

CORPUS INFORMATION FORM

Check (x) or fill in the blank (_____) if appropriate.

1. Name of the Corpus: PSCLT (Singapore Primary School Chinese Language
Text)

2. IPR Holder: Chinese and Oriental Language
Information Processing Society,
Singapore and Institute of
Computational Linguistics, Peking
University, China

3. Corpus Type:

- Speech Corpus ()
- Text Corpus (x)

4. If it is a speech corpus:

- Purpose:
 - ASR ()
 - TTS ()
 - Other, please specify _____
- Language:
 - Putonghua (SC) ()
 - Mandarin in Taiwan (TC) ()
 - Cantonese in HK (TC) ()
 - Other, please specify _____
- Style:
 - Read speech ()
 - Spontaneous speech ()
 - Conversational speech ()
 - Other, please specify _____
- Channel:
 - Close-talk Microphone ()
 - Telephone ()
 - Mobile phone ()
 - Other, please specify _____

-
- Sampling Rate: _____ k Hz
 - Sampling Precision:
 - PCM (), _____ bits per sample
 - A-law ()
 - Miu-law ()
 - Other, please specify _____
 - Corpus size: _____ hours _____ speakers
 - SNR level: _____ dB
 - Transcriptions:
 - Character tier (SC) ()
 - Character tier (TC) ()
 - Canonical Pinyin tier ()
 - Other canonical pronunciation tier, please specify _____
 - Surface form IF tier ()
 - Surface form IPA tier ()
 - Surface form SAMPA-C tier ()
 - Other surface form tier, please specify _____
 - Other transcription, please specify _____
 - Other transcription, please specify _____
 - Other transcription, please specify _____

5. If it is a text corpus:

- Language:
 - SC (x)
 - TS ()
 - Other, please specify _____
- Domain:
 - Culture ()
 - Economy ()
 - Military ()
 - News ()
 - Politics ()
 - Sciences ()
 - Sports ()
 - Other, please specify Primary school Chinese language text
 - Other, please specify _____
 - Other, please specify _____
 - Other, please specify _____
 - Other, please specify _____
 - Other, please specify _____
- Corpus size: 0.035 Mega characters
- Tag Information:
 - Word segmentation: (X)
 - Part-of-Speech (X)

- Other, please specify Syntax_____
- Other, please specify _____
- Other, please specify _____
- Other, please specify _____

6. A brief Description of the Corpus:

We hand tagged 12 volumes of Chinese language text used in the Singapore schools. It covers levels from Primary 1 to Primary 6 and contains about 35,000 simplified Chinese characters. In part-of-speech, we follow the system from Institute of Computational Linguistics, Peking University. Syntax trees are given.