

SPEX QQC REPORT

TITLE DATABASE: BDLEX-50000, v2.1
 DATABASE OWNER / PRODUCER: IRIT – Université Paul Sabatier
 ELRA CATALOGUE NUMBER: S0004

AUTHORS OF QQC REPORT: Henk van den Heuvel
 DATE: 11 Sep. 2002
 VERSION: 1.0

SUMMARY SHEET:

Database part	Applicability (y/n)	Quality value		
		*	**	***
1. Documentation	Y		**	
2. Format	Y			***
3. Design & contents	N			
4. Speech signals	N			
5. Annotation files	N			
6. Speakers	N			
7. Environments	N			
8. Transcriptions	N			
9. Lexicon	Y			***

For each applicable part a star assessment is given.

1. * This value is given if there is not a proper and reliable fit of the contents of the SLR and the information about this part as presented in the documentation.
2. ** This value is given if the documentation well accounts for the contents of the SLR. Some small deviations are permitted.
3. *** This value is given if there is no mismatch between the documentation and the contents of the SLR.

1 Quick Quality Check Report

1.1 Documentation

The most important topics should be covered and clearly described in the documentation:

The documentation is provided in the file \DOCUMENTATION\Lisezmoi-BDLex-V2.en.doc and also in a French translation.

- db layout and media

OK

- application potential for the SLR

⇒ Not mentioned

- directory structure and file names

OK

- recording equipment

N/A

- design and contents of the recordings

⇒ The origin of the words (in terms of text corpora) is not mentioned

- coding and format of the speech files

N/A

- contents and format of the label files

N/A

- speakers

N/A

- recording environments distinguished

N/A

- transcription conventions

N/A

- lexicon: format and transcriptions included
OK, there is an extensive description of transcription conventions.

1.2 Format

- The file names and directory structure should correspond to the documentation
OK

1.3 Design and contents

- All mandatory items according to the documentation should be included
N/A
- Number of effectively missing files per corpus item should be appropriate
N/A

1.4 Speech signals

- For 2 CDs of the SLR acoustic measurements on the speech files will be made, and the results reported. The acoustical measurements involved are:
 - o Clipping rate
 - o SNR
 - o Mean amplitude*N/A*

1.5 Annotation files

- A random selection of the annotation/label files will be checked. They should be
 - o Readable
 - o Contain the information described in the documentation*N/A*

N/A

1.6 Speakers

- Speaker distributions should be in agreement with documentation
N/A

1.7 Environments

- Environment distributions should be in agreement with documentation
N/A

1.8 Transcription

- how many speech files miss an orthographic transcription?
N/A
- All non-speech markers should be described in documentation
N/A

1.9 Lexicon

- The correct set of phone symbols should be used (according to documentation)
OK
- All words in the (orth.) transcriptions should be present in the lexicon
OK

The contents of the corresponding files in the directories BDLex and BDLex-syll are identical, except for the phonemic transcriptions (as expected).

The contents of the phonemic fields of the corresponding files BDLex and BDLex-syll are identical, except for the syllable boundaries (as expected).

The PHONO field contains only the phonemes indicated in section 4 of the documentation file.

*⇒ The FPH field contains one value that is not explained in section 5 of the documentation file:
j", only in "gentil;j"*

The CS field contains only values that are explained in section 6.1 of the documentation file.

Field 5 (gender, number, person) contains only values that are explained in section 6.2 of the documentation file.

Field 6 (syntactic subcategories) contains only values that are explained in section 6.3 of the documentation file.

1.10 Other remarks

⇒ *The lexicon is rather redundant. All information appears about two times because the directory BDLex-syll adds the syllable boundaries in the phonetic transcriptions. It would be more efficient to have included the phone transcription with syllable boundaries as extra field in the files in the BDLex directory.*

2 Recommendations

None