

# CHECKS OF SPEECH ANNOTATION OF AMI MEETINGS

Ing. Kamil CHALUPNÍČEK, Doctoral Degree Programme (1)  
Dept. of Information Systems, FIT, BUT  
E-mail: chalupni@fit.vutbr.cz

Supervised by: Dr. Jan Černocký

## ABSTRACT

This paper deals with checking annotation of AMI meetings and with a procedure of quality assessment applied to the annotation of these meetings. We are presenting semi-automated quality checks implemented on the web.

## 1 INTRODUCTION

AMI (Augmented Multiparty Interaction) is a project running in several countries. Its purpose is improving interaction between people in real time. It is engaged in recording multimodal meeting data. This project consists of several parts. Real meeting is recorded in the first part. Speech data is transcribed into text representation in the second part. This is task for annotators. People naturally make errors and transcriptions are quite exhausting. Automatic processing needs clean output, otherwise we have problems with recognizer training and testing tools [1]. Therefore, all transcriptions must be checked for syntax errors. We created the package of scripts for syntax checks and lexical analysis.

## 2 RULES FOR SPEECH ANNOTATION

Rules are defined for transcription of speech data to text [2]. These are:

- Noise marks. These marks describe sounds that are made using the mouth (or nose) but that do not have standard lexical representations. In these transcriptions the number of these which will be annotated is reduced to four, each of which has a simple symbolic representation. These will be \$ : laugh, % : cough, & : sniff, # : other prominent vocal noise (including creaky voice, aspiration, yawn, throat clearing, tongue click, etc).
- Mispronunciations. If a speaker mispronounces a word and it is possible word, which was intended, the annotator writes the word as it should be spelled and marks it with

an asterisk after the last letter, e.g. “spaghetti\*”. If it is not possible to determine which word was intended, the annotator transcribes what he/she was hearing and marks it with parentheses, e.g. “(fligop)”.

- Numbers. Number sequences must be spelled out, e.g. “forty four” not “44”.
- Spelling. Acronyms should be spelt as they are pronounced, e.g. “NASA” or “U\_S\_A\_”. When a letter sequence is used as part of a word, the annotator adds the inflection after the underscore: “They I\_D\_ed him”. Other possible variant is: “I\_D\_’ed”, “U\_S\_A\_s”, “U\_S\_A\_’s”.
- Unfinished words. If a speaker does not finish a word, and the annotator thinks he/she knows what the word was, he/she can spell out as much of the word as was pronounced by inserting a single dash as the last letter of the word, e.g. “I thin basic-”.
- Punctuation. Punctuation should be limited in the corpus. The annotator should only use commas, full stops (periods), and question marks to punctuate a ‘sentence’. The only exception is dashes, in compound words that are traditionally written with them, e.g. ‘passer-by’, but annotator makes sure there is no space between the dash and the words.
- Unfinished sentences. If the speaker is cut off or doesn’t finish his/her sentence, a dash should be inserted at the end of statement, e.g. “I was going to do that, but then -” (remember, if there is no space between the last word and the dash, this indicates that the word was not finished).
- Fully unintelligible words. If after listening to a “word” several times the annotator is still not sure what it is, he/she should mark the word as “(??)”.
- Non-English words. If the sound is a word or phrase in foreign language, then if annotator can transcribe the sounds, he/she does so, marking them on both ends with the carat sign “^”. If the annotator does not want to hazard a guess, and he/she is certain it is a non-English word, he/she marks the sounds as follows “^(??)^”.
- Others.
  - Standard spoken language should be transcribed as it is spoken, e.g. “gonna” not “going to”, “wanna” not “want to”, “kinda” not “kind of”, etc.
  - Annotators should avoid word abbreviations, i.e “doctor” not “Dr”, and “mountain” not “Mt”.
  - Annotators use normal capitalization on proper nouns and at the beginning of sentences.
  - While punctuation should generally follow (simplified) standard English usage, particular care should be taken at two locations: a) at regions of disfluency, where the speaker interrupts himself to correct or restart or repeat, use a dash (if there is not already a dash from a word fragment), e.g. “I just meant - I mean ...”, and b) at the end of a speech segment, punctuation should be used

to make it clear whether the next turn is a continuation of the current one. If the speaker continues with the same utterance, the annotators punctuate as they would if there was no break (including potentially having no punctuation at all at the end of the initial turn); if the speaker breaks off and does not continue the sentence in this next turn, then they indicate this with a dash.

### 3 ANNOTATION QUALITY ASSESSMENT

We present our annotation processing procedure which is designed for achieving of maximal quality of annotations. The work is based on previous work by Černocký and Pollák [3]. We are providing several automated, semi-automated, and manual checks in the following steps:

- I. Each annotator is working on a set of meetings. Usually, one single meeting is checked by the automatic procedure.
- II. The first step in quality assessment is based on syntax test. The most evident errors should be found here, typically usage of allowed characters, correct usage of special marks, empty annotation fields, etc.
- III. As the second step of quality check, the semi-automated lexical test is performed. This test is described by flow chart in Fig.1a. and it contains the following principal substeps:
  - A. Every transcription file is compared with the reference lexicon with already checked and approved words and differential dictionary is created. Differential dictionary contains only unknown words. This test is implemented by a call of ispell [4], the reference dictionary is composed of standard ispell dictionary and AMI dictionary where approved new words are stored.
  - B. The annotator manually checks unknown words, and errors are marked. The script generates the transcription file with markers at places of 'suspect' words so that they can be easily checked and eventually corrected.
  - C. Listening of utterances with marked strange words is done by the annotator and bad words are corrected.
  - D. Corrected new words are added to reference AMI lexicon.
- IV. The annotation package is accepted only if both tests are successfully passed.

The implementation is in Perl language, because Perl is suitable for working with text.

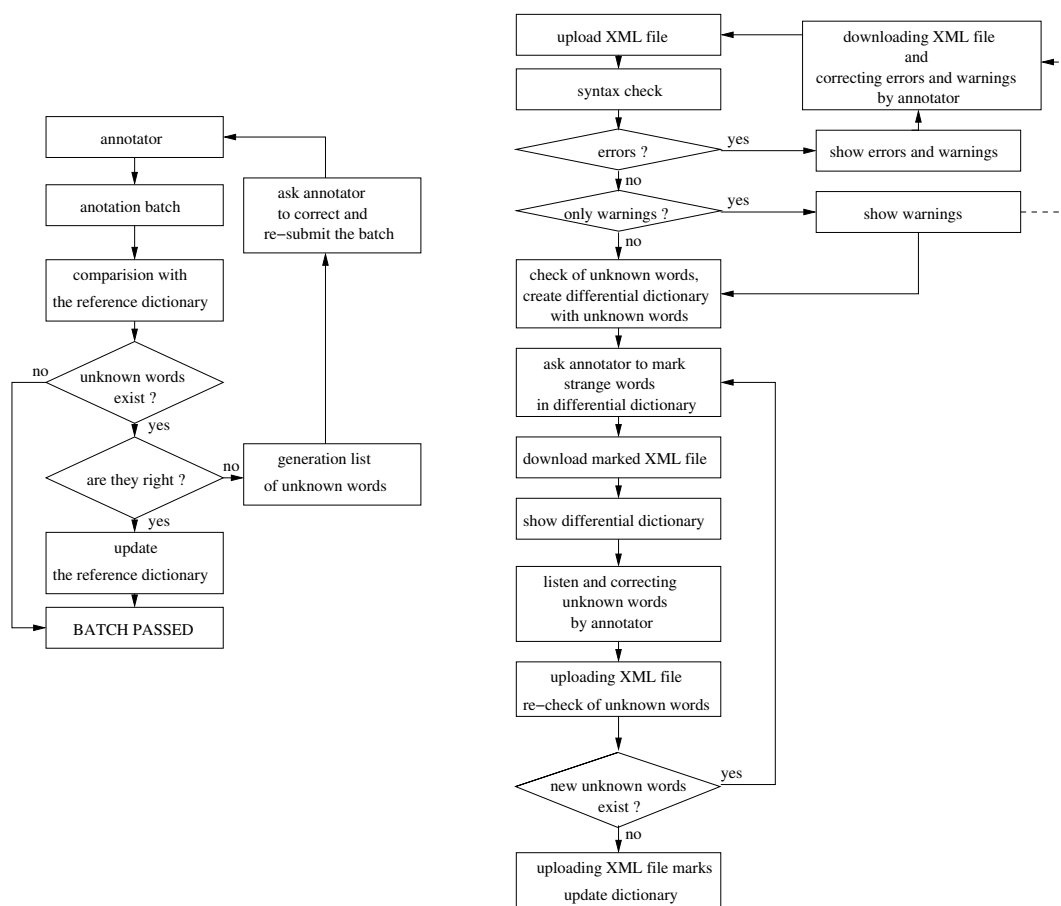


Figure 1: a) Lexical analysis process b) Web checks process

#### 4 WEB INTERFACE FOR THE QUALITY ASSESSMENT TOOLS

Syntax checks and lexical checks are accessible through the web (Fig. 1b.). At the beginning, the annotator uploads annotation file with speech transcription on the web site. File is checked and annotator downloads the annotation file with marks (Fig. 2.). These marks are at places with errors. Annotator corrects these errors and he/she re-uploads annotation file for re-check. If annotation file contains errors now, then they must be corrected. If annotation file doesn't contain errors, the procedure proceeds to lexical analysis. Annotation file is compared with the reference lexicon and differential dictionary is created. The annotator manually checks unknown words, listening of utterances with marked strange words is done and bad words are corrected. Correct new words are added to reference lexicon.

#### 5 CONCLUSIONS

In this paper, we are presenting set of tools for semiautomatic checks of speech annotations. The most important results of this work are:

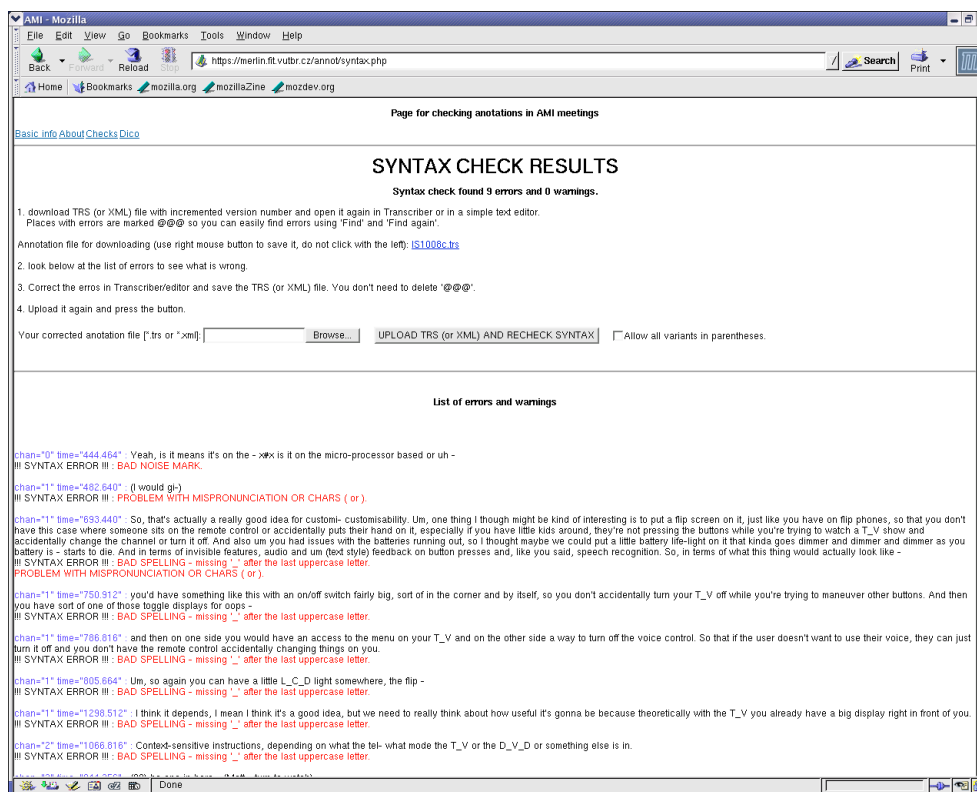


Figure 2: Check page on web

1. The procedure for annotation quality assessment was defined.
2. With checks on the web, the annotation of one meeting will be accelerated.

## ACKNOWLEDGEMENTS

This work was partially supported by EC project Augmented Multi-party Interaction (AMI), No. 506811 and Grant Agency of Czech Republic under project No. 102/05/0278. This article has been supported by project FRVŠ No. 3449/2005.

## REFERENCES

- [1] Young, S.: The HTK Book, Entropics Ltd., 1999
- [2] More, J., Kronenthal, M., Ashby, S.: Guidelines for AMI speech transcriptions, IDIAP, Switzerland, 2005
- [3] Pollák, P., Černocký, J.: Orthographic and Phonetic Annotation of Very Large Czech Corpora with Quality Assessment, LREC Lisbon, 2004
- [4] Nelson, B., Atkinson, K.: Ispell/Aspell manual page.