

CORPUS INFORMATION FORM

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

1. **Name of the Corpus:** CCC-CEB1 (Chinese Event Bank: Part 1)

2. **IPR Holder:** Chinese Corpus Consortium (CCC)

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: _____ 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 (x)
 - Economy ()
 - Military ()
 - News (x)
 - Politics ()
 - Sciences (x)
 - Sports ()
 - Other, please specify literatures
 - Other, please specify _____
 - Other, please specify _____
 - Other, please specify _____
 - Other, please specify _____
 - Other, please specify _____
- Corpus size: about 1 Mega characters
- Tag Information:
 - Word segmentation: (x)
 - Part-of-Speech (x)
 - Other, please specify: sense codes of target verbs to describe special event situations
 - Other, please specify: syntactic and semantic tags of each event frame arguments
 - Other, please specify _____
 - Other, please specify _____

6. A brief Description of the Corpus:

This corpus was designed Chinese event detection tasks. We selected the real text sentence of some basic event situations, including possession transferring, existence, space and time transferring, and annotated detailed event description information for each sentence. In CEB-1, all sentences are extracted from Tsinghua Chinese Treebank (TCT).

Firstly, for each target verb (verb phrase), we annotate its suitable sense codes defined in four semantic lexicons, including a situation network, a computer-oriented common-sense KB (HowNet), a Chinese thesaurus dictionary (CiLin) and the Contemporary Chinese dictionary (Xianhan).

Secondly, we annotate the suitable semantic role tags for the core or peripheral arguments of the special events evoked by the target verbs (or verb phrases). In CEB-1, we designed a semantic argument tag set with 5 tags (x, y, z, L, O), three ones for core arguments and two ones for peripheral arguments.

This corpus can be used for target verb Word Sense Disambiguation (WSD) and sentence Semantic Role Labeling (SRL) in real world texts.