Human Communication 1
Lecture 21

Annotation of Speech Data: Monologue and Dialogue

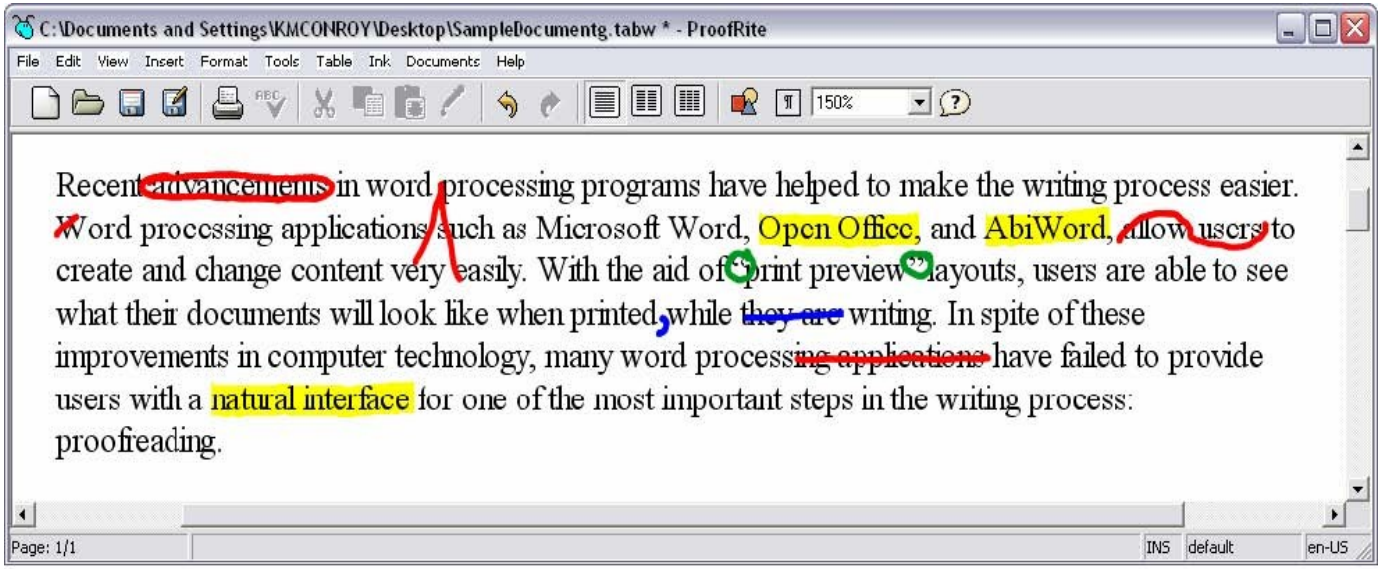
What is annotation?

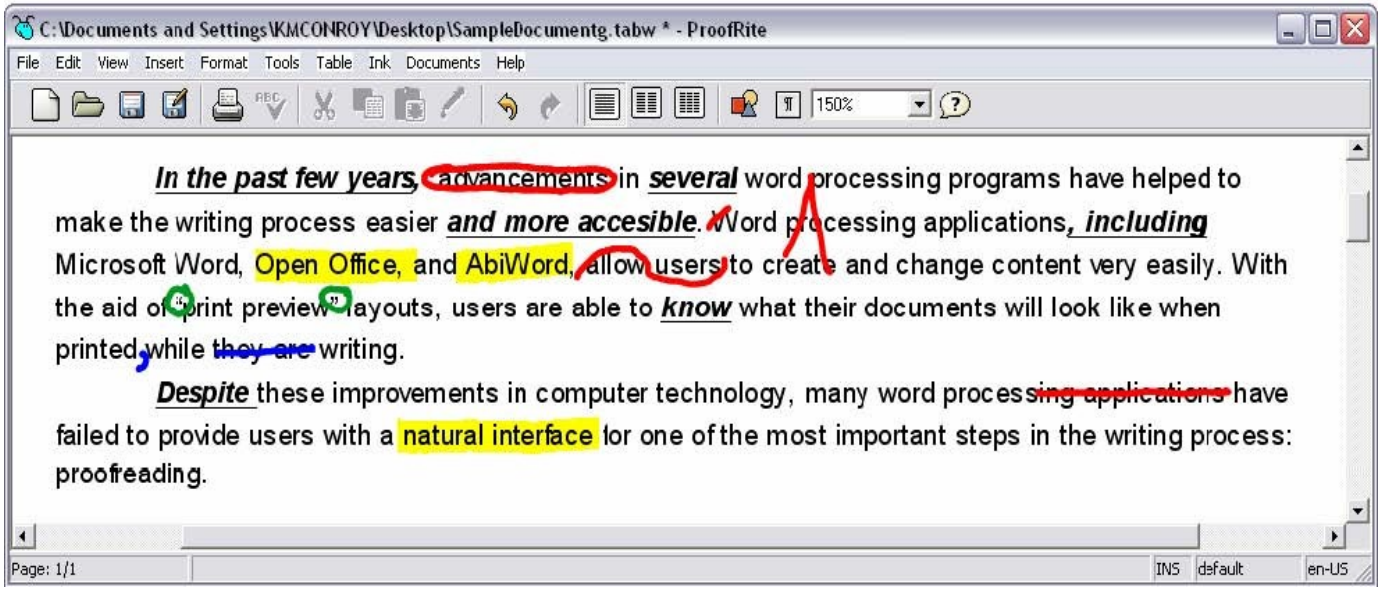
- The word annotation is defined as:

  The act or process of furnishing critical commentary or explanatory notes

Annotating text

- When we annotate text we usually add notes to an already written text, these notes can contain comments or various forms of *meta-data*





Annotating Speech Data

- First of all:

  You are *transcribing* the sounds of speech in monologue or dialogue

  - There are different ways to transcribe speech data:

    Phonetic

    Orthographic

- Secondly:

  you can *annotate* what you transcribed

What do I need to annotate speech

- To be able to annotate any speech data you need:

  A spoken language corpus

- Will any kind of spoken language corpus do it for me?

- No

  You need a corpus that reflects e.g the problem you are working on, the language you want to research, etc.

What Spoken Language Corpus?

- A foreign language, for example a dying indigenous American dialect, that has been sound-recorded and will now be phonetically transcribed with annotation added
- A patient with a certain motor speech disorder: to analyse the disorder better you record the speech of your patient and transcribe it phonetically
- Large amounts of speech data for use in developing computer systems that understand human speech
- ...

Purpose of Spoken Language Corpus

Study aspects of language with your transcribed corpora

– In addition, there are studies of:

- Discourse
- Phonology
- Dialect
- Diachronic change
- ...

Creating a spoken language corpus

- First of all you record speech through a microphone (or microphones) onto a storage medium (tape, hard drive, etc.)
- You have to decide who you want to record and for what p u r p ʊ s e: therapy, linguistics (e.g. preserving) or speech processing
- Purpose may dictate the quality of recording needed, number and type of microphones, etc.
- In most cases, you ideally need more than the data from a single speaker (or pair, etc.), which can sometimes be problematic

How bring spoken data into a corpus?

Digitise sound/video

- Transcribe speech
  - Choose transcription
    - Phonetic
    - Orthographic
- Time code transcription for alignment with sound file
- Annotate as required
  - Extra-linguistic features (laughter, coughing)
  - Semi-linguistc features (meaningful sounds)
  - Gestures
  - Grammatical features

Monologue

- Speech data of patients with a speech and language disorder.
- Speaker of a rare or dying language/dialect
- Speaker reading text from a document

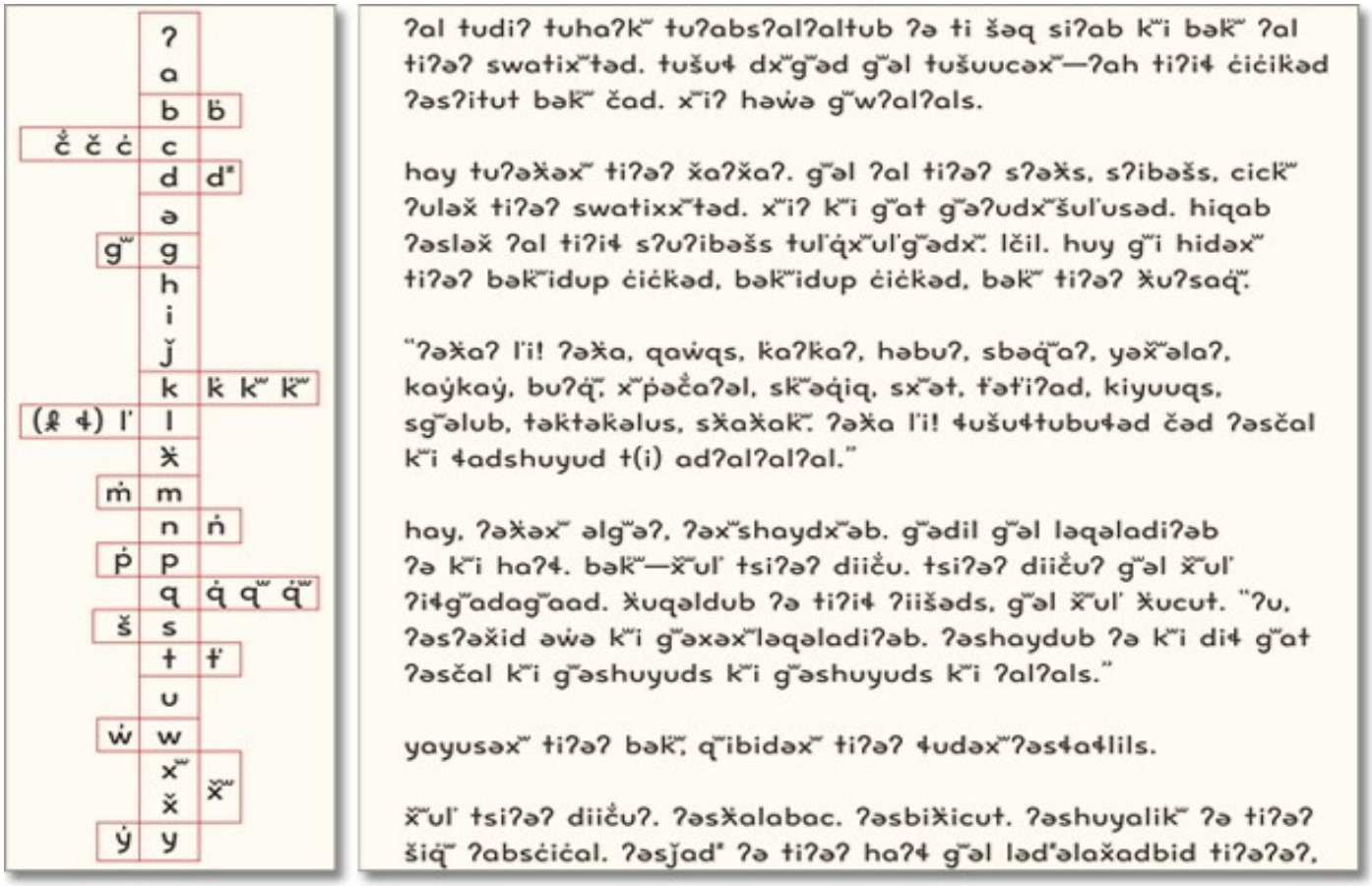
Example 1

- Motor speech disorder:

Cluttering

http://www.youtube.com/watch?v=HhMWHAW8v4Y

Example 2



In the state of Washington,type designer Juliet Shen has been working with a Native American tribe, called Tulalip, to create a font for the Lushootseed language. This is based on a corpus of transcribed Lushootseed speech.

See http://www.atissuejournal.com/2009/10/13/designing-a-font-to-preserve-a-vanishing-language/

Example 3: a Dialogue

On 1 February 2007, Richard Warman testified under oath in the *Warman V. Lemire* case before the *Canadian Human Rights Tribunal* hearing (Case T1073/5405) and clearly stated that he never signed up an account on the "Freedomsite message Board"

MS KULASZKA: Did you ever sign up a user account at the message board?
MR. WARMAN: No, I don't believe I did.
THE CHAIRPERSON: I want to be clear I understand that. Did he sign up a user account?
MS KULASZKA: A user account at the message board.
THE CHAIRPERSON: That would enable you to participate in the messaging. Is that what that would do?
MR. WARMAN: Yes. No, I didn't need to. I could access everything that I needed to as a guest.
THE CHAIRPERSON: As a guest, so only to view. But you did not sign up in order to add material to it?
MR. WARMAN: No, I did not.

Lemire Transcript, Volume 4, Page 769

Respondents counsel, Barbara Kulaszka showed Mr. Warman the incontrovertible proof that he indeed did sign up an account and had deceived the tribunal.

MS KULASZKA: If you could go to tab 3, the respondent's binder. Three pages from the back. Do you recognize this user account called Lucy?

MR. WARMAN: Sorry, I stand corrected. Yes, I do. That is an e-mail address that I used.
MS KULASZKA: When did you use it?
MR. WARMAN: It indicates on Saturday, November 15th, 2003, at 1:03 p.m., and on Saturday November, 15th, 2003 at 3:03 p.m.
MS KULASZKA: You made two log-ins?
MR. WARMAN: That's what it states here.

Lemire Transcript, Volume 4, Page 770

Other spoken language corpora

- For example for Speech Processing:

  Interesting dialogue annotation

- Why?

  The computer needs to learn conversational strategies

  The computer needs to learn to distinguish between speech and other noise

  Dialogue is the most natural speech behaviour of a human being

Types of Annotation

2 fundamental types:

- Segmentation and
- Linguistic annotation

Segmentation

- Delimits linguistic elements that appear in the primary data, like:
  - Continuous segments
  - Super- and sub-segments
  - Discontinuous segments
  - Landmarks (e.g. time stamps)

Linguistic Annotation

- It provides linguistic information about the segments in the primary data, e.g. a morpho-syntactic annotation in which a part of speech and lemma are associated with each segment in the data
- Note that the identification of a segment as a word, sentence, noun-phrase, etc. also constitutes linguistic annotation
- Might also identify dialogue moves, rhetorical structure, speech acts, etc.

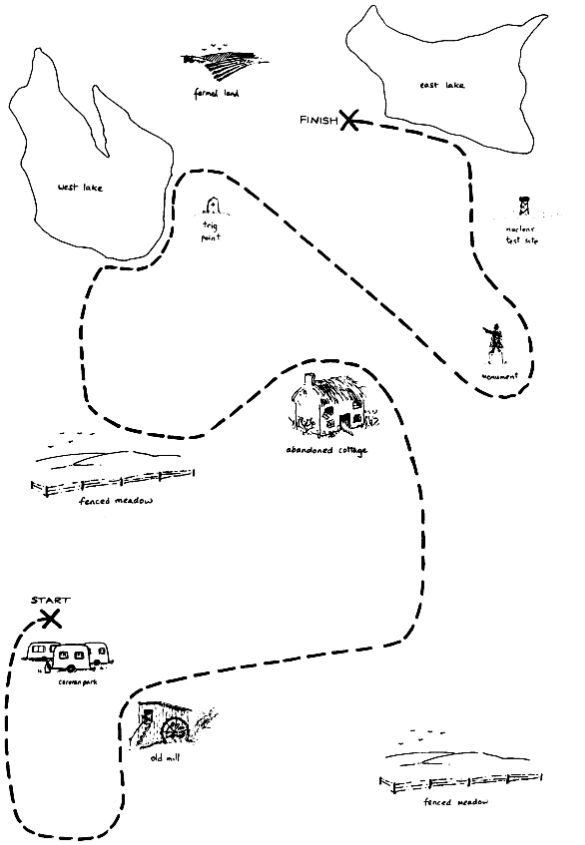
Transcription/Annotation tools

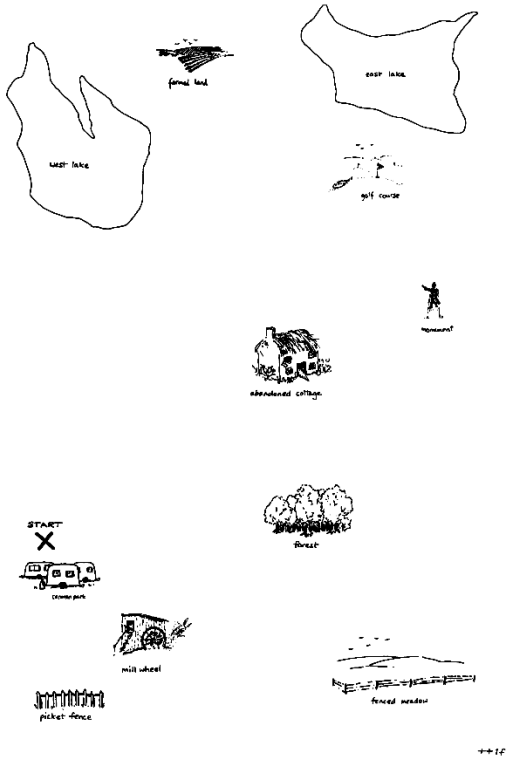
- Nowadays, transcription and annotation are normally computer-assisted
- Tools automate time-alignment of transcription with a recording
- Tools may support many kinds of “stand-off markup”: annotations of basic transcription
- Annotations are usually encoded in XML (eXtensible Markup Language), a highly standardised syntax initially designed for this purpose (though now used for much else)

Example: HCRC Map Task Corpus

- Very large corpus of dialogues based on route-finding; many different annotations

<u who=G n=3> <sfo samp=107715> <bo id=o75a> About half an inch above it, we've got an {lelx} marking start. Have
<u who=F n=4> <sfo samp=208987> Yes.
<u who=G n=5> you got that?
<eo id=o75a>





Communication only Spoken Language?

No! Communication is multimodal

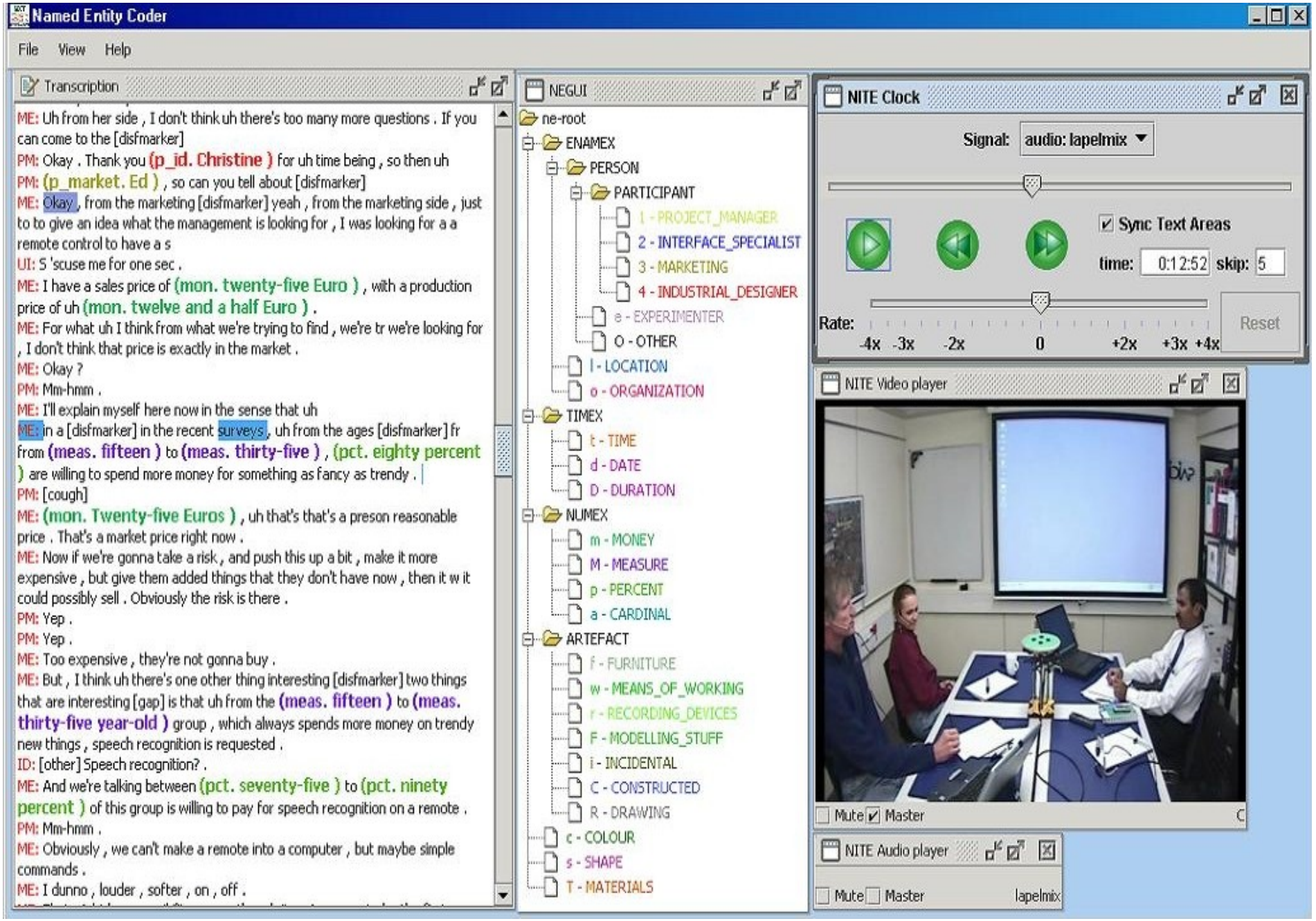
- We have often seen in this course that communication consists of more than speech
- For example:
  - Gesture
  - Facial Expression
  - Eye Gaze/Movement
  - ...

So why so many spoken language corpora?

- Multimodality (speech, gesture, facial expression) is difficult and time consuming to annotate
- Focusing on one modality like speech allows us to transcribe and annotate more thoroughly (though still expensive in time and money)
- Annotated corpora are the essential basis of *machine learning*
  - annotating speech helps us to create automatic annotating systems that can then handle enormous numbers of corpora

Multimodal Corpora

- Nonetheless, there is more and more interest in multimodal communication
- More and more multimodal corpora are being created and annotated
- The rising interest in multimodality in communication is connected to the explosion in processing, storage and use of digital video
- Multimodal annotation is *not necessarily based on a transcription*: may use a dynamic video browser


Named Entity Coder
File View Help
Transcription
ME: Uh from her side , I don't think uh there's too many more questions . If you can come to the [disfmarker]
PM: Okay . Thank you (p_id. Christine) for uh time being , so then uh
PM: (p_market. Ed) , so can you tell about [disfmarker]
ME: Okay , from the marketing [disfmarker] yeah , from the marketing side , just to to give an idea what the management is looking for , I was looking for a a remote control to have a s
UI: S 'scuse me for one sec .
ME: I have a sales price of (mon. twenty-five Euro) , with a production price of uh (mon. twelve and a half Euro) .
ME: For what uh I think from what we're trying to find , we're tr we're looking for , I don't think that price is exactly in the market .
ME: Okay ?
PM: Mm-hmm .
ME: I'll explain myself here now in the sense that uh
ME: in a [disfmarker] in the recent surveys , uh from the ages [disfmarker] fr from (meas. fifteen) to (meas. thirty-five) , (pct. eighty percent) are willing to spend more money for something as fancy as trendy .
PM: [cough]
ME: (mon. Twenty-five Euros) , uh that's that's a preson reasonable price . That's a market price right now .
ME: Now if we're gonna take a risk , and push this up a bit , make it more expensive , but give them added things that they don't have now , then it w it could possibly sell . Obviously the risk is there .
PM: Yep .
PM: Yep .
ME: Too expensive , they're not gonna buy .
ME: But , I think uh there's one other thing interesting [disfmarker] two things that are interesting [gap] is that uh from the (meas. fifteen) to (meas. thirty-five year-old) group , which always spends more money on trendy new things , speech recognition is requested .
ID: [other] Speech recognition? .
ME: And we're talking between (pct. seventy-five) to (pct. ninety percent) of this group is willing to pay for speech recognition on a remote .
PM: Mm-hmm .
ME: Obviously , we can't make a remote into a computer , but maybe simple commands .
ME: I dunno , louder , softer , on , off .
NEGUI
ne-root
ENAMEX
PERSON
PARTICIPANT
1 - PROJECT_MANAGER
2 - INTERFACE_SPECIALIST
3 - MARKETING
4 - INDUSTRIAL_DESIGNER
e - EXPERIMENTER
O - OTHER
l - LOCATION
o - ORGANIZATION
TIMEX
t - TIME
d - DATE
D - DURATION
NUMEX
m - MONEY
M - MEASURE
p - PERCENT
a - CARDINAL
ARTEFACT
f - FURNITURE
w - MEANS_OF_WORKING
r - RECORDING_DEVICES
F - MODELLING_STUFF
i - INCIDENTAL
C - CONSTRUCTED
R - DRAWING
c - COLOUR
s - SHAPE
T - MATERIALS
NITE Clock
Signal: audio: lapelmix
Sync Text Areas
time: 0:12:52 skip: 5
Rate:
-4x -3x -2x 0 +2x +3x +4x
Reset
NITE Video player
Mute Master
NITE Audio player
Mute Master lapelmix