characteristics of the constituent modules
Characteristics of the constituent modules
Characteristics of the constituent modules

- Each module has a characteristic finite set of primitives and, by way of the Operational component these primitives are combined into an infinite set of non-primitives and further structures
Characteristics of the constituent modules

- Each module has a characteristic finite set of primitives and, by way of the Operational component these primitives are combined into an infinite set of non-primitives and further structures.

- The basic structure generates all and only those structures (configurations of primitives) that are formally possible (‘grammatical’) in any communicative event.
Each module has a characteristic finite set of *primitives* and, by way of the *Operational component* these primitives are combined into an infinite set of non-primitives and further structures.

The *basic structure* generates all and only those structures (configurations of primitives) that are formally possible (‘grammatical’) in any communicative event.

Eg. The ‘start’ of an event can be followed by ‘end’ of an event, but the inverse order is not possible (‘ungrammatical’).
Characteristics of the constituent modules

- Each module has a characteristic finite set of *primitives* and, by way of the *Operational component* these primitives are combined into an infinite set of non-primitives and further structures.

- The *basic structure* generates all and only those structures (configurations of primitives) that are formally possible (‘grammatical’) in any communicative event.

  Eg. The ‘start’ of an event can be followed by ‘end’ of an event, but the inverse order is not possible (‘ungrammatical’).

- The *functional extension* assigns all possible communicative functions to any given structure generated by the basic structure and only those
Each module has a characteristic finite set of primitives and, by way of the Operational component these primitives are combined into an infinite set of non-primitives and further structures.

The basic structure generates all and only those structures (configurations of primitives) that are formally possible (‘grammatical’) in any communicative event.

Eg. The ‘start’ of an event can be followed by ‘end’ of an event, but the inverse order is not possible (‘ungrammatical’).

The functional extension assigns all possible communicative functions to any given structure generated by the basic structure and only those.

Eg. the restart node in the basic structure can be associated with the continuity function, but the assignment of the function turn-taking to it is not possible.
Characteristics of the constituent modules
Characteristics of the constituent modules
Characteristics of the constituent modules

- The *pragmatic extension* actualizes the input from the functional extension for the given actual communicative event by selecting the appropriate *markers* and their appropriate *values* based on the given scenario and ontology behind the event.
Characteristics of the constituent modules

- The *pragmatic extension* actualizes the input from the functional extension for the given actual communicative event by selecting the appropriate *markers* and their appropriate *values* based on the given scenario and ontology behind the event.

  Eg. The function of ‘happiness’ is expressed by the appropriate value of some modal marker(s): facial, gestural, audio, lexical or some/all of them.
The pragmatic extension is the interface to technology:

- Functions are translated into their technological counterparts as parameters through data fusion.
- The pragmatic extension selects the modalities and their markers to represent the given function.
- Actual occurrences of markers are represented by the corresponding parameter values.
- Technology receives these parameter values as input and operates on them.
Experimental settings and annotation
Annotation

unimodal (video, audio)
multimodal (video + audio)

manual
automatic

description of physical properties (esp. video)
interpretative annotations

with focus on emotions and the multimodal alignment of video and audio

special features of annotation:
  pragmatic
  syntactic
  prosodic
Levels of annotation and attributes (labels)

Audio
Levels of annotation and attributes (labels)

Audio

*IP-level: HC, SC, EM, IN, BC, HE, RE, IT, SL, V*
Levels of annotation and attributes (labels)

Audio

**IP-level**: HC, SC, EM, IN, BC, HE, RE, IT, SL, V

**discourse-level**: TT, TK, BC, SL
Levels of annotation and attributes (labels)

Audio

IP-level: HC, SC, EM, IN, BC, HE, RE, IT, SL, V

discourse-level: TT, TK, BC, SL

emotion-level: neutral, sad, happy/laughing, surprised, recall, tensed (and degrees of: strong, moderate, reduced), other, silence
Levels of annotation and attributes (labels)

Video
Levels of annotation and attributes (labels)

Video

*comevent*: start, end
Levels of annotation and attributes (labels)

**Video**

comevent: start, end

deictic: addressee, self, measure, object, shape; left, right, both
Levels of annotation and attributes (labels)

Video

comevent: start, end

deictic: addressee, self, measure, object, shape; left, right, both

emblems: attention, agree, block, disagree, doubt, doubt-shrug, refusal, surprise, more-or-less, number, finger-ring, hands up, one hand other hand, other
Levels of annotation and attributes (labels)

Video
Levels of annotation and attributes (labels)

Video

*emotions*: natural, happy, recall, sad, surprise, tense; (and degrees of: strong, moderate, reduced)
Levels of annotation and attributes (labels)

Video

emotions: natural, happy, recall, sad, surprise, tense; (and degrees of: strong, moderate, reduced)

headshift: lower, turn, raise, shake, nod; sideways, left, right
Levels of annotation and attributes (labels)

**Video**

*emotions*: natural, happy, recall, sad, surprise, tense; (and degrees of: strong, moderate, reduced)

*headshift*: lower, turn, raise, shake, nod; sideways, left, right

*touchmotion*: hair, leg, arm, face, eye, ear, chin, mouth, neck, bust, forehead, nose, glasses; tap, scratch; left, right
Levels of annotation and attributes (labels)
Levels of annotation and attributes (labels)

Video

*posture*: crossing arm, holding head, lean back, lean forward, lean left, lean right, rotate right, rotate left, shoulder up, upright
Levels of annotation and attributes (labels)

Video

*posture*: crossing arm, holding head, lean back, lean forward, lean left, lean right, rotate right, rotate left, shoulder up, upright

*handshape*: breaking, fist, crossing fingers, open flat, open spread, thumb out, index out; left, right, both
Levels of annotation and attributes (labels)

Video
Levels of annotation and attributes (labels)

Video

facial expressions: natural, happy, recall, sad, surprise, tense (and degrees of: moderate, reduced, strong)
Levels of annotation and attributes (labels)

Video

*facial expressions*: natural, happy, recall, sad, surprise, tense (and degrees of: moderate, reduced, strong)

*eyebrows*: scowl, up; left, right, both
Levels of annotation and attributes (labels)

Video

*facial expressions*: natural, happy, recall, sad, surprise, tense (and degrees of: moderate, reduced, strong)

eyebrows: scowl, up; left, right, both

gaze: blink, up, down, left, right, forwards, left-up, left-down, right-up, right-down
Levels of annotation and attributes (labels)

Syntax

structural segmentation:
Levels of annotation and attributes (labels)

Syntax

structural segmentation:

clause boundaries
Levels of annotation and attributes (labels)

Syntax

structural segmentation:

clause boundaries

hierarchical arrangement of clauses
Levels of annotation and attributes (labels)

Syntax

structural segmentation:

clause boundaries

hierarchical arrangement of clauses

internal structure of clauses (esp. missing elements)
Levels of annotation and attributes (labels)

Syntax vs. prosody

(prosody of clauses)
Levels of annotation and attributes (labels)

Syntax vs. prosody

(prosody of clauses)

pitch movement: rise, fall, stagnant + finer distinctions
Levels of annotation and attributes (labels)

Syntax vs. prosody
(prosody of clauses)

*pitch movement*: rise, fall, stagnant + finer distinctions

*intensity*: increase, decrease, stagnant + finer distinctions
Levels of annotation and attributes (labels)

Syntax vs. prosody

(prosody of clauses)

pitch movement: rise, fall, stagnant + finer distinctions

intensity: increase, decrease, stagnant + finer distinctions

pause/duration: increase, decrease, stagnant + finer distinctions
Levels of annotation and attributes (labels)

Pragmatics - multimodal
Levels of annotation and attributes (labels)

Pragmatics - multimodal

Annotation: DiAMSL for text based eventseral on several layers, esp. audio: turn management, discourse
Levels of annotation and attributes (labels)

Pragmatics - multimodal

Annotation: DiAMSL for text based eventseral on several layers, esp. audio: turn management, discourse

Multimodality - the complex of audio + video multimodal communicative act (annotation based on Bach-Harnish)
Levels of annotation and attributes (labels)

Pragmatics - multimodal
Levels of annotation and attributes (labels)

Pragmatics - multimodal

communicative act types: constatives directives, comissives, acknowledgements, none
Levels of annotation and attributes (labels)

Pragmatics - multimodal

*communicative act types*: constatives, directives, comissives, acknowledgements, none

*supporting events of communicative acts*: backchannel, *politeness markers*, corrections, no support
Levels of annotation and attributes (labels)

Pragmatics - multimodal
Levels of annotation and attributes (labels)

Pragmatics - multimodal

*thematic control*: topic initiation, elaboration, topic change (contextual, non-contextual)
Levels of annotation and attributes (labels)

Pragmatics - multimodal

*thematic control*: topic initiation, elaboration, topic change (contextual, non-contextual)

*information structure*: given vs. new information
Levels of annotation and attributes (labels)

Pragmatics - unimodal
Levels of annotation and attributes (labels)

Pragmatics - unimodal

agreement: uninterested, disagree, block, uncertainty; full, partial
Levels of annotation and attributes (labels)

Pragmatics - unimodal

*agreement*: uninterested, disagree, block, uncertainty; full, partial

*attention*: calling, paying
Levels of annotation and attributes (labels)

Pragmatics - unimodal

*agreement*: uninterested, disagree, block, uncertainty; full, partial

*attention*: calling, paying

*deixis*
Levels of annotation and attributes (labels)

Pragmatics - unimodal
Levels of annotation and attributes (labels)

Pragmatics - unimodal

information: received novelty
Levels of annotation and attributes (labels)

Pragmatics - unimodal

*information*: received novelty

*turn-management*: intending to start speaking, start speaking successfully, end speaking, breaking in
Sample data for multimodal alignments

Turn management
Sample data for multimodal alignments

Turn management

*turn-give*: forward, blink, down, left-down, right-down
Sample data for multimodal alignments

Turn management

*turn-give*: forward, blink, down, left-down, right-down

*turn-take*: forward, blink, down, left-down, right-down
Sample data for multimodal alignments

**Turn management**

*turn-give*: forward, blink, down, left-down, right-down

*turn-take*: forward, blink, down, left-down, right-down

*break-in_turn-keep*: forwards, blink, up, down, left-down, right-down
Sample data for multimodal alignments

Emotions vs. gestures
Sample data for multimodal alignments

Emotions vs. gestures

uncertainty is mostly found to be associated with the hand gesture open spread, less frequently with crossing fingers.
Sample data for multimodal alignments

Emotions vs. gestures

Uncertainty is mostly found to be associated with the hand gesture open spread, less frequently with crossing fingers.

Agreement is also associated with open spread and crossing fingers.
Sample data for multimodal alignments

Emotions vs. gestures
Sample data for multimodal alignments

**Emotions vs. gestures**

doubt is found to be associated with *open spread*, *crossing fingers* and *sideways* as well
# Video annotation (manual vs. automatic)

<table>
<thead>
<tr>
<th>Annotation Method</th>
<th>Advantage/Disadvantage (+/-)</th>
<th>Physical Values</th>
<th>Interpretative Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manual</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Automatic</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

**Essential difference between the two:** automatic annotation is ‘digital’ following framewise judgement across a predefined number of frames, whereas manual annotation is ‘analog’
Automatic annotation: Noldus FaceReader
Settings and sample output

Values assigned to single frames hence only begin time of the given value
Binary value assignment
Timeline: natural vs. happy
Valence
Summary statistics

Analysis Block 1

- Happy (14.4%)
- Neutral (49.4%)
- Scared (13.9%)
- Other (3.3%)
- Angry (10.4%)
- Sad (8.6%)
Comparison: automatic vs. manual emotion recognition

Analysis Block 1

- Happy (14.4%)
- Neutral (49.4%)
- Scared (13.9%)
- Angry (10.4%)
- Sad (8.6%)
- Other (3.3%)

Manual annotation

- Happy (45%)
- Natural (42%)
- Tense
- Recall
- Surprise

Wednesday, April 24, 13
Comparison: automatic vs. manual emotion recognition

Although both systems are based on the FACS model of emotions, different categories (emotions) recognised

Whereas both systems assign interpretative values, manual annotation selects ‘more difficult’ ones

Manual annotation offers subjectively observed degrees of emotions (strong, moderate, reduced), for automatic annotation thresholds for being ‘happy’, ‘angry’ etc. are determined statistically > smaller degrees are left out
Comparison: automatic vs. manual emotion recognition

Eventual unrealistic values in automatic annotation are the result of the single frames approach.

Duration is not marked > offset of an emotion cannot be determined in case of non-continuous labels.

Most agreement between the two approaches: happy, natural.
Annotation of spoken syntax and its relation to prosody in the HuComTech corpus
• Aim: language technology (speech-to-text)
• communication studies (alignment of multimodal markers for communicative acts and emotions)
• linguistics (the syntax-prosody interface)
Syntactic data from our annotation

Spoken language vs. written language

Grammar: same/different?

Same underlying principles:

grouping of elements
hierarchical organisation of groups

Difference: two additional dimensions of spoken language:

- time
- grouping has language specific means
<table>
<thead>
<tr>
<th># of clauses in a sentence</th>
<th>Informal dialogs</th>
<th>Formal dialogs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2933</td>
<td>688</td>
</tr>
<tr>
<td>2</td>
<td>289</td>
<td>82</td>
</tr>
<tr>
<td>3</td>
<td>163</td>
<td>24</td>
</tr>
<tr>
<td>4</td>
<td>92</td>
<td>26</td>
</tr>
<tr>
<td>5</td>
<td>44</td>
<td>9</td>
</tr>
<tr>
<td>6</td>
<td>37</td>
<td>9</td>
</tr>
<tr>
<td>7</td>
<td>24</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>24</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>17</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>21</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>22</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

![Graph showing the distribution of clauses per sentence for informal and formal dialogs](image-url)
Structural relations (hierarchy)

clause sequences with no structural relation
has no subordinate clause
has no coordinate clause
has neither subordinate, nor coordinate clause

embeddings
insertions
multiple subordination (recursion)
## Type of missing element according to syntactic code

<table>
<thead>
<tr>
<th>Type of missing element</th>
<th>Informal dialogs</th>
<th>%</th>
<th>Formal dialogs</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. nothing missing</td>
<td>2664</td>
<td>35.59</td>
<td>758</td>
<td>34.6</td>
</tr>
<tr>
<td>2. main clause</td>
<td>37</td>
<td>0.49</td>
<td>15</td>
<td>0.69</td>
</tr>
<tr>
<td>3. preceding clause</td>
<td>58</td>
<td>0.77</td>
<td>6</td>
<td>0.27</td>
</tr>
<tr>
<td>4. relative pronoun</td>
<td>89</td>
<td>1.19</td>
<td>22</td>
<td>1.01</td>
</tr>
<tr>
<td>5. conjunction</td>
<td>22</td>
<td>0.29</td>
<td>4</td>
<td>0.18</td>
</tr>
<tr>
<td>6. subject (grammatical)</td>
<td>3178</td>
<td>42.37</td>
<td>1167</td>
<td>54.14</td>
</tr>
<tr>
<td>7. subject (logical)</td>
<td>274</td>
<td>3.66</td>
<td>113</td>
<td>5.17</td>
</tr>
<tr>
<td>8. predicate</td>
<td>214</td>
<td>2.87</td>
<td>72</td>
<td>3.29</td>
</tr>
<tr>
<td>9. object</td>
<td>102</td>
<td>1.36</td>
<td>45</td>
<td>2.06</td>
</tr>
<tr>
<td>10. adverb</td>
<td>11</td>
<td>0.15</td>
<td>4</td>
<td>0.18</td>
</tr>
<tr>
<td>11. attribute</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12. verb</td>
<td>10</td>
<td>0.13</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13. unfinished clause</td>
<td>728</td>
<td>9.7</td>
<td>167</td>
<td>13.1</td>
</tr>
<tr>
<td>14 missing element not relevant</td>
<td>3375</td>
<td>45.05</td>
<td>769</td>
<td>35.21</td>
</tr>
</tbody>
</table>

**Sum:** 143.62  149.9

## Type of missing element by frequency

<table>
<thead>
<tr>
<th>Type of missing element</th>
<th>Informal dialogs</th>
<th>%</th>
<th>Formal dialogs</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>14 missing element not relevant</td>
<td>3375</td>
<td>45.05</td>
<td>769</td>
<td>35.21</td>
</tr>
<tr>
<td>6. subject (grammatical)</td>
<td>3178</td>
<td>42.37</td>
<td>1167</td>
<td>54.14</td>
</tr>
<tr>
<td>1. nothing missing</td>
<td>2664</td>
<td>35.59</td>
<td>758</td>
<td>34.6</td>
</tr>
<tr>
<td>13. unfinished clause</td>
<td>728</td>
<td>9.7</td>
<td>167</td>
<td>13.1</td>
</tr>
<tr>
<td>7. subject (logical)</td>
<td>274</td>
<td>3.66</td>
<td>113</td>
<td>5.17</td>
</tr>
<tr>
<td>8. predicate</td>
<td>214</td>
<td>2.87</td>
<td>72</td>
<td>3.29</td>
</tr>
<tr>
<td>9. object</td>
<td>102</td>
<td>1.36</td>
<td>45</td>
<td>2.06</td>
</tr>
<tr>
<td>4. relative pronoun</td>
<td>89</td>
<td>1.19</td>
<td>22</td>
<td>1.01</td>
</tr>
<tr>
<td>3. preceding clause</td>
<td>58</td>
<td>0.77</td>
<td>6</td>
<td>0.27</td>
</tr>
<tr>
<td>2. main clause</td>
<td>37</td>
<td>0.49</td>
<td>15</td>
<td>0.69</td>
</tr>
<tr>
<td>5. conjunction</td>
<td>22</td>
<td>0.29</td>
<td>4</td>
<td>0.18</td>
</tr>
<tr>
<td>10. adverb</td>
<td>11</td>
<td>0.15</td>
<td>4</td>
<td>0.18</td>
</tr>
<tr>
<td>12. verb</td>
<td>10</td>
<td>0.13</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11. attribute</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Sum:** 142.62  149.9
Syntactic types vs. gestures

Alignment of syntactic type and gestures can offer an insight into certain cognitive processes in communication:

- speech dynamics
- error detection
- gesturing the “untold”
Syntactic types vs. gestures

A very interesting finding: *pause* (*silence*) and *gaze* can be mutually supplementary:

We found very few instances of gazing at the right overlap of an unfinished clause followed by a pause, but there was frequent gazing if there was no pause in a similar position.

Also, *looking up* as gazing direction was specific to alignment with the end of an unfinished clause but was quite rare at the end of a finished/complete clause.
Annotation of prosody in the HuComTech corpus
• the “IP-level” (based on F0-contour and pause, manual)
• pitch movement (automatic)
• intensity change (automatic, in progress)
• accent/stress detection (automatic, in progress)
Detection of pitch movement

- aim: to generate data on pitch movement trends (actual movement type, tone range)
- capture F0-properties of syntactic types
- assign communicative functions (including emotions)
The calculation of the trend of pitch movement

- based on Praat (algorithm and scripts)
- trend of syllable based stylization (Szekrényes, 2012)
- classification
Calculations associated with syntactic type but not based on it
<table>
<thead>
<tr>
<th>Filename</th>
<th>StartTime</th>
<th>EndTime</th>
<th>Duration</th>
<th>StartValue</th>
<th>EndValue</th>
<th>Absolute Difference</th>
<th>Change across time (Hz/msec)</th>
<th>Movement</th>
<th>ActualF0Range</th>
<th>Sentence #</th>
<th>Clause analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>057fc30_I_shure</td>
<td>136.44</td>
<td>137.01</td>
<td>0.57</td>
<td>236.31</td>
<td>172.81</td>
<td>63.5</td>
<td>0.111411</td>
<td>fall</td>
<td>MM</td>
<td>s30</td>
<td>1.0.0.0.0.6.</td>
</tr>
<tr>
<td>057fc30_I_shure</td>
<td>137.01</td>
<td>137.06</td>
<td>0.05</td>
<td>172.81</td>
<td>221.4</td>
<td>48.59</td>
<td>1.079783</td>
<td>rise</td>
<td>MM</td>
<td>s30</td>
<td>1.0.0.0.0.6.</td>
</tr>
<tr>
<td>057fc30_I_shure</td>
<td>137.06</td>
<td>137.29</td>
<td>0.23</td>
<td>221.4</td>
<td>255.44</td>
<td>34.04</td>
<td>0.14798</td>
<td>rise</td>
<td>MH2</td>
<td>s30</td>
<td>1.0.0.0.0.6.</td>
</tr>
<tr>
<td>057fc30_I_shure</td>
<td>137.29</td>
<td>137.78</td>
<td>0.49</td>
<td>255.44</td>
<td>179.93</td>
<td>75.5</td>
<td>0.153578</td>
<td>fall</td>
<td>H1M</td>
<td>s30</td>
<td>1.0.0.0.0.6.</td>
</tr>
<tr>
<td>057fc30_I_shure</td>
<td>137.78</td>
<td>138.32</td>
<td>0.54</td>
<td>179.93</td>
<td>186.13</td>
<td>6.2</td>
<td>0.011479</td>
<td>stagnant</td>
<td>MM</td>
<td>s31</td>
<td>1.2.0.0.0.4,6,9.</td>
</tr>
<tr>
<td>057fc30_I_shure</td>
<td>138.32</td>
<td>138.42</td>
<td>0.1</td>
<td>186.13</td>
<td>229.14</td>
<td>43.01</td>
<td>0.452767</td>
<td>rise</td>
<td>MM</td>
<td>s31</td>
<td>1.2.0.0.0.4,6,9.</td>
</tr>
<tr>
<td>057fc30_I_shure</td>
<td>138.42</td>
<td>139.03</td>
<td>0.61</td>
<td>229.14</td>
<td>168.1</td>
<td>61.05</td>
<td>0.10008</td>
<td>fall</td>
<td>ML1</td>
<td>s31</td>
<td>1.2.0.0.0.4,6,9.</td>
</tr>
<tr>
<td>057fc30_I_shure</td>
<td>139.03</td>
<td>139.08</td>
<td>0.05</td>
<td>168.1</td>
<td>198.17</td>
<td>30.07</td>
<td>0.6014</td>
<td>rise</td>
<td>L1M</td>
<td>s31</td>
<td>2.0.0.1.0.6.</td>
</tr>
<tr>
<td>057fc30_I_shure</td>
<td>139.08</td>
<td>139.2</td>
<td>0.12</td>
<td>198.17</td>
<td>165.49</td>
<td>32.68</td>
<td>0.272326</td>
<td>fall</td>
<td>ML1</td>
<td>s31</td>
<td>2.0.0.1.0.6.</td>
</tr>
<tr>
<td>057fc30_I_shure</td>
<td>139.2</td>
<td>139.56</td>
<td>0.36</td>
<td>165.49</td>
<td>169.56</td>
<td>4.07</td>
<td>0.011147</td>
<td>stagnant</td>
<td>L1M</td>
<td>s31</td>
<td>2.0.0.1.0.6.</td>
</tr>
<tr>
<td>057fc30_I_shure</td>
<td>139.56</td>
<td>139.63</td>
<td>0.07</td>
<td>169.56</td>
<td>201.24</td>
<td>31.68</td>
<td>0.452586</td>
<td>rise</td>
<td>MM</td>
<td>s31</td>
<td>2.0.0.1.0.6.</td>
</tr>
<tr>
<td>057fc30_I_shure</td>
<td>139.63</td>
<td>140.2</td>
<td>0.56</td>
<td>201.24</td>
<td>208.74</td>
<td>7.5</td>
<td>0.013276</td>
<td>stagnant</td>
<td>MM</td>
<td>s31</td>
<td>2.0.0.1.0.6.</td>
</tr>
</tbody>
</table>

5 pitch levels: L2, L1, M, H1, H2
Slope (Hz/ms)
Absolute values
Syntactic types vs. F0

Formal type 13 (unfinished clause)
- Fall: 6%
- Rise: 23%
- Stagnant: 70%

Informal type 13 (unfinished clause)
- Fall: 14%
- Rise: 25%
- Stagnant: 61%
Syntactic types vs. F0

Formal type 2 (main clause missing)
- 44% fall
- 40% rise
- 16% stagnant

Informal type 2 (main clause missing)
- 56% fall
- 29% rise
- 16% stagnant
Syntactic types vs. F0

Formal type 3 (subord. clause missing)
- Fall: 70%
- Rise: 23%
- Stagnant: 8%

Informal type 2 (subord. clause missing)
- Fall: 56%
- Rise: 29%
- Stagnant: 16%
Detection of intensity change

methodology based on the calculation of the trend of pitch movement (currently being implemented)
Detection of accent/stress

- based on Hunyadi 2002
- PET: pitch and energy over time
- accent/stress is the result of the interaction of pitch and intensity: relative prominence
- absolute PET-value + duration
Ő látta?
he saw
‘Did HE see it?’
‘Did he SEE it?’
Kati mondja.
Kate says
‘It is Kate who says it.’

mondja
‘says it.’
THANK YOU!

http://hucomtech.unideb.hu/hucomtech