

# A Corpus-driven Study of Contrastive Markers in Cantonese–English Political Interpreting

*Jun Pan*

Hong Kong Baptist University, Hong Kong  
Baptist University Rd, Kowloon Tong, Hong Kong  
janicepan@hkbu.edu.hk

*Billy Tak Ming Wong*

Open University of Hong Kong, Hong Kong  
30 Good Shepherd St, Ho Man Tin, Hong Kong. Tel.: +852 2711 2100  
billywtm@gmail.com

## Abstract

Political interpreting, as a significant means for foreign language speakers to access a government's official policies, has been regarded as an intensive and stressful task. Any single misinterpretation or misuse of strategy can lead to regional and even international disputes. It will therefore be interesting to study the pragmatic strategies applied by interpreters working in political settings, especially when they render propositions that may sound unfavorable or contrastive to people's presuppositions. In this regard, the use of contrastive markers, an important type of pragmatic markers, serves as an important linguistic indicator of the application of such strategies. Nevertheless, not much has been explored in this aspect. This paper, therefore, studies the use of contrastive markers in the interpreting of policy addresses from Cantonese to English. A parallel corpus, consisting of policy addresses delivered by Chief Executives in Hong Kong (about 0.22 million words) and their English interpretations (about 0.29 million words), was used in the study. The Cantonese contrastive markers — *bat gwo* and *daan (hai)* (comparable in meaning to *however* and *but* respectively in English), and their renditions in English — were compared and analyzed. The two Cantonese contrastive markers were found to correspond to a variety of renditions in English. These findings show how interpreters apply pragmatic strategies when dealing with the extreme situations in political interpreting. They shed light on the development of e-learning for pragmatic competence training of interpreters working in political settings, as well as natural language processing applications for handling such a high-level linguistic feature.

**Keywords:** interpreting, linguistics, e-learning, natural language processing

## 1. Introduction

Political interpreting serves as a significant means for foreign language speakers to access a government's official policies (Gagnon, 2010; Schäffner & Bassnett, 2010). Interpreters of senior government officials therefore often face great challenges and work under extreme stress, as any misinterpretation or misuse of interpreting strategy can lead to regional and even international disputes. Apart from excellent language abilities and a broad knowledge base, interpreters working in political settings need to be sensitive to the pragmatics of political discourse, including but not limited to “*understatement, unspoken assumptions or subtle emphases, innuendo and hedging, or things left unsaid*”, and render them appropriately in the target language (Buri, 2015, para.17). It is therefore important to study the pragmatic strategies applied by interpreters in political settings, especially when they render propositions that may sound unfavorable or contrastive to people's presuppositions. In this regard, the use of contrastive markers — a major type of pragmatic markers — serves as a key linguistic indicator of the application of such strategies. Nevertheless, not much has been explored in this aspect.

In addition, Chinese is often regarded as a language that is typically implicit and pragmatically significantly different from English, which can be evidenced by its politeness maxims of self-denigration, address-term, refinement, agreement and virtue (Gu, 1992). An examination of the use of contrastive markers that signal non-agreement in this language and their renditions in

English can offer special insights to the pragmatics of political discourse and to natural language processing applications with language pairs that are pragmatically significantly different.

Cantonese is a variety of Chinese that is spoken in the Pearl River Delta and is the principal language used in Hong Kong. As a spoken variety of Chinese used in south China, Cantonese is considered demonstrating greater subtlety than Standard Chinese, or Putonghua does, especially at the pragmatic level. Against this background, this study was conducted to investigate the use of contrastive markers in Cantonese–English political interpreting.

## 2. Literature review

### 2.1. Political interpreting

Situated at the intersection between political and media discourse, political interpreting poses great challenges to the interpreters due to its complicated role as “an integral part of political activity” (Schäffner & Bassnett, 2010, p. 13):

Where foreign policy of individual states is concerned, translation becomes relevant, for example, for delivering speeches during state visits. Translations of such speeches are made available on government or embassy websites and are sometimes also published in bulletins or the media. In this way, a government can communicate its political aims and decisions to the outside world.

Gagnon (2010, p. 255) also suggests that political interpreting or translation may show shifts that “had an impact on the target society”.

According to Xu (2000, p. 38, as in Yang, 2012, p. 16), “a diplomatic translator is usually a diplomat, who is required to translate or interpret not only the leaders’ speeches but also their attitude and mood, and even the political atmosphere on the spot”.

Apart from the great challenges at the pragmatic level, interpreting for senior government officials is often regarded as an intensive and stressful task. Buri (2015, para. 14), for instance, illustrates that interpreters working in political or diplomatic settings are “under continuous scrutiny” and may be “easily transformed into scapegoats especially when there are misunderstandings or friction between parties — straightforwardly attributed to misinterpretation”.

Yang (2012, p. 12) emphasizes that diplomatic translators and interpreters should be able to employ the tools of discourse analysis and “analyze the political meaning of the diplomatic language by reading between the lines”. Wang (2008), in particular, calls for a textual perspective in Chinese-to-English translation of political documents to ensure accuracy and effectiveness in the rendition of meaning.

In sum, previous studies on political interpreting have highlighted its great challenges at the pragmatic level and the significance of interpreters’ ability in pragmatic and discourse analysis. However, not much work has been done to investigate interpreters’ pragmatic strategies in dealing with the challenges of rendering the non-propositional meaning in political and diplomatic settings, leaving a significant void to fill in the training of political interpreters in discourse competence and the enhancement of the quality of interpreter-mediated political discourse.

### 2.2. Contrastive markers

Fraser (1996, p. 168) defines pragmatic markers as “the linguistically encoded clues which signal the speaker’s potential communicative intentions”. He provides a four-level scheme to cover the types of pragmatic markers, including basic markers (structural, lexical and hybrid), commentary markers, parallel markers and discourse markers. Discourse markers are further divided into four categories, namely, topic change markers, contrastive markers, elaborative markers, and inferential markers. As a subset of discourse markers, contrastive markers can be illustrated by the use of expressions such as *however*, *but*, *on the contrary*, etc.

Contrastive markers, often used as adversative conjunctions, constitute a general meaning that is “contrary to expectation” (Halliday & Hasan, 1976, p. 250). They signal that “the utterance following is either a denial or a contrast of some proposition associated with the preceding

discourse” (Fraser, 1996, p. 187). Contrastive markers also indicate “the unexpected, surprising nature of what is being said in view of what was said before” (Biber et al., 2002, p. 878).

Two typical contrastive markers, i.e., *but* and *however*, are among the most well researched in this category. In particular, *but* is regarded as signaling denial of expectation and contrast (Blakesmore, 1989). It is less subtle as compared to *however*, which communicates the pragmatic meaning of concession (Quirk et al., 1985), implies the closing of a topic (of a digression) or reintroduces a prior topic (Bublitz, 1988). Corresponding to these two, the expressions *bat gwo* (不過, comparable to *however* in English) and *daan (hai)* (但[是], comparable to *but* in English) constitute the two most representative contrastive markers in Cantonese, with the former more subtle than the latter (see CUHK, 2014).

Despite their important roles in signaling contrastive meaning with different degrees of subtlety, the translation or interpretation of the Cantonese pair of contrastive markers *bat gwo* and *daan (hai)* has been seldom discussed, in particular in the sensitive and subtle political discourse.

### 3. The study

This study aims to investigate the use of contrastive markers in Cantonese–English political interpreting. It focuses particularly on two Cantonese contrastive markers — *bat gwo* and *daan (hai)* — and their closest equivalences in English — *however* and *but*, respectively. The two pairs of contrastive markers are chosen for their representativeness in the two languages.

#### 3.1. Research questions

To uncover the use of contrastive markers in Cantonese and English, the study focuses on the following research questions:

- i. How frequent is the use of the four contrastive markers (two in Cantonese and two in English) in Cantonese and English political speeches?
- ii. How are the four contrastive markers translated/interpreted into another language?

#### 3.2. Corpus data

A parallel corpus was built up as the source of data for the study. The corpus includes the annual policy address speeches from 1997 to 2017, which were delivered in Cantonese by Chief Executives of the Hong Kong SAR Government. Simultaneous interpreting in English was provided during the speeches.

The Cantonese and English texts for the corpus were collected from the government website, which provides the published version of the policy addresses in Chinese and English and the video recordings of the speeches in the two languages. The published version was used as a basis and checked against the video recordings. The features of oral speeches which were removed or amended in the published version, such as the words which occur only in Cantonese speeches but not in standard written Chinese, were recovered and included in the texts for the corpus.

Further processing was performed on the Chinese and English texts. Word segmentation was done for the Cantonese texts using the software tool *SegmentAnt* (Anthony, 2015), so that both the Cantonese and English texts have a consistent form of word segmentation with space used as a delimiter. All the texts were checked manually to ensure that they are aligned to the counterparts in another language at the paragraph level.

#### 3.3. Analysis methods

The Cantonese and English texts were compiled into the parallel corpus using the software tool *AntPConc* (Anthony, 2014). Instances of each of the four contrastive markers were searched and located using the tool, which retrieved both the paragraphs containing the markers and the corresponding paragraphs in another language. Manual checking was performed to identify the translation of the contrastive markers. Then the frequencies of the contrastive markers and their translations were counted.

## 4. Findings

### 4.1. Corpus statistics

Table 1 presents the overall corpus statistics for the Cantonese and English texts. The Cantonese and English parts have a total of about 0.22 million tokens and 0.29 million tokens, respectively. The Cantonese part has a higher type-token ratio (TTR) than the English part — 8.4% and 3.0%, respectively — which means that the former has a larger lexical variety in general.

Table 1. Overall corpus statistics

	<i>Token</i>	<i>Type</i>	<i>TTR</i>
Cantonese	223,082 (word)	18,836 (word)	8.4%
English	291,286	8,855	3.0%

Table 2 shows the frequency of the contrastive markers in the corpus, including the total occurrence frequency and the average frequency per 10K words. Despite that the contrastive markers are the major ones in the two languages, their use in the data varies. The Cantonese marker *bat gwo* was only used about 2 times per 10K on average, comparing with *daan (hai)* which was used more than 12 times on average. The two English markers show less variation in their frequency of use, with *however* being used for about 3.5 times and *but* for about 5 times on average. The results suggest that each contrastive marker has its own patterns of use in each language.

Table 2. Frequency of the contrastive markers in the corpus

	<i>Contrastive markers</i>	<i>Total freq. (all)</i>	<i>Avg. freq. (per 10K)</i>
Cantonese	<i>bat gwo</i>	45	2.02
	<i>daan (hai)</i>	287	12.87
English	<i>however</i>	101	3.47
	<i>but</i>	147	5.05

### 4.2. The renditions of “*bat gwo*” and “*daan (hai)*”

The renditions of *bat gwo* and *daan (hai)* were closely examined by looking into how they were interpreted. Figure 1 lists all the renditions of *bat gwo* and their frequency. The figures show that *bat gwo* was most often interpreted into *however* (frequency=22) — its closest equivalence in English. There were, however, 7 cases that *bat gwo* was interpreted into *but* — its stronger and less subtle correspondence that indicates denial and contrast. Apart from rendering into these two most common English contrastive markers, there were also 5 cases that *bat gwo* was not interpreted at all, suggesting a possible mitigation strategy employed by the interpreter(s). Likewise, the rest of *bat gwo* were rendered into other markers including *nevertheless*, *nonetheless*, *while*, *yet*, *having said that / that said*, all of which indicate concession and topic change, yet with an even higher degree of subtlety as compared to *however*.

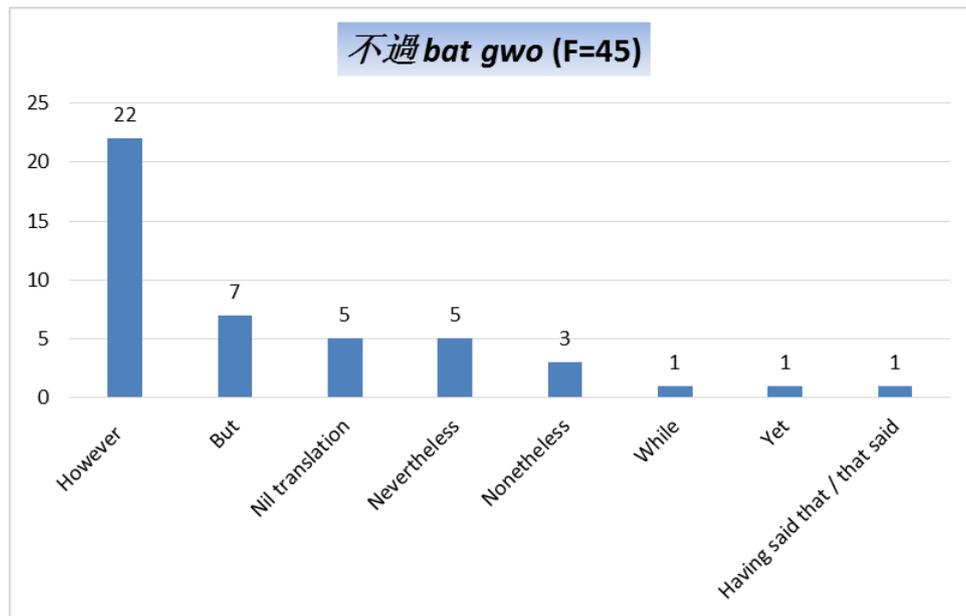


Figure 1. The frequency of different renditions of “bat gwo” in the English sub-corpus

The renditions of *daan (hai)* show a similar pattern (Figure 2). Among the 287 cases of *daan (hai)*, the majority were interpreted into *but* (frequency=74) — its closest equivalence in English that signals denial and contrast. A total of 60 *daan (hai)* received no interpretation at all, which, again, indicates a possible mitigation strategy employed by the interpreter(s). Similar to the case of *bat gwo*, *however* — the other frequently used English marker apart from *but* — was the next most often employed rendition of *daan (hai)* in the English sub-corpus. In addition to *but* and *however*, *daan (hai)* was also rendered into the following English markers: *although*, *despite*, *while*, *nevertheless*, *yet*, *having said that / that said*, *nonetheless*, *though*, *notwithstanding*, *on the other*, *regardless*, *even so*, and *after*, most of which signal concession and topic change, with an even higher degree of subtlety as compared to *however*.

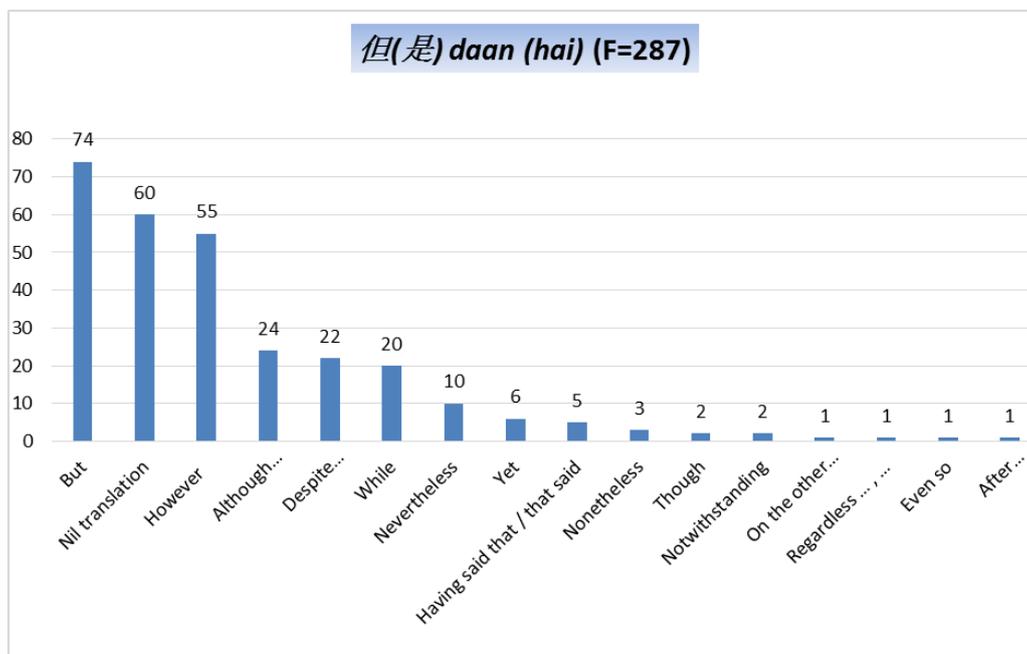


Figure 2. The frequency of different renditions of “daan (hai)” in the English sub-corpus

### 4.3. The use of “however” and “but”

The use of *however* and *but* in the English sub-corpus was also investigated through looking into their source segments in the Cantonese sub-corpus. Figure 3 lists all the Cantonese source segments of *however* and their frequency. The figures show that the use of *however* mostly resulted from the employment of *daan (hai)* in the Cantonese source text (frequency=52). There were 24 instances of *however* which corresponded to no source segment at all, indicating a possible trend of explicitation — a feature of interpreted or translated language. Only 21 cases of *however* were caused by the use of *bat gwo* — its closest equivalence in Cantonese. The rest 4 cases resulted from the use of the Cantonese marker *ho si (可是)* (frequency=3; close in meaning to both *however* and *but* in English, indicating concession or contrast) and *zeon gun (儘管)* (frequency=1; close in meaning to *although* or *in spite of* in English, indicating concession or topic change).

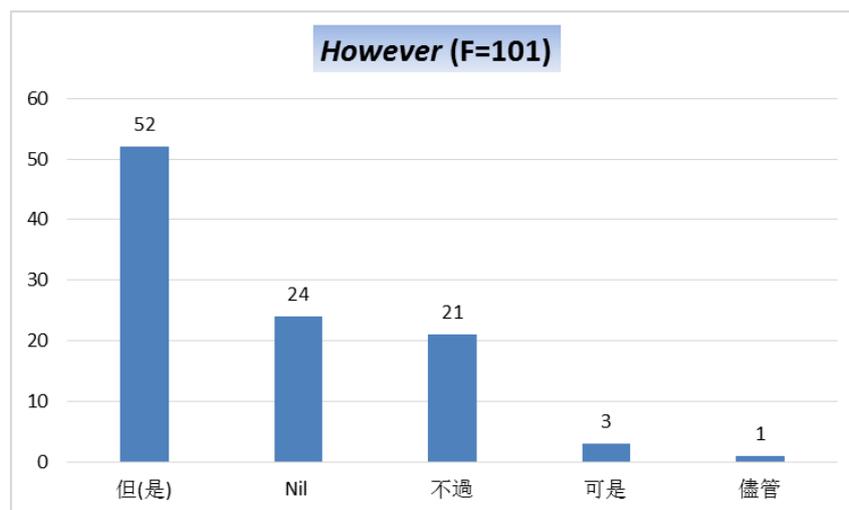


Figure 3. The frequency of different source segments of “however” in the Cantonese sub-corpus

Likewise, Figure 4 lists all the Cantonese source segments of *but* and their frequency. The results show that most of the use of *but* was contributed by *daan (hai)* — its closest equivalence in Cantonese. However, there were 52 instances of the use of *but* which corresponded to no source segments at all in the Cantonese sub-corpus, indicating, again, the possibility of explicitation. The rest of the 16 instances of *but* were attributed by the use of the Cantonese markers *ji (而)* (frequency=8; close in meaning to *however* and *but*, indicating concession or contrast), followed by *bat gwo* (frequency=7) and *koek (畀)* (frequency=1; close in meaning to *however* and *but*, indicating concession or contrast).

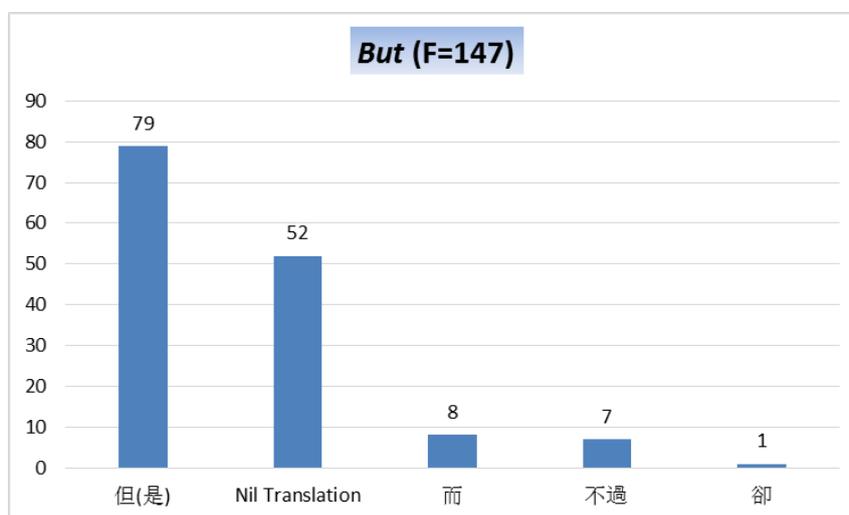


Figure 4. The frequency of different source segments of “but” in the Cantonese sub-corpus

## 5. Conclusions and implications

This study investigates the use of contrastive markers, in particular the Cantonese markers *bat gwo* and *daan (hai)* and their closest equivalences, *however* and *but*, in English, through a parallel corpus consisting of policy addresses delivered by Chief Executives in Hong Kong and their English interpretations. The findings suggest that interpreters tend to employ mitigation strategies in political interpreting, and confirm a possible trend of explicitation in interpreted language from Cantonese to English (cf. Shlesinger, 2009). These findings shed light on the development of e-learning for pragmatic competence training of interpreters working in political settings, as well as natural language processing applications for handling such a high level linguistic feature.

### 5.1. Pragmatic strategies applied by interpreters

The markers *bat gwo* and *daan (hai)*, and their closest equivalences in English, i.e., *however* and *but*, were found to be the most frequently used contrastive markers in the Cantonese and English political speeches. The markers *daan (hai)* and *but*, both indicating denial of expectation and contrast, were found to be even more frequently used than the subtler markers *bat gwo* and *however*. The use of *daan (hai)* even exceeds that of *but* as indicated by the average frequency per 10K, indicating the possible adoption of mitigation strategies by interpreters working from Cantonese to English in political settings.

This trend of mitigation is also evidenced in the renditions of *bat gwo* and *daan (hai)* in the English sub-corpus by the many instances of nil translation, and the substantial number of *however* attributed by the use of *daan (hai)* in the Cantonese sub-corpus.

To conclude, the findings suggest that interpreters tend to intentionally monitor the use and rendition of contrastive markers when interpreting the speeches delivered by senior government officials in Cantonese by reducing the use of contrastive markers showing denial and contrast, and employing additional contrastive markers that suggest concession and topic change. These findings provide evidence to the delicacies of political interpreting and show the importance of pragmatic awareness enhancement in the training of political interpreters.

The study also shows that both speakers and interpreters employ different varieties of contrastive markers in political speeches, in addition to the most representative ones *bat gwo* and *daan (hai)* in Cantonese, and their closest equivalences, *however* and *but*, in English:

- Contrastive markers in Cantonese (by frequency): *daan (hai)* 但(是); *bat gwo* 不過; *ho si* 可是; *zeon gun* 儘管; *ji 而*; and *koek 卻*.
- Contrastive markers in English (by frequency): *but*; *however*; *although*....,....; *despite*....,....; *while*; *nevertheless*; *yet*; *having said that / that said*; *nonetheless*; *though*; *notwithstanding*; *on the other hand*; *regardless ... , ...*; *even so*; and *after all*.

These findings suggest that political interpreters should be equipped with a large repertoire of pragmatic resources in both languages in order to render appropriately or even shift slightly the pragmatic meaning of source texts in political settings.

### 5.2. E-learning for pragmatic competence training and natural language processing applications

This study has shown the diversity in how contrastive markers are translated/interpreted into another language. The results suggest that real texts are needed to be provided as a reference for learners to explore the context of use for the varieties of contrastive markers. This would require corpus analysis tools, such as *concordancer*, to facilitate learners to search and examine corpus data. The parallel texts and instances of contrastive markers collected in this study thus serve as a foundation. A Cantonese–English parallel corpus with contrastive markers annotated can be built up as a corpus-based e-learning resource that helps student interpreters to master relevant interpreting strategies to handle each of the contrastive markers in different contexts. Its benefits have already been emphasized in literature, that “the advantage of using translation corpora is especially strong

for pragmatic markers, precisely because of their underspecified core meaning and their polysemous nature” (Aijmer & Simon-Vandenberg, 2006, p. 6).

For the development of natural language processing applications, processing of high-level linguistic features at the discourse level has been challenging (Wong & Kit, 2012; Wong & Lee, 2013). It has been suffering from the lack of linguistic rules guiding the analysis and translation of the discourse markers. This study contributes to showing there is no simple one-to-one mapping between English and Cantonese contrastive markers. It has also resulted in a collection of corpus-based instances regarding the use and translation of the markers, which facilitates future work on examining more contextual features to identify the parallel relationships between the markers in the two languages.

What should be noted is that the source texts came from a small number of speakers — only three Chief Executives in the past 20 years, and thus the use of contrastive markers may be influenced by personal spoken styles of the speakers. This may be verified by future work on comparing the use of contrastive markers by the different speakers.

Another possible extension of the study is to examine the use of the same contrastive markers in a comparable corpus consisting of similar speeches delivered by native speakers of English, which will provide further evidence as to the featured use of pragmatic markers in differentiating interpreted speeches and non-interpreted speeches delivered in English. Such findings will provide further reference to setting criteria for strengthening the political pragmatic competence of student interpreters towards a native-like level (cf. Pan & Wong, 2015a, 2015b).

Despite the limitations, the present study offers a close examination of the use of contrastive markers in both Cantonese and English political speeches. Its findings show the delicacies of political discourse and indicate the significance of interpreters (and translators) in the process of co-constructing such a discourse. They also provide important implications for e-learning design for pragmatic competence training and natural language processing applications.

### Acknowledgements

This study is sponsored by HKSAR’s Research Grants Council under its Early Career Scheme (Project Number: 22608716).

### References

- Aijmer, K. & Simon-Vandenberg, A.-M. (2006). Introduction. In Karin Aijmer & Anne-Marie Simon-Vandenberg (Eds.), *Pragmatic markers in contrast* (pp. 1–10). Amsterdam: Elsevier.
- Anthony, L. (2014). *AntPConc* (Version 1.1.0) [Computer Software]. Tokyo, Japan: Waseda University. Retrieved from <http://www.antlab.sci.waseda.ac.jp/>
- Anthony, L. (2015). *SegmentAnt* (Version 1.1.0) [Computer Software]. Tokyo, Japan: Waseda University. Retrieved from <http://www.laurenceanthony.net/>
- Biber, D., Conrad, S., Reppen, R., Byrd, P., & Helt, M. (2002). Speaking and writing in the university: A multidimensional comparison. *TESOL Quarterly*, 36(1), 9–48.
- Blakesmore, D. (1989). Denial and contrast: A relevance theoretic account of *but*. *Linguistics and Philosophy*, 12(1), 15–37.
- Bublitz, W. (1988). *Supportive fellow-speakers and cooperative conversations*. Amsterdam: John Benjamins.
- Buri, M. R. (2015). Interpreting in diplomatic settings. *aiic.net*. Retrieved from <http://aiic.net/p/7349>.
- CUHK. (2014). Conjunctions. In *A comparative database of Modern Chinese and Cantonese*. Retrieved from <http://apps.itsc.cuhk.edu.hk/hanyu/Page/Terms.aspx?target=%E9%80%A3%E8%A9%9E>.
- Fraser, B. (1996). Pragmatic markers. *Pragmatics*, 6, 167–190.
- Gagnon, C. (2010). Political translation. In Y. Gambier & L. van Doorslaer (Eds.), *Handbook of translation studies* (pp. 252–256). Amsterdam: John Benjamins.

- Gu Y. G. (1992). Pragmatic politeness and culture. *Foreign Language Teaching and Research*, 4, 30–32.
- Halliday, M.A.K. & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Pan, J., & Wong, B. T.M. (2015a). Investigating pragmatic markers in interpreted political speeches from Chinese to English: A preliminary study. In “*Found in Translation*” – *Translations are the Children of Their Time. Synchrony and Diachrony in Translation, Interpretation and Terminology* (p. 6), Bucharest, Romania.
- Pan, J., & Wong, B. T.M. (2015b). Pragmatic markers in interpreted political discourse: A corpus-driven study. In *International Conference on Corpus Linguistics and Technology Advancement 2015* (p. 48), Hong Kong, China.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. London: Longman.
- Schäffner, C. & Bassnett, S. (2010). Politics, media and translation: Exploring synergies. In C. Schäffner & S. Bassnett (Eds.), *Political discourse, media and translation* (pp. 1–31). Newcastle: Cambridge Scholars Publishing.
- Shlesinger, M. (2009). Towards a definition of Interpretese: An intermodal, corpus-based study. In G. Hansen, A. Chesterman, & H. Gerzymisch-Arbogast (Eds.), *Efforts and models in interpreting and translation research: A tribute to Daniel Gile* (pp. 237–253). Amsterdam: John Benjamins.
- Wang, X. P. (2008). A textual perspective on the translation of political documents. *Journal of Nanchang University*, 39, 151–155.
- Wong, B. T. M. & Kit, C. (2012). Extending machine translation evaluation metrics with lexical cohesion to document level. In *The Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)* (pp. 1060–1068). Jeju, Korea.
- Wong, B. T. M. & Lee, S. Y. M. (2013). Annotating legitimate disagreement in corpus construction. In *The 11th Workshop on Asian Language Resources* (pp. 51–57). Nagoya, Japan.
- Yang, M. (2012). The principles and tactics on diplomatic translation: A Chinese perspective. *Babel*, 58(1), 1–18.
- Xu, Y. (2000). Features in diplomatic interpretation of translation and diplomatic interpreters’ qualities. *Chinese Translators Journal*, (3). 35–38.



**Jun PAN** is Assistant Professor in the Translation Programme and Director of the M.A. Programme in Translation and Bilingual Communication at Hong Kong Baptist University. She is an experienced conference interpreter and interpreter trainer and has been researching in interpreting and translation for many years. Her research interests include corpus-based interpreting studies, learner factors in interpreter training, and professionalism in interpreting.



**Billy Tak Ming WONG** is Research Fellow of the Open University of Hong Kong. He has been teaching and conducting research on computational linguistics, computer-aided translation and educational technology for years. He has widely published on evaluation of machine translation quality and the impacts of technology on education.