The Chinese/English Political Interpreting Corpus (CEPIC)

- Brand new: developed in 2019
- 16 corpora included (i.e. HK SAR Policy Addresses; HKPA)
- 6.5 million word tokens
- Data from 1997-2017
- Languages including Cantonese, Putonghua and English
- Transcripts of speeches by top political figures from Hongkong, Beijing, Washington DC and London & translated/interpreted texts of the above
- Different prosodic and paralinguistic features included for the study of spoken languages as well as interpreting
- Users can also download search results from the corpus for their own teaching/research purposes.

Annotation

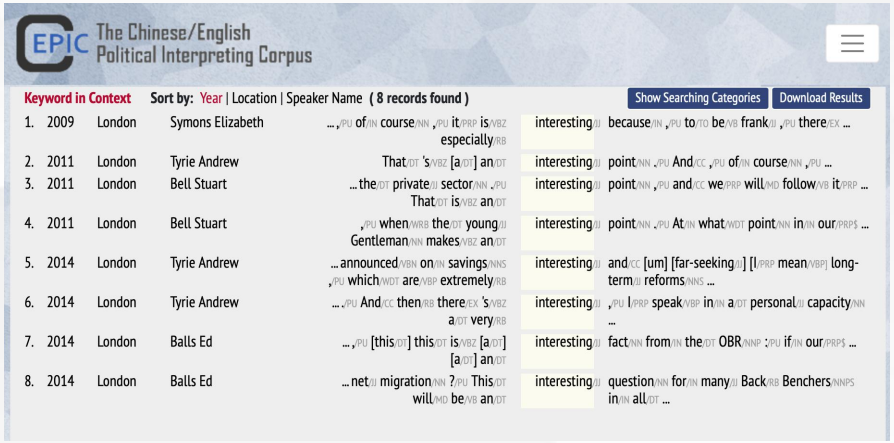
- POS tagging with Stanford CoreNLP 3.9.2 (Manning et al. 2014)
- Speeches of CEPIC were manually revised or transcribed based on audios/videos with the speeches and their interpreting.
- Apart from following a standardised process, the transcription of CEPIC aims to represent the spoken text as close as it was delivered.
- Text and audio/video links were also included for those who may be interested in the sources of the speeches.

An example of annotation:

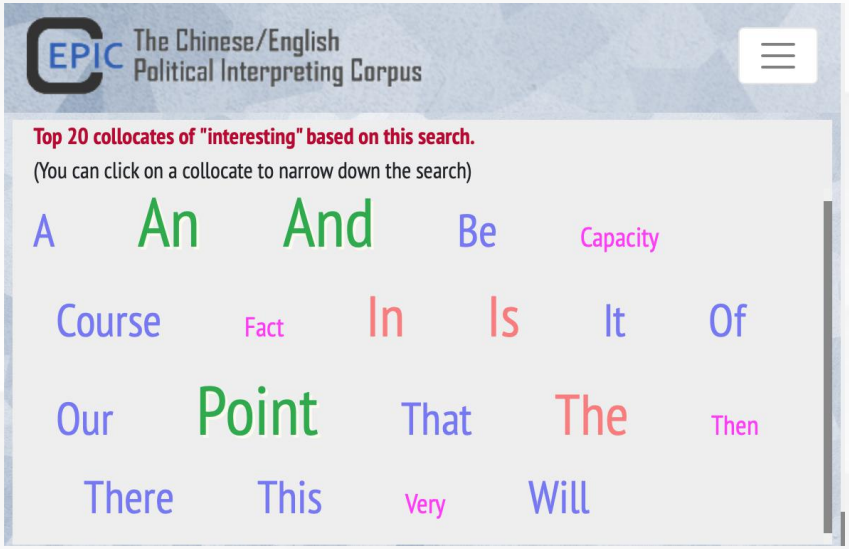
Raw	Annotated
So that is the big difference in our approach and the approach that I think might have been debated about. (Press Conference of US Budget Speech, 1997-02-06)	[er] So [that] that is the big difference [er] in our approach and the approach [er] that [er] I think [er] might have been debated about. (Press Conference of US Budget Speech, 1997-02-06)

Usage

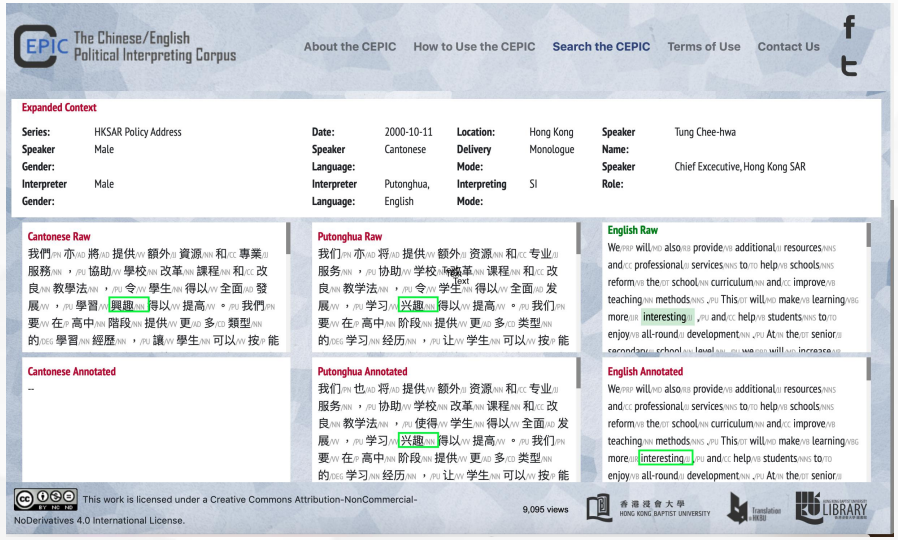
- Keyword Search: A lexical associative function.



- Word collocation



- Expanded Keyword in context



Usage

Sample study: Pragmatic competence in C-E Retour interpreting of political Speeches

- Retour interpreting: From mother tongue to non-mother tongue interpreting.
- To address what and how pragmatic competence (PM) is used in interpreting.
- The use of a set of PMs was examined, including syntactic markers, lexical markers, contrastive makers, and elaborative markers.

Category	Selected PMs
Syntactic markers	I know
	I think
	I suppose
Lexical markers	actually
	kind of
	sort of
	then
Contrastive markers	but
	instead of
	however
Elaborative markers	above all
	what is more
	in other words

- Two sub corpora were compared:

(1)The Corpus of Interpreted Political Speeches from Chinese to English (CIPSCE): interpreted English

(2)The Corpus of English Political Speeches (CEPS): native English

- Findings: A general underuse of most PMs, in particular syntactic markers and lexical markers, in the CIPSCE than in the CEPS

References

- Pan, J. (2019, 5-6 September). The Chinese/English Political Interpreting Corpus (CEPIC): A new electronic resource for translators and interpreters. Paper presented at The Second Workshop on Human-Informed Translation and Interpreting Technology (HiT-IT 2019), Varna, Bulgaria.

- Pan, J., & Wong, T. M. (2019). Developing Pragmatic Competence in Chinese–English Political Retour Interpreting: A Corpus-Driven Exploratory Study of Pragmatic Markers., inTRAlinea Special Issue: New Insights into Translator Training.

- Pan, J. (2019). The Chinese/English Political Interpreting Corpus (CEPIC). Hong Kong Baptist University Library, [Retrieved Date], Accessed from https://digital.lib.hkbu.edu.hk/cepic/

- Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J. R., Bethard, S., & McClosky, D. (2014, June). The Stanford CoreNLP natural language processing toolkit. In Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations (pp. 55-60).

ISLRN Metadata

Title CEPIC	CEPIC
Full Official Name	The Chinese/English Political Interpreting Corpus
Resource Type	Primary text
Source/URL	https://digital.lib.hkbu.edu.hk/cepic
Format/MIMEType	text/xml
Size/Duration	6,393,994 tokens; 116.060 types
Access Medium	on-line
Description	A new electronic and open access resource developed for translators and interpreters, especially those working with political text types.
Version	1.0
Media type	Text
Language(s)	Chinese (Cantonese, Putonghua); English
Resource Creator	Dr. Jun PAN (Associate Professor, Translation Programme, Hong Kong Baptist University; HKBU)
Distributor	HKBU Library
Rights Holder	The Principal Investigator (Dr. Jun PAN) and Hong Kong Baptist University Library
Relation	The Corpus of Political Speeches; The WAW corpus