

HQSYN16 - Task #3709

Task # 3672 (Assigned): RA1d - Automatic cleaning of speech corpora

Task # 3704 (New): Detection and correction of prosodic structures

Merge ASF files (segmentations) and SNT files (annotations)

21.01.2016 11:06 - Hanzlíček Zdeněk

Status:	Closed	Start date:	05.04.2016
Priority:	Normal	Due date:	
Assignee:	Hanzlíček Zdeněk	% Done:	100%
Category:		Estimated time:	0.00 hour
Target version:	RA1: Analysis of artifacts in synthetic speech		
Description Merge ASF files (segmentations) and SNT files (annotations): 1. Find differences between words in ASF and SNT and update SNT in SVN repository. 2. Add new columns into ASF: punctuation and pronunciation.			
Subtasks: Task # 3851: Merge ASF and SNT files for MM voice (Slovak) Closed			
Related issues: Related to HQSYN16 - Task #3761: Create script for conversion ASF to SNT Closed 23.02.2016			

History

#1 - 21.01.2016 12:23 - Hanzlíček Zdeněk

Add missing non-speech events from SNT into ASF with zero duration.

#2 - 21.01.2016 16:30 - Hanzlíček Zdeněk

Use upper-case chars in words in ASF (copy from SNT).

#3 - 25.01.2016 16:25 - Hanzlíček Zdeněk

- % Done changed from 0 to 70

Scripts for this task are placed in SVN repository ARTIC_UTILS/trunk/hmm_synth/LabLight.

A simple description of the merging procedure:

1. Script diff_asf_snt.py performs a simple comparison between ASF and SNT files, prints out suspicious inconsistent utterances (pauses are ignored). This comparison reveals only some basic types of inconsistency.
2. Manual correction of SNT file (when needed).
3. Script merge_asf_snt.py merges ASF and SNT files, a new ASF file is created.
 1. Join verbs and enclitic "li" into one word (e.g. bude-li).
 2. Use words from SNT (with capital letters).
 3. Add punctuation and pronunciation columns.

Two ASF files were processed (voices MR and TJ). New ASF and SNT files are placed in ARTIC directory /artic/Experiments/asf.snt.merge.

#4 - 29.01.2016 14:56 - Hanzlíček Zdeněk

- % Done changed from 70 to 80

Two more voices were processed: KI and AJ.

NOTE: Several specific corrections had to be done manually. Thus, the new ASFs should replace the default ASFs in the SVN repository. Otherwise, sooner or later, we could have two parallel inconsistent ASF versions.

#5 - 23.02.2016 09:53 - Hanzlíček Zdeněk

- Related to Task #3761: Create script for conversion ASF to SNT added

#6 - 24.06.2016 15:41 - Hanzlíček Zdeněk

- *Status changed from Assigned to Resolved*

Summary

Voices with merged ASFs:

- Czech voices: AJ, JS, KI, MR, SK, TJ
- Slovak voice: MM

ASF file for the Czech female voice PP was not merged, since the annotation file is not available.

#7 - 24.06.2016 15:42 - Hanzlíček Zdeněk

- *Status changed from Resolved to Closed*