International Journal of

Computational Linguistics & Chinese Language Processing

中文計算語言學期刊

A Publication of the Association for Computational Linguistics and Chinese Language Processing

This journal is included in THCI Core, Linguistics Abstracts, and ACL Anthology.

易繫辭曰上古結繩而治後世聖人易之以書契百官以治萬民以察説文敘曰蓋文字者經藝之本宣教明化之始前人所以垂後後人所以識古故曰本立而道生知天下之至賾而不可亂也教化既萌文心雕龍則謂人之立言因字而生句積句而成章積章而成篇篇之彪炳

Vol.14 No.4 December 2009 ISSN: 1027-376X

International Journal of Computational Linguistics & Chinese Language Processing

International Journal of

Computational Linguistics & Chinese Language Processing

The editing of this journal is subsidized by Center for Humanities Research National Science Council in 2009.

Aims and Scope

International Journal of Computational Linguistics and Chinese Language Processing (IJCLCLP) is an international journal published by the Association for Computational Linguistics and Chinese Language Processing (ACLCLP). This journal was founded in August 1996 and is published four issues per year since 2005. This journal covers all aspects related to computational linguistics and speech/text processing of all natural languages. Possible topics for manuscript submitted to the journal include, but are not limited to:

- Computational Linguistics
- Natural Language Processing
- Machine Translation
- Language Generation
- Language Learning
- Speech Analysis/Synthesis
- Speech Recognition/Understanding
- Spoken Dialog Systems
- Information Retrieval and Extraction
- Web Information Extraction/Mining
- Corpus Linguistics
- Multilingual/Cross-lingual Language Processing

Membership & Subscriptions

If you are interested in joining ACLCLP, please see appendix for further information.

Copyright

Cover

Calligraphy by Professor Ching-Chun Hsieh, founding president of ACLCLP
Text excerpted and compiled from ancient Chinese classics, dating back to 700 B.C.
This calligraphy honors the interaction and influence between text and language

International Journal of
Computational Linguistics and Chinese Language Processing
vol. 14, no. 4, December 2009

Contents

Papers

The Association for Computational Linguistics and Chinese Language Processing

Computational Linguistics and Chinese Language Processing
Vol. 14, No. 4, December 2009, pp. 341-362

Tonal Effects on Voice Onset Time

Jui-Feng Peng[*], Li-mei Chen[*], and Chia-Cheng Lee[*]

Abstract

This study examines the influence of lexical tone on voice onset time (VOT) in Mandarin and Hakka spoken in Taiwan. The examination of VOT values for Mandarin and Hakka word-initial stops /p, t, k, p^h, t^h, k^h/ followed by three vowels /i, u, a/ in different lexical tones revealed that lexical tone has significant influence on the VOT values for stops. The results are important as they suggest that future studies should take the influence of lexical tone into account when studying VOT values and when designing wordlists for stops in tonal languages. In Mandarin, stops' VOT values, from the longest to the shortest, are in MR, FR, HL, and HF tones. This sequence is the same as in Liu, Ng, Wan, Wang, and Zhang (2008). Later, however, it was found that it is very likely that the sequence results from the existence of non-words. In order to produce non-words correctly, participants tended to pronounce them at a slower speed, especially those in MR tone. Therefore, we further examined the data without non-words, in which no clear sequence was found. For Hakka, post-hoc tests (Scheffe) show that aspirated stops in entering tones, which are syllables ending with a stop, have significantly shorter VOT values than they have in other tones. Although the tonal effects on VOT values are not consistently found in different sets of data, probably due to a methodology problem, the possibility of tonal effect on VOT values could not be excluded. Tonal effect, thus, should be taken into consideration in designing word lists for VOT studies. Moreover, further studies should include both real words and non-words in separate sets of word lists to verify the current study results.

Keywords: Voice Onset Time, Hakka Stops, Mandarin Stops, Tonal Effect

[*] Department of Foreign Languages and Literature, National Cheng Kung University, 1 University Rd., Tainan, Taiwan. Telephone: (06)2757575 ext. 52231
E-mail: leemay@mail.ncku.edu.tw
The author for correspondence is Li-mei Chen.

1. Introduction

The aim of this paper is to explore whether lexical tones influence the VOT values for word-initial stops. This issue is important because VOT is considered one of the reliable acoustic features for differentiating consonant stops (Cho & Ladefoged, 1999; Gósy, 2001; Lisker & Abramson, 1964; Riney, Takagi, Ota, & Uchida, 2007; Rochet & Fei, 1991; Zheng & Li, 2005) and it has been applied recently to the study of the language production of patients with language deficits or disorders (Auzou, Ozsancak, Morris, Jan, Eustache, & Hannequin, 2000; Jäncke, 1994). Among the languages being investigated, some are tonal languages, *i.e*. Mandarin, Cantonese, and Taiwanese. In a tonal language, the duration of each lexical tone (which can change the meaning of a word) differs slightly. Consequently, it is possible that lexical tone will affect stops' VOT to some extent; nevertheless, few studies have taken this factor into consideration when studying tonal languages. Therefore, the current study examines two tonal languages, Mandarin and Hakka spoken in Taiwan, to verify the effects of lexical tone. It is hoped that the results of the current study can establish the groundwork for future studies related to VOTs in tonal languages. If lexical tone does influence VOT, it should be considered when creating speech materials in future studies for tonal languages.

2. Literature Review

2.1 Voice Onset Time (VOT)

Lisker and Abramson (1964) defined VOT as the temporal interval from the release of an initial stop to the onset of glottal pulsing for a following vowel. VOT has been considered a reliable phonetic cue for categorizing stop consonants (*i.e*., voiced versus voiceless or unaspirated versus aspirated) in various languages (Cho & Ladefoged, 1999; Gósy, 2001; Keating, Linker, & Huffman, 1983; Lisker & Abramson, 1964; Riney *et al*., 2007; Rochet & Fei, 1991; Zheng & Li, 2005). In addition, by comparing VOT values for stops produced by native and non-native speakers for specific languages, researchers have put forth specific suggestions for language learning and teaching (Liao, 2005; Riney & Takagi, 1999; Zheng & Li, 2005). Moreover, recently, researchers have studied production deficits of aphasia, apraxia, and stuttering patients by observing their VOT values for stops (Jäncke, 1994; Auzou *et al*., 2000; Tsen, 1994).

2.2 Factors Affecting Voice Onset Time

When investigating stops, researchers found that the VOT values for stops varied in relation to the place of articulation. Lisker and Abramson (1964) demonstrated that, for both unaspirated and aspirated stops, velar stops have longer mean VOT values than alveolar and bilabial stops.

In the languages examined, except for Tamil, Cantonese, and Eastern Armenian, alveolar stops tend to have longer mean VOT values than bilabial stops. Cho and Ladefoged's (1999) study further revealed that velar stops have the longest mean VOT values, alveolar stops have intermediate mean VOTs, and bilabial stops have the shortest mean VOTs, with the exception of Navajo and Dahalo. The fact that VOT values get longer when the place of articulation moves from an anterior to a posterior position is confirmed in most languages (Cho & Ladefoged, 1999; Lisker & Abramson, 1964; Rosner, López- Bascuas, García-Albea, & Fahey, 2000; Zheng & Li, 2005); nevertheless, some exceptions exist, including Hungarian, Japanese, and Mandarin.

As for the influence of vowel context, Lisker and Abramson (1967) reported that the vowels following the consonants do not have a significant effect on stops' VOTs. Recently, however, other researchers have made opposing claims. Morris, McCrea, & Herring (2008), who studied English word-initial stops, claimed that stops preceding the high vowels /i/ and /u/ had longer VOTs than stops preceding the low vowel /a/. Similar results were revealed in Rochet and Fei (1991), Chao, Khattab, and Chen (2006), and Chen, Chao, and Peng (2007) studies of Mandarin and Gósy's (2001) study of Hungarian. Furthermore, Gósy (2001) indicated that the higher the tongue position, the longer the VOTs for the preceding voiceless stops. Fant (1973), however, found the opposite to be true in a study of Swedish: the VOTs for aspirated stops preceding /a/ were longer than the VOTs for stops preceding /i/ and /u/. Fant's results are extraordinary, as most studies report that stops preceding high vowels tend to have longer VOTs than stops preceding low vowels.

Moreover, speaking rate might have influences on stops' VOTs. Kessinger and Blumstein (1997), who investigated English, French, and Thai, claimed that the speaking rate affected VOT values for long lag stops in Thai and English and for pre-voiced stops in Thai and French, but did not influence VOTs in the short lag category. Magloire and Green (1999) suggested that the speaking rate affected English monolinguals' VOT production and Spanish monolinguals' production of pre-voicing of the voiced stops. By examining English, Kessinger and Blumstein (1998) also reported that both VOT and vowel duration increased as the speaking rate slowed down. Gósy's (2001) study results further proved this. Gósy found that Hungarian bilabial and velar stops had significantly shorter mean VOTs in natural fluent speech than in carefully produced speech. Therefore, it is reasonable to expect that, in careful speech, the speaking rate will decrease and the accompanying VOT will get longer.

The VOT values for word-initial stops in various languages have been extensively investigated. Although some of the languages studied are tonal languages (*e.g.*, Mandarin, Taiwanese, and Cantonese), few studies have considered the effects of lexical tone when designing speech materials (Chao *et al.*, 2006; Chen *et al.*, 2007; Liao, 2005; Lisker & Abramson, 1964; Rochet & Fei, 1991). Gu (2005) claimed that tone is affected primarily by pitch. Different tones have different pitch levels, which are determined by the vibrating frequency of the vocal cord. When the vocal cord tenses, the frequency of vibration increases, resulting in a higher pitch level. Conversely, the pitch level is low when the vocal cord is loose. Liu, Ng, Wan, Wang, and Zhang (2008) speculated that VOT durations may be affected by tone, as different tones have different fundamental frequencies and pitch levels, which are determined primarily by the tension of the vibrating structure. In order to achieve different levels of tension, different amounts of time might be needed. Consequently, VOT values may vary when they occur in different lexical tones. Gu (2005) further indicated that duration affects lexical tone to some extent; for example, in Hakka, the entering tone is short and rapid, meaning less time is needed to produce it. In a tonal language, the durations for each lexical tone are slightly different; therefore, it is reasonable that lexical tone might have some effects on stops' VOTs. Liu *et al*. (2008), who studied the effect of tonal changes on VOTs between normal laryngeal and superior esophageal speakers of Mandarin Chinese, reported an important finding. Normal laryngeal speakers produce significant differences in VOT values as a result of lexical tones. According to their results (Figure 1), stops in the High-falling tone have significantly shorter mean VOT values than stops in the Mid-rising tone and Falling-rising tone. Nevertheless, it should be noted that, in Liu *et al*.'s study, some of the speech materials were non-words. The researchers did not determine whether participants produced real words and non-words differently; therefore, more studies examining the influences of tone are needed. By carrying out a systematic study with respect to the influence of lexical tone on a stop's VOT using two tonal languages (*i.e*., Mandarin and Hakka spoken in Taiwan), the current study aims to create a foundation for future linguistic studies focused on tonal languages.

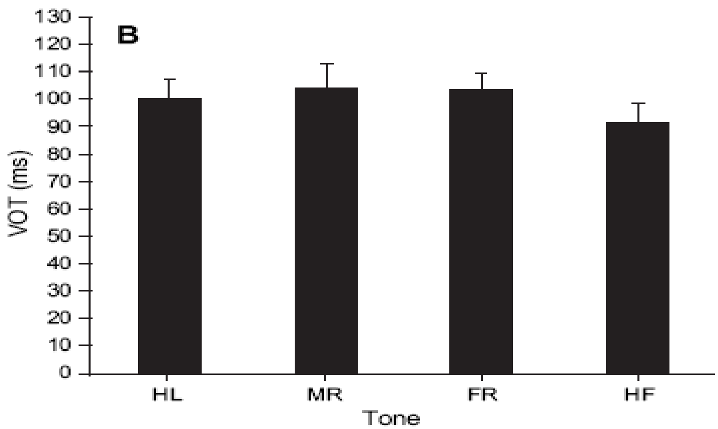


Figure 1. VOTs for Mandarin stops in individual tones produced by normal laryngeal speakers (Taken from Liu et al., 2008).

2.3 Tones in Mandarin and Hakka Spoken in Taiwan

Mandarin and Hakka only have voiceless stops; therefore, the current study investigates the unaspirated stops /p, t, k/ and aspirated stops /p^h, t^h, k^h/. In addition, Mandarin and Hakka are tonal languages, in which a word's meaning can be changed by the tone in which it is pronounced. Chao (1967) suggested a numerical notation for lexical tones, dividing a speaker's pitch range into four equal intervals by five points: 1 (low), 2 (half-low), 3 (middle), 4 (half-high), and 5 (high). The numerical notation indicates how the pitches of a lexical tone change. For example, the numerical notation for a Mid-rising tone in Mandarin is 35, which indicates that the pitch will go from middle to high. Mandarin has four contrasting lexical tones: High-level (HL) (55), Mid-rising (MR) (35), Falling-rising (FR) (214), and High-falling (HF) (51). Sixian Hakka has six contrasting lexical tones: low-rising (LR) (24), mid-falling (MF) (31), high-level (HL) (55), low-entering (LE) (32), low-level (LL) (11), and high-entering (HE) (55). Among them, LE and HE tones are short and rapid, and the words in these two tones end in a stop, like /p/, /t/, /k/.

Mandarin Chinese and Hakka have specific tone sandhi rules. In Mandarin, FR tone, which has the longest duration among the four lexical tones, becomes MR tone when followed by another FR tone (Cheng, 1973). In Sixian Hakka, LR tone becomes LL tone when preceding a LR tone, HL tone, or HE tone. Therefore, tone sandhi rules are taken into consideration when developing speech materials in order to avoid the combinations that might cause tonal change.

Mandarin (Chao, 1967)

FR Tone → MR Tone / _____ {FR Tone}

Sixian Hakka (Chung, 2004)

LR Tone → LL Tone / _____ {LR Tone, HL Tone, HE Tone}

3. Methodology

This study examined word-initial unaspirated stops /p, t, k/, and aspirated stops /p^h, t^h, k^h/, in combination with three corner vowels /i, u, a/ in Mandarin and Hakka spoken in Taiwan. Except for participants and speech materials, the methodology employed for both languages was the same.

3.1 Participants

In this study, the Mandarin and Hakka participants were different. The Mandarin participants included 15 male and 15 female Mandarin speakers from Tainan City with an age range from 23 to 33 years (mean = 27.2 years). All participants had grown up in Taiwan and had no hearing or speech defects. For Hakka, Sixian Hakka was chosen because it is the most extensively used Hakka dialect in Taiwan. The average age of the 21 participants - 11 men and 10 women - was 51 years, with the oldest being 80 and the youngest being 36. All of the participants for Hakka were also fluent Mandarin speakers as Mandarin is the official language in Taiwan. In the current study, the age range of Mandarin participants is controlled to within 10 years to avoid the effect of age difference. As for Hakka participants, the age-range was quite wide because it is not easy to find fluent Hakka speakers.

3.2 Data Collection

The speech materials in both languages were combinations of six stops /p, t, k, $p^{\mathbf{h}}$, $t^{\mathbf{h}}$, $k^{\mathbf{h}}$/ and three vowels /i, u, a/, resulting in 18 combinations. Mandarin's 4 contrasting lexical tones meant that a total of 72 monosyllabic words were created; among them, 18 combinations do not have corresponding Chinese characters in Mandarin. The 6 contrasting lexical tones in Sixian Hakka resulted in 108 monosyllabic words, 12 of which do not have corresponding Chinese characters. Chen *et al*. (2007) claimed that disyllabic words can create a more natural-like context for participants. Therefore, in order to make speakers produce the words more naturally, all of the words were followed by another word in order to create meaningful disyllables, including non-words. For example, the Mandarin word /pi/ was followed by another word /p^huo/ to become the existing disyllable /pi p^huo/ "force." Even non-words were arranged in disyllabic forms to give them a more natural-like quality. Since the neutral tone in Mandarin never occurs in phrase-initial position, it was not evaluated in this study. The structure of non-words was the same as real words, which is a CV syllable with one consonant (stop) and one vowel (corner vowel). For example, there is no /k^ha/ in MR tone in Mandarin or /pu/ in MF tone in Hakka, so non-words were created for these combinations. The way we measure VOTs of non-words is the same as with the real words. For example, /k^ha/ is measured from target consonant /k^h/ to /a/.

The corpus was arranged randomly. Participants were asked to read the words out loud in a normal voice and at a comfortable rate. After finishing, the participants were asked to read the words a second time. Therefore, two groups of data were gathered for each participant. All speech was recorded using a 24 bit WAVE/MPS recorder, connected to AKG C520 Head-Worn Condenser Microphone positioned approximately 10 to 15 centimeters from the participant's mouth in a quiet room.

3.3 Data Measurement and Analysis

After recording, data were edited into individual files and analyzed using the Praat software. VOT, measured in milliseconds (ms), was obtained by measuring the temporal interval between the beginning of the release burst and the onset of the following vowel, as shown in Figure 2. The values of both the waveform and spectrogram were recorded, but the VOTs were determined primarily through waveform analysis, with the values in the spectrogram being provided as references. If the values in waveform differed from the values in the spectrogram by more than five milliseconds, the data were re-measured to verify accuracy.

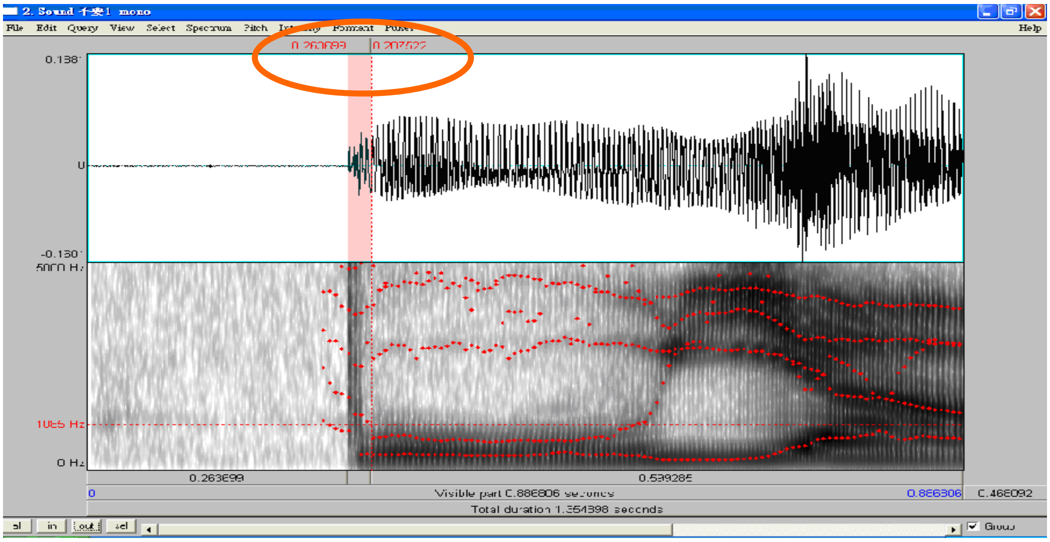


Figure 2. Spectrogram and waveform for Mandarin word /pu iau/ 'don't want'. The values in the circle are the starting and endpoints of VOT.

The VOT values were measured by one investigator. Furthermore, 10% of each recording (selected randomly) was re-measured by another investigator to verify the reliability of the results. Ultimately, 7 Mandarin words and 11 Hakka words for each recording were re-measured. Pearson's product-moment correlations (Gravetter & Wallnau, 2008) indicated high inter-rater agreement for both Mandarin and Hakka data (Mandarin: r = .995, p<.001; Hakka: r = .978, p<.001).

When analyzing the data, VOT values for mispronounced words were omitted. Moreover, data for Hakka /pi/ in HE Tone were not analyzed due to incorrect word choices. A four-way mixed factorial ANOVA (Montgomery, 2009; the same test was used in Francis, Ciocca, & Yu, 2003) (place of articulation by vowel context by lexical tone by gender) was used to examine whether the variables significantly influenced each stop's VOT. In addition, differences between the examined targets were analyzed using T-test or post-hoc tests

(Scheffe) (Gravetter & Wallnau, 2008); results were considered significant when the *p* value was less than 0.05.

Four-way mixed factorial ANOVA can be illustrated in the following formula (Montgomery, 2009).

$$y_{ijklm} = \mu+\tau_i+\beta_j+\gamma_k+\sigma_l+(\tau\beta)_{ij}+(\tau\gamma)_{ik}+(\tau\sigma)_{il}+(\beta\gamma)_{jk}+(\beta\sigma)_{il}+(\gamma\sigma)_{kl}+(\tau\beta\gamma)_{ijk}+(\tau\beta\sigma)_{ijl}+(\beta\gamma\sigma)_{jkl}+(\tau\gamma\sigma)_{ikl}+(\tau\beta\gamma\sigma)_{ijkl}+\varepsilon_{ijklm}$$

i= 1, 2, 3 place
j= 1, 2, 3 vowel
k= 1, 2, 3, 4 tone
l= 1, 2 gender
m= 1, 2, …30 subjects

4. Results and Discussion

When examining the VOT values for Mandarin stops, it became apparent that they tend to be longer than the mean VOT values reported in the studies by Liao (2005), Chao *et al*. (2006), and Chen *et al*. (2007) and shorter than those reported by Rochet and Fei (1991). Examining the methodologies in these previous studies indicated that the speech materials in Rochet and Fei's study were monosyllabic, but disyllabic in the remaining four studies. Gósy (2001) claimed that speakers speak in a careful and disciplined way while uttering syllables and words in isolation. Therefore, participants are expected to produce monosyllables in a more careful manner as their speech tempo decreases. According to Kessinger and Blumstein (1998), VOTs get longer while the speaking rate slows down. This may explain why the mean VOTs for Mandarin stops in Rochet and Fei's study were longer than in other reports. Another possible explanation might be regional differences in the target language. Although the target language is the same, participants in Rochet and Fei's study grew up in Mainland China, while participants in the other studies were raised in Taiwan. Consequently, regional differences might be another origin of the variations. As a result, comparing the results of Liao (2005), Chao *et al*. (2006), and Chen *et al*. (2007) to those of the present study indicates that the mean VOT values for the stops in the present study tend to be longer than their counterparts in the other studies. Such differences stem primarily from the existence of non-words in the current study. During the recording process, speakers tended to produce non-words more carefully and at a lower speed because they were not familiar with the non-words. An examination of the data both with and without non-words separately demonstrated that the existence of non-words resulted in stops having longer mean VOTs.

4.1 Mandarin Chinese

The statistical analyses were conducted using a four-way mixed factorial ANOVA. For Mandarin unaspirated stops, the results showed a primary effect of place of articulation, $F(2, 972) = 522.9680$; vowel context, $F(2, 972) = 117.3569$; lexical tone, $F(3, 972) = 6.5506$; and gender $F(1, 972) = 56.9180$ (all $p < .001$). These results indicate that the stop place of articulation (bilabial, alveolar, and velar), vowel context (/i/, /u/, /a/), lexical tone (HL, MR, FR, HF), and gender do create significant differences in VOT values of word-initial unaspirated stops. Furthermore, significant two-way interactions were also observed (Figures 3-4) between place of articulation and vowel context and between vowel context and gender. A complete ANOVA table showing interaction between variables is listed in Appendix 1.

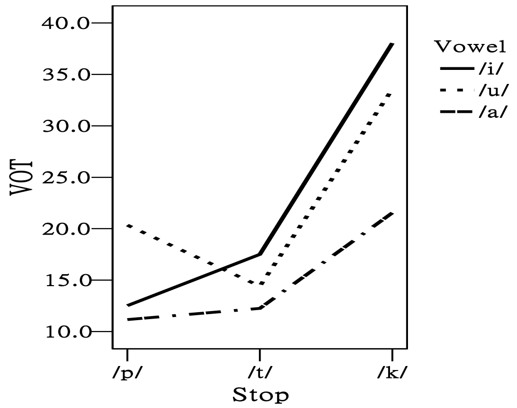


Figure 3. Significant interaction between place of articulation and vowel context in unaspirated stops in Mandarin

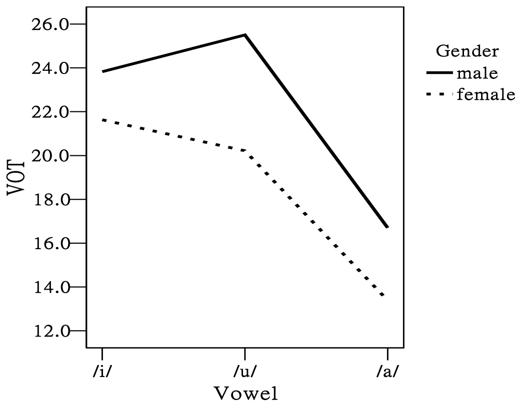


Figure 4. Significant interaction between vowel context and gender in unaspirated stops in Mandarin

As for aspirated stops, the results demonstrated a main effect of place of articulation, $F(2, 972) = 95.2742$; vowel context, $F(2, 972) = 43.5079$; lexical tone, $F(3, 972) = 12.0121$; and gender $F(1, 972) = 20.3186$, thereby indicating that VOT values of word-initial aspirated stops would vary in accordance with place of articulation, vowel context, lexical tone and gender (all $p < .001$). Significant two-way interactions occurred between place of articulation and vowel context, between place of articulation and lexical tone, and between vowel context and gender (Figures 5-7). For both unaspirated and aspirated stops in Mandarin, no significant three-way and four-way interactions were evident between variables. A complete ANOVA table showing interaction between variables is listed in Appendix 2.

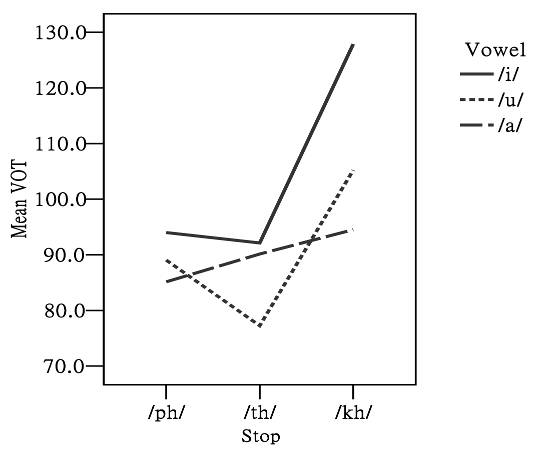


Figure 5. Significant interaction between place of articulation and vowel context in aspirated stops in Mandarin

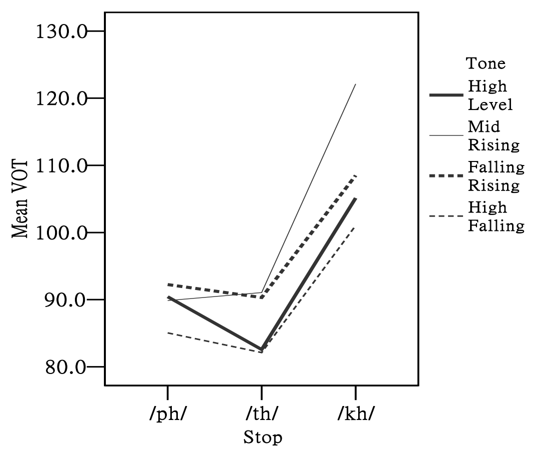


Figure 6. Significant interaction between place of articulation and lexical tones in aspirated stops in Mandarin

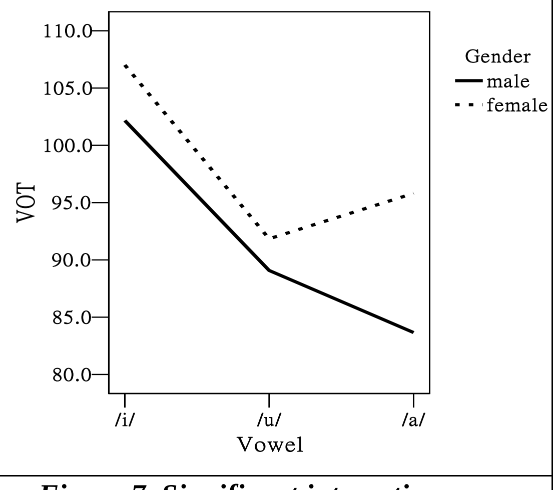


Figure 7. Significant interaction between vowel context and gender in aspirated stops in Mandarin

Table 1 lists mean VOT values and standard deviation of Mandarin stops in each lexical tone. ANOVA tests revealed that the lexical tone significantly influences VOTs of stops [unaspirated stops, $F(3, 972) = 6.5506$, $p < .001$; aspirated stops, $F(3, 972) = 12.0121$, $p < .001$]. A post-hoc test reveals that aspirated stops in HF have significantly shorter mean

VOTs than stops in MR and FR (all $p < .05$). In addition, for both unaspirated and aspirated stops, stops in MR have the longest mean VOTs while stops in HF have the shortest mean VOTs. VOTs, from longest to shortest, occurred in MR, FR, HL, and HF - the same sequence as in Pearce (2009) and Liu *et al*. (2008). Yet, it is worth noting that, in both studies, some of the speech materials were non-words. It was subsequently determined that the sequence results from the existence of non-words because, in order to produce non-words correctly, participants tended to pronounce them at a slower speed, making the VOTs longer.

Therefore, the current study further examined the data without non-words, in which the main variation occurred in participants' productions of stops in MR tone. When analyzing the data with non-words, Mandarin unaspirated and aspirated stops in MR tone had the longest mean VOTs. Nevertheless, when excluding non-words, the results revealed that stops in MR tone did not have the longest mean VOTs, and unaspirated stops in MR tone even had significantly shorter mean VOTs than in HL and FR tones. The divergence revealed that participants' productions of real words or non-words in MR tone were quite different. In addition, ANOVA tests revealed that lexical tone does not significantly influence stops' VOTs by analyzing the data without non-words. Further studies are needed to have separate sets of wordlists for real words and non-words to verify the current findings.

Table 1. Mean VOT values of Mandarin stops with different lexical tones. All measurements are in milliseconds (ms).

	With non-words				Without non-words			
	unaspirated stops		aspirated stops		unaspirated stops		aspirated stops	
	Mean	*SD*	mean	*SD*	mean	*SD*	mean	*SD*
HL	20.20	(11.90)	92.72	(25.53)	17.71	(9.95)	88.69	(20.4)
MR	21.10	(12.68)	101.02	(30.21)	13.99	(6.03)	89.47	(23.31)
FR	20.89	(13.35)	97.03	(27.75)	17.00	(10.98)	92.30	(23.49)
HF	18.42	(9.94)	89.4	(25.72)	16.32	(9.07)	85.62	(24.18)

Table 2. Mean VOT values of six Mandarin stops with different lexical tones. All measurements are in milliseconds (ms).

	With non-words (mean)						Without non-words (mean)					
	p	t	k	p^h	t^h	k^h	p	t	k	p^h	t^h	k^h
HL	14.4	14.7	31.5	90.4	82.6	105.1	12.0	14.7	27.9	90.4	82.6	95.2
MR	16.0	15.1	32.2	89.9	91.1	122.1	12.3	15.1	*	89.9	88.9	97.4
FR	15.0	14.9	32.7	92.3	90.3	108.5	11.3	14.9	34.7	90.2	90.3	97.8
HF	13.3	14.2	27.7	85.0	82.1	101.0	11.8	14.2	31.8	85.0	82.1	*

*No real word data

In the data with non-words in Table 2, individual stops mostly follow the pattern of MR with the longest VOT and HF with the shortest VOT, especially for /p, t, t^h, k^h/. As for the data without non-words, the general pattern is also found in /p, t/.

4.2 Lexical Tone and VOT in Hakka

For Hakka unaspirated stops, the results showed a main effect of place of articulation, $F(2, 843) = 404.3395$; vowel context, $F(2, 843) = 69.8958$; lexical tone, $F(5, 843) = 6.3054$; and gender $F(1, 843) = 34.2724$ (all $p < .001$). Similar to the findings in Mandarin stops, these results indicate that stop place of articulation (bilabial, alveolar, and velar), vowel context (/i/, /u/, /a/), lexical tone (HL, MR, FR, HF), and gender do make significant differences in VOT values of word-initial unaspirated stops. Significant two-way interactions occurred between the place of articulation and lexical tone (Figure 8), and significant three-way interactions occurred among place of articulation, vowel context, and lexical tone. A complete ANOVA table showing interaction between variables is listed in Appendix 3.

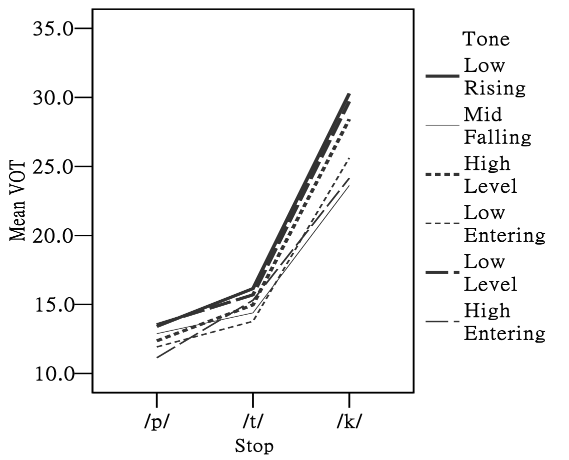


Figure 8. Significant interaction between place of articulation and lexical tone in unaspirated stops in Hakka

As for aspirated stops, the results also demonstrated a main effect of place of articulation, $F(2, 798) = 55.6543$; vowel context, $F(2, 798) = 44.7708$; lexical tone, $F(5, 798) = 46.4587$; and gender $F(1, 798) = 42.0266$, thereby indicating that VOT values of word-initial aspirated stops would vary in accordance with place of articulation, vowel context, lexical tone, and gender (all $p < .001$). Significant two-way interactions existed between place of articulation and vowel context, between place of articulation and lexical tone, and between vowel context and lexical tone (Figures 9-11). In addition, significant three-way interactions occurred among place of articulation, vowel context, and lexical tone. A complete ANOVA table showing interaction between variables is listed in Appendix 4.

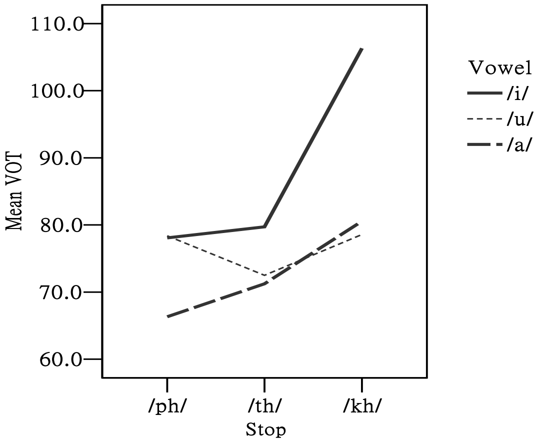


Figure 9. Significant interaction between place of articulation and vowel context in aspirated stops in Hakka

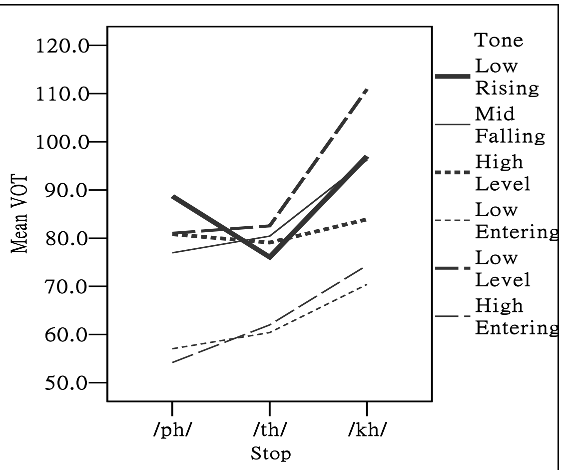


Figure 10. Significant interaction between place of articulation and lexical tone in aspirated stops in Hakka.

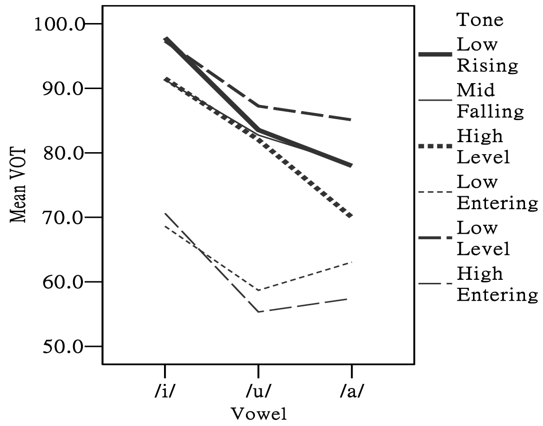


Figure 11. Significant interaction between vowel context and lexical tones in aspirated stops in Hakka.

The mean VOT values and standard deviations for Hakka stops in each lexical tone are shown in Table 3. ANOVA tests indicated that the lexical tone significantly influences stops'

VOTs [unaspirated stops, $F(5, 843) = 6.3054$, $p < .001$; aspirated stops, $F(5, 798) = 46.4587$, $p < .001$]. Unaspirated and aspirated stops in LR and LL have longer mean VOTs than stops in other tones, whereas the shortest mean VOTs for both unaspirated and aspirated stops are in HE. The post-hoc test revealed that aspirated stops in HE and LE have significantly shorter mean VOTs than those in other tones (all $p < .001$). In Hakka, HE and LE tones are entering tones, which are short, rapid, and end in a stop like /p, t, k/. Gu (2005) further claimed that the durations for entering tones are shorter than the durations for other tones. Therefore, the VOTs for stops in the entering tones are shorter than stops in other tones.

Table 3. Mean VOT values of Hakka stops with different lexical tones. All measurements are in milliseconds (ms)

	Unaspirated stops		Aspirated stops	
	mean	(*SD*)	mean	(*SD*)
LR	20	(11.56)	86.83	(25.8)
MF	16.94	(8)	84.67	(26.56)
HL	18.88	(11.02)	81.32	(23.73)
LE	17.19	(9.44)	62.93	(18.36)
LL	19.4	(11.43)	90.08	(27.08)
HE	16.11	(7.98)	61.53	(20.36)

Table 4. Mean VOT values of six Hakka stops with different lexical tones. All measurements are in milliseconds (ms)

	Without non-words (mean)					
	p	t	k	p^h	t^h	k^h
LR	13.4	16.2	30.3	88.7	76.1	97.0
MF	12.9	14.4	23.6	77.0	80.4	96.2
HL	12.4	15.0	28.4	80.8	79.1	83.9
LE	11.9	13.8	25.6	57.0	60.4	70.4
LL	13.5	15.7	29.7	81.0	82.5	110.9
HE	11.1	15.3	24.2	54.2	62.0	74.4

Table 4 shows that almost all of the stops follow the general pattern where LE and HE are among the shortest.

5. Conclusion

The study results revealed that lexical tones significantly influence VOT values for stops in Mandarin (with both real words and non-words), and significant tonal effect was also found in

Hakka data of real words. Nevertheless, there is no significant tonal effect on VOT in Mandarin data with only real words.

The study results are important as they suggest that future studies should take the influence of lexical tones into account when studying VOT values and when designing wordlists for stops in tonal languages. Although the tonal effects on VOT values are not consistently found in different sets of data, probably due to a methodology problem, the possibility of tonal effect on VOT values could not be excluded. Several factors might contribute to this inconsistency. First, we used different methods in eliciting non-words production in these two languages. In Mandarin, we used Zhuyin Fuhao to guide non-words productions, which might force participants to take a few seconds to figure out the new combinations of Chinese phonetic symbols. In contrast, we asked participants to read a real word first as a clue in producing a target non-word in Hakka. Second, many of the participants in Hakka were at the age range of 50-80, and we found the participants were not flexible enough to comprehend our instructions in producing non-words. Therefore, we decided not to take the data of non-words (mostly by guessing with uncertainly) into analysis in order to have a reliable result. (The discarded cells (non-words) are less than 25% of the total, which fulfills the requirements of ANOVA test for reliable test results (Gravetter & Wallnau, 2008). Future studies should keep the method of elicitation consistent across languages, recruit participants of similar age range, and include both real words and non-words in separate sets of word lists in order to verify the current study findings. Moreover, although the results of this study indicate that lexical tones influence VOT of stops to some extent, we cannot exclude the possibility of correlation between VOT and the duration of lexical tones. Further study can explore this possibility by adopting mean durations of each tone in Mandarin as the normalized parameters upon calculating mean VOTs of stops.

References

Auzou, P., Ozsancak, C., Morris, R. J., Jan, M., Eustache, F., & Hannequin, D. (2000). Voice onset time in aphasia, apraxia of speech and dysarthria: a review. *Clinical Linguistics & Phonetics*, 14(2), 131-150.

Chao, K.-Y., Khattab, G., & Chen, L.- M. (2006). Comparison of VOT patterns in Mandarin Chinese and in English. *Proceedings of the 4th Annual Hawaii International Conference on Arts and Humanities*, 840-859.

Chao, Y. R. (1967). *Mandarin Primer*. Cambridge: Harvard University Press.

Chen, L.- M., Chao, K.- Y., & Peng, J.- F. (2007). VOT productions of word-initial stops in Mandarin and English: a cross-language study. *Proceedings of the 19th Conference on Computational Linguistics and Speech Processing*, 303-317.

Cheng, C.- C. (1973). *A Synchronic Phonology of Mandarin Chinese*. The Hague: Mouton.

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics*, 27, 207-229.

Chung, R.- F. (2004). *An introduction to Taiwan Hakka phonology*. Taipei: Wu-Nan Book.

Fant, G. (1973). *Speech sounds and features*. Cambridge: MIT Press.

Francis, A. L., Ciocca, V., & Yu, J. M. C. (2003). Accuracy and variability of acoustic measures of voicing onset. *Journal of the Acoustical Society of America*, 113(2), 1025-1032.

Gósy, M. (2001). The VOT of the Hungarian voiceless plosives in words and in spontaneous Speech. *International Journal of Speech Technology*, 4, 75-85.

Gravetter, F. J., & Wallnau, L. B. (2008). *Statistics for the behavioral sciences* (8^{th} ed.). Belmont, California: Wadsworth.

Gu, G.- S. (2005). Phonology of Taiwan Hakka. In Gu, G.-S. (Ed.), *Introduction to Taiwan Hakka*, 117-161. Taipei: Wu-Nan.

Jäncke, L. (1994). Variability and duration of voice onset time and phonation in stuttering and nonstuttering adults. *Journal of Fluency Disorders*, 19(1), 21-37.

Keating, P. A., Linker, W., & Huffman, M. (1983). Patterns in allophone distribution for voiced and voiceless stops. *Journal of Phonetics*, 11, 277-290.

Kessinger, R. H., & Blumstein, S. E. (1997). Effects of speaking rate on voice-onset time in Thai, French, and English. *Journal of Phonetics*, 25, 143-168.

Kessinger, R. H., & Blumstein, S. E. (1998). Effects of speaking rate on voice-onset time and vowel production: some implications for perception studies. *Journal of Phonetics*, 26, 117-128.

Liao, S. J. (2005). *Interlanguage production of English stop consonants: a VOT analysis.* Unpublished master's thesis, National Kaohsiung Normal University, Kaohsiung, Taiwan.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: acoustical measurements. *Word*, 20, 384-422.

Lisker, L., & Abramson, A. S. (1967). Some effects of context on voice onset time in English stops. *Language Speech*, 10, 1-28.

Liu, H., Ng, M. L., Wan, M., Wang, S., & Zhang,Y. (2008). The effect of tonal changes on voice onset time in Mandarin esophageal speech. *Journal of Voice*, 22(2), 210-218.

Magloire, J., & Green, K. P. (1999). A cross-language comparison of speaking rate effects on the production of voice onset time in English and Spanish. *Phonetica*, 56, 158-185.

Montgomery, D. C. (2009). *Design and analysis of experiments, student solutions manual* (7^{th} ed.). Hoboken, N.J. :Wiley.

Morris, R. J., McCrea, C. R., & Herring, K. D. (2008). Voice onset time differences between adult males and females: isolated syllables. *Journal of Phonetics*, 36, 308-317.

Pearce, M. (2009). Kera tone and voicing interaction. *Lingua,* 119(6), 846-864.

Riney, T. J., & Takagi, N. (1999). Global foreign accent and voice onset time among Japanese EFL speakers. *Language Learning*, 49(2), 275-302.

Riney, T. J., Takagi, N., Ota, K., & Uchida, Y. (2007). The intermediate degree of VOT in Japanese initial voiceless stops. *Journal of Phonetics*, 35, 439-443.

Rochet, B. L., & Fei, Y. (1991). Effect of consonant and vowel context on Mandarin Chinese VOT: production and perception. *Canadian Acoustics*, 19(4), 105-106.

Rosner, B. S., López- Bascuas, L. E., García-Albea, J. E., & Fahey, R. P. (2000). Voice-onset times for Castilian Spanish initial stops. *Journal of Phonetics*, 28, 217-224.

Tsen, C.Y. (1994). An investigation on voice onset time and tone production of Chinese aphasia. *Bulletin of the Institute of History and Philology Academia Sinica*, 65(1), 37-79.

Zheng, X.- R., & Li, Y.- H. (2005). A contrastive study of VOT of English and Korean stops. *Journal of Yanbian University*, 38(4), 99-102.

Appendix 1.

Four-way factorial ANOVA for Mandarin unaspirated stops

Source of Variation	SS	df	MS	F
Place of articulation	62001	2	31000	522.9680
Vowel	13913	2	6957	117.3569
Gender	3374	1	3374	56.9180
Tone	1165	3	388	6.5506
Vowel×Place	10289	4	2572	43.3945
Gender×Place	242	2	121	2.0372
Gender×Vowel	427	2	214	3.6038
Tone×Place	544	6	91	1.5286
Tone×Vowel	252	6	42	0.7081
Tone×Gender	130	3	43	0.7283
Gender×Vowel×Place	349	4	87	1.4725
Vowel×Place×Tone	1016	12	85	1.4283
Place×Tone×Gender	148	6	25	0.4153
Tone×Gender×Vowel	104	6	17	0.2910
Tone×Gender×Vowel×Place	216	12	18	0.3035
Residual	57618	972	59	

Appendix 2

Four-way factorial ANOVA for Mandarin aspirated stops

Source of Variation	SS	df	MS	F
Place of articulation	106145	2	53073	95.2742
Vowel	48472	21	24236	43.5079
Gender	11318	1	11318	20.3186
Tone	20074	3	6691	12.0121
Vowel×Place	38558	4	9640	17.3046
Gender×Place	1273	2	636	1.1422
Gender×Vowel	4216	2	2108	3.7840
Tone×Place	10302	6	1717	3.0824
Tone×Vowel	2378	6	396	0.7115
Tone×Gender	419	3	140	0.2508
Gender×Vowel×Place	3253	4	813	1.4598
Vowel×Place×Tone	5303	12	442	0.7933
Place×Tone×Gender	1880	6	313	0.5626
Tone×Gender×Vowel	552	6	92	0.1650
Tone×Gender×Vowel×Place	3451	12	288	0.5162
Residual	541453	972	557	

Appendix 3.

Four-way factorial ANOVA for Hakka unaspirated stops

Source of Variation	SS	df	MS	F
Place of articulation	38120	2	19060	404.3395
Vowel	6590	2	3295	69.8958
Gender	1616	1	1616	34.2724
Tone	1486	5	297	6.3054
Vowel×Place	4613	4	1153	24.4639
Gender×Place	61	2	31	0.6502
Gender×Vowel	50	2	25	0.5267
Tone×Place	1104	10	110	2.3427
Tone×Vowel	837	10	84	1.7750
Tone×Gender	436	5	87	1.8519
Gender×Vowel×Place	293	4	73	1.5519
Vowel×Place×Tone	2132	19	112	2.3799
Place×Tone×Gender	350	10	35	0.7415
Tone×Gender×Vowel	235	10	23	0.4982
Tone×Gender×Vowel×Place	475	19	25	0.5302
Residual	39738	843	47	

Appendix 4.

Four-way factorial ANOVA for Hakka aspirated stops

Source of Variation	SS	df	MS	F
Place of articulation	46848	2	23424	55.6543
Vowel	37687	2	18843	44.7708
Gender	17688	1	17688	42.0266
Tone	97769	5	19554	46.4587
Vowel×Place	22984	4	5746	13.6520
Gender×Place	123	2	62	0.1465
Gender×Vowel	451	2	225	0.5357
Tone×Place	17302	10	1730	4.1108
Tone×Vowel	8157	10	816	1.9381
Tone×Gender	2030	5	406	0.9649
Gender×Vowel×Place	1436	4	359	0.8532
Vowel×Place×Tone	35729	20	1786	4.2445
Place×Tone×Gender	2244	10	224	0.5331
Tone×Gender×Vowel	4006	10	401	0.9517
Tone×Gender×Vowel×Place	4705	20	235	0.5590
Residual	335866	798	421	

Computational Linguistics and Chinese Language Processing
Vol. 14, No. 4, December 2009, pp. 363-382

A Discrete-cepstrum Based Spectrum-envelope Estimation Scheme and Its Example Application of Voice Transformation

Hung-Yan Gu*and Sung-Feng Tsai*

Abstract

Approximating a spectral envelope via regularized discrete cepstrum coefficients has been proposed by previous researchers. In this paper, we study two problems encountered in practice when adopting this approach to estimate the spectral envelope. The first is which spectral peaks should be selected, and the second is which frequency axis scaling function should be adopted. After some efforts of trying and experiments, we propose two feasible solution methods for these two problems. Then, we combine these solution methods with the methods for regularizing and computing discrete cepstrum coefficients to form a spectral-envelope estimation scheme. This scheme has been verified, by measuring spectral-envelope approximation error, as being much better than the original scheme. Furthermore, we have applied this scheme to building a system for voice timbre transformation. The performance of this system demonstrates the effectiveness of the proposed spectral-envelope estimation scheme.

Keywords: Spectral Envelope, Discrete Cepstrum, Harmonic-plus-noise Model, Voice Timbre Transformation.

1. Introduction

Here, a spectral envelope means a magnitude-spectrum envelope. Various methods have been proposed to estimate the spectral envelope of a speech frame. For example, in LPC (linear prediction coding) based methods (O'Shaughnessy, 2000; Schwarz & Rodet, 1999), the frequency response of an all-pole model is used to approximate the spectral envelope of a speech frame. Nevertheless, the frequency response curve of an LPC all-pole model will usually go below the true envelope around speech formants, and go above the regions where

* Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, 43 Keelung Rd., Sec. 4, Taipei, Taiwan.
E-mail: {guhy, M9615069}@mail.ntust.edu.tw

spectrum magnitudes fall suddenly. This is illustrated in Figure 1 using a frame sliced from an utterance of /i/. Therefore, the mismatches between the LPC envelope curve and the true curve cannot be ignored.

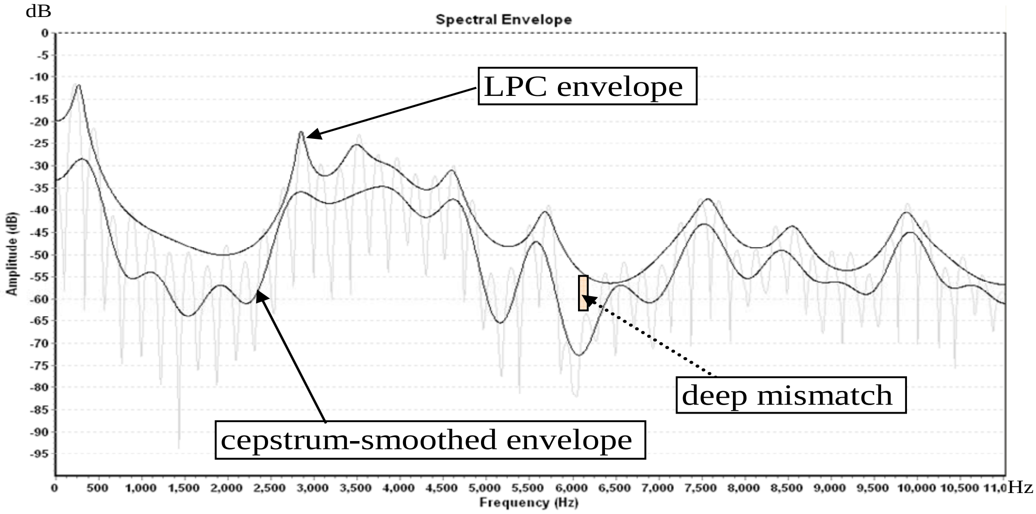


Figure 1. LPC and cepstrum smoothed spectral curves for a frame from /i/.

Besides LPC, several estimation methods are based on cepstrum analysis. The simplest one is to keep some leading cepstrum coefficients but truncate the remanded ones, *i.e.* replace them with zeros. Then, DFT (discrete Fourier transform) is used to transform the cepstrum coefficients back to the spectrum domain to obtain a smoothed spectrum curve. Nevertheless, such a smoothed spectrum curve is not really an envelope curve because it goes between the peaks and valleys of a DFT spectrum. One example is the lower smoothed curve in Figure 1. Therefore, a real cepstrum-based method to estimate a spectral envelope was proposed later by Imai and Abe (Imai & Abe, 1979; Robel & Rodet, 2005). They call this method true envelope estimation. In our opinion, this method is good but lacking in efficiency because a lot of computations are required. Similarly, the method proposed by Kawahara, Masuda-katsuse, and Cheveign (1999), STRAIGHT, is very accurate in its estimated spectral envelope. Nevertheless, it also requires a considerable number of computations and cannot be used to implement real-time systems currently. On the other hand, Galas and Rodet (1990) proposed the concept of discrete cepstrum and designed a feasible estimation method with this concept. Later, Cappé and Moulines (1996) improved this estimation method by adding a regularization technique to prevent unstable vibrating of the envelope curve from occurring. We think that estimating a spectral envelope with discrete cepstrum is a good approach if the feasibility and accuracy issues must be considered simultaneously. Therefore, we began to study the problems that will be encountered in practice.

As an overview, the spectral envelope estimation scheme proposed here is shown in Figure 2. When a speech frame is given, its fundamental frequency is first detected in the first block. If a frame is decided to be voiced, its estimated fundamental frequency will be used later in the block, “spectral peaks selection”. Here, a method combining an autocorrelation function and AMDF (absolute magnitude difference function) is adopted to detect a frame’s fundamental frequency (Kim *et al.*, 1998; Gu, Chang & Wu, 2004). Next, the frame is Hanning windowed and appended with zeros to form a sequence of 1,024 signal samples. This sequence is then transformed to frequency domain with FFT (fast Fourier transform) to obtain its magnitude spectrum. Given the magnitude spectrum, the block “spectral peaks selection” will determine which spectral peaks should be selected according to a method proposed here. After spectral peaks are selected, the frequency value of each selected peak is mapped to its target value with a frequency-axis scaling function proposed here. As the final step, the block “discrete cepstrum computation” will adopt an envelope-approximation criterion (Cappé & Moulines, 1996) to compute discrete cepstrum coefficients according to the selected and mapped spectral peaks.

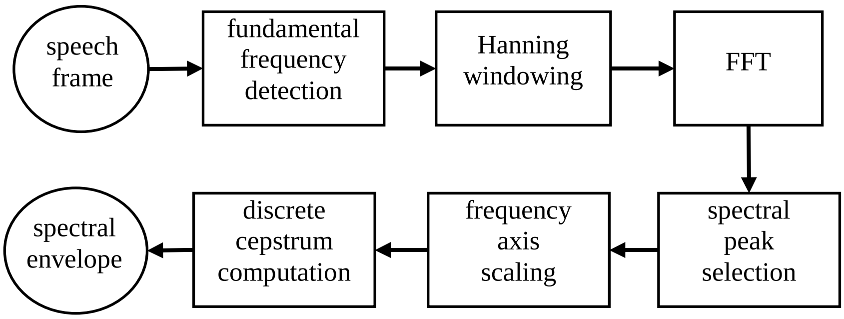


Figure 2. Main flow of the spectral-envelope estimation scheme.

In Figure 2, discrete-cepstrum computation is the main block, which already has been solved by other researchers (Cappé & Moulines, 1996). Nevertheless, the blocks, spectral-peak selection and frequency-axis scaling, still play important roles. When inappropriate peaks are selected or the frequency-axis is not scaled appropriately, the approximated spectral envelope will noticeably deviate from the true envelope. Therefore, we began to study these two blocks’ problems, and the results are presented in Sections 3 and 4, respectively. As to discrete cepstrum, its computation and regularization will be reviewed in Section 2. In Section 5, the proposed scheme is practically evaluated by applying the scheme to build a voice timbre transformation system.

2. Spectral-envelope Estimation with Discrete Cepstrum

2.1 Discrete Cepstrum

The concept of discrete cepstrum was proposed by Galas and Rodet (1990). They adopted the least square criterion to a given set of spectral peaks to derive cepstrum coefficients. Such a derivation method is different from the conventional one. The conventional method transforms the logarithmic magnitude-spectrum with inverse DFT (IDFT) to get its cepstrum coefficients. In the conventional method, let the obtained cepstrum coefficients be c_0, c_1, …, $c_{N\text{-}1}$ where N is the length of the signal sample sequence. According to these cepstrum coefficients, the original logarithmic magnitude-spectrum can be restored with DFT, *i.e.*

$$\log\left|X(k)\right| = \sum_{n=0}^{N-1} c_n e^{-j\frac{2\pi}{N}kn}, \quad 0 \le k \le N-1 \tag{1}$$

where $|X(k)|$, k=0, 1, …, N-1 represent the magnitude spectrum. Since $\log|X(k)|$ is even symmetric, *i.e.* $\log|X(k)| = \log|X(N\text{-}k)|$, the derived cepstrum coefficients are also even symmetric, $c_k = c_{N\text{-}k}$. Therefore, Equation (1) can be rewritten as:

$$\log\left|X(k)\right| = c_0 + 2\sum_{n=1}^{\frac{N}{2}-1} c_n \cos(\frac{2\pi}{N}kn) + c_{N/2}\cos(\pi k), \quad 0 \le k \le N-1\,. \tag{2}$$

If most of the terms on the right side of Equation (2) are cancelled except the leading terms (*e.g.* p+1 terms), the magnitude spectrum computed, $\log S(f)$, would be a smoothed version of the original, $\log|X(f)|$. Here, the index variable, k, in Equations (1) and (2) is replaced with f in order to change the frequency scale from bins to the normalized frequency range from 0 to 1. Accordingly, $\log S(f)$ is computed as:

$$\log S(f) = c_0 + 2\sum_{n=1}^{p} c_n \cdot \cos(2\pi f n)\,, \quad f = \tfrac{0}{N}, \tfrac{1}{N}, \tfrac{2}{N}, ..., \tfrac{N-1}{N}\,. \tag{3}$$

Based on Equation (3), some researchers have proposed to approximate the spectral envelope of $\log|X(f)|$ with $\log S(f)$. Nevertheless, the coefficients, c_n, in Equation (3) cannot be derived directly with IDFT. One derivation method proposed by Galas and Rodet is to define a set of envelope constraints and find the values of the coefficients, c_n, that can best satisfy the envelope constraints. In this manner, the derived coefficients, c_n, n=0, 1, …, p, are called the discrete cepstrum for $\log|X(f)|$.

The envelope constraints just mentioned are actually L pairs of (f_k, a_k) for L representative spectral peaks selected from the original spectrum $\log|X(f)|$. Here, f_k and a_k represent the frequency (already normalized to the value range from 0 to 1) and amplitude of the k-th spectral peak, respectively. Note that L is usually larger than the cepstrum order, p. Hence, a least-squares criterion is adopted to minimize the approximation errors between $S(f_k)$

and a_k, k=1, 2, …, L. That is, the approximation error computed as

$$\varepsilon = \sum_{k=1}^{L} \left|\log a_k - \log S(f_k)\right|^2 \tag{4}$$

is to be minimized. This equation can be rewritten in a matrix form

$$\varepsilon = (A - M \cdot C)^{\mathrm{T}} (A - M \cdot C) \tag{5}$$

where A = [log(a_1), log(a_2), …, log(a_L)]$^{\mathrm{T}}$, C is a column vector of (p+1) discrete cepstrum coefficients to be derived, *i.e.* $C = [c_0, c_1, \ldots, c_p]^{\mathrm{T}}$, and

$$M = \begin{bmatrix} 1 & 2\cos(2\pi f_1) & 2\cos(2\pi f_1 \cdot 2) & \cdots & 2\cos(2\pi f_1 \cdot p) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 2\cos(2\pi f_L) & 2\cos(2\pi f_L \cdot 2) & \cdots & 2\cos(2\pi f_L \cdot p) \end{bmatrix}$$

When the error ε of Equation (4) is minimized with the least-square criterion, the optimal values of the discrete cepstrum coefficients can be derived to be

$$C = (M^{\mathrm{T}} \cdot M)^{-1} \cdot M^{\mathrm{T}} \cdot A \tag{6}$$

That is, by just executing the operations of matrix inversion and multiplication, the values of the discrete cepstrum coefficients can be obtained.

2.2 Regularization of Discrete Cepstrum

According to Equations (5) and (6), it is seen that the discrete cepstrum coefficients are derived in the frequency domain with the least-square criterion. Nevertheless, such a derivation method may encounter a vital problem in practice. That is, the spectral envelope computed according to Equation (3) may vibrate radically and have very large approximation error at some frequencies slightly away from the selected spectral-peak frequencies, f_k. This is because the direct estimation method (*i.e.* Equation (6)) may sometimes be ill-conditioned. That is, slightly varying the frequency values of the detected spectral peaks may result in a very different spectral envelope curve being obtained. For example, look at the dash-lined spectral envelope in Figure 3. This curve is computed with 40 derived discrete cepstrum coefficients. Even though the first 7 spectral peaks are passed by this curve, the curve vibrates radically between two adjacent peaks. In practical applications, such a spectral envelope as the one in Figure 3 cannot be tolerated.

Therefore, Cappé and Moulines (1996) proposed a regularization technique to prevent such radical vibrations from occurring. To do this, they added a curve-sharpness penalty term to the approximation-error calculation equation, *i.e.* Equation (4), thereby making the approximation-error calculation equation:

$$\varepsilon = \sum_{k=1}^{L} \left| \log a_k - \log S(f_k) \right|^2 \ + \ \lambda \cdot R\big(S(f)\big) \tag{7}$$

where the function $R(\cdot)$ is intended to measure the sharpness of the spectral-envelope curve $S(f)$, and λ is a parameter to adjust the relative weight of the value returned by $R(\cdot)$. A typical function suggested for $R(\cdot)$ is

$$R\big(S(f)\big) = \int_0^{\pi} \left[\frac{d}{df} S(f) \right]^2 df \quad . \tag{8}$$

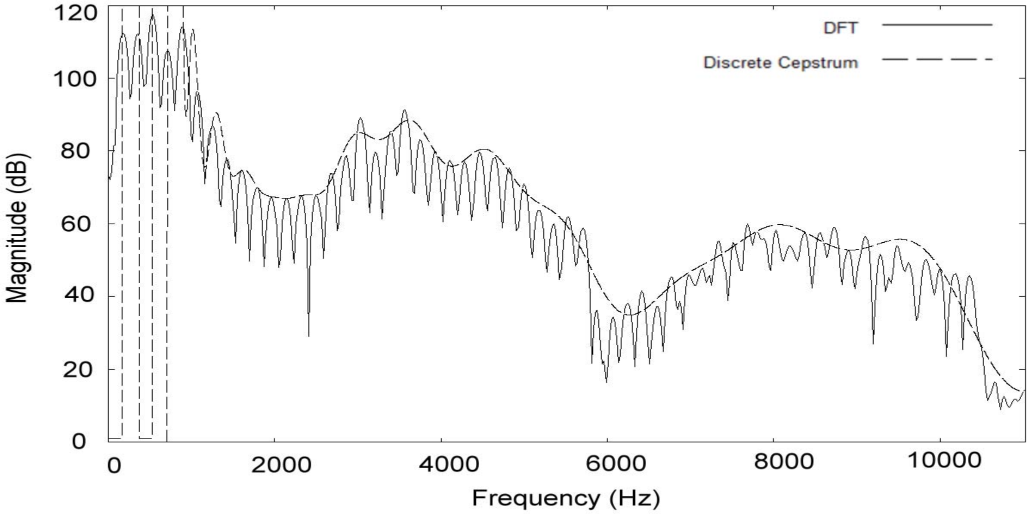


Figure 3. A spectral envelope computed with 40 non-regularized discrete cepstrum coefficients.

When the definition of $S(f)$ given in Equation (3) is taken into Equation (8), the following equation,

$$R\big(S(f)\big) = C^{\mathrm{T}} \cdot U \cdot C \quad , \quad U = 8\pi^2 \begin{bmatrix} 0 & & & 0 \\ & 1^2 & & \\ & & \ddots & \\ 0 & & & p^2 \end{bmatrix}, \tag{9}$$

can be derived (Cappé & Moulines, 1996; Stylianou, 1996). Then, the optimal solution that minimizes the error calculated in Equation (7) can be derived via:

$$\begin{aligned} &\varepsilon = (A - MC)^{\mathrm{T}} (A - MC) + \lambda C^{\mathrm{T}} UC \ ; \\ &\text{Let} \quad \frac{\partial \varepsilon}{\partial C} \ = \ -2 \cdot M^{\mathrm{T}} (A - MC) + 2 \cdot \lambda UC = 0 \ ; \\ &(M^{\mathrm{T}} M + \lambda U) \cdot C = M^{\mathrm{T}} A \ ; \\ &C = (M^{\mathrm{T}} M \ + \ \lambda U)^{-1} \cdot M^{\mathrm{T}} A \quad . \end{aligned} \tag{10}$$

According to empirical experience, the parameter λ is better set to a value around 0.0001. Thereafter, the ill-conditioning problem can be solved, and a regularized spectral envelope curve will be obtained. An example of a regularized spectral envelope is the dash-lined curve in Figure 4, where the solid-lined curve is the same as the one in Figure 3. Apparently, the phenomenon of radical vibration is not seen in this figure. Note that frequency axis scaling (to be discussed in Section 4.2) is already applied in addition to regularization to obtain the spectral envelope in Figure 4.

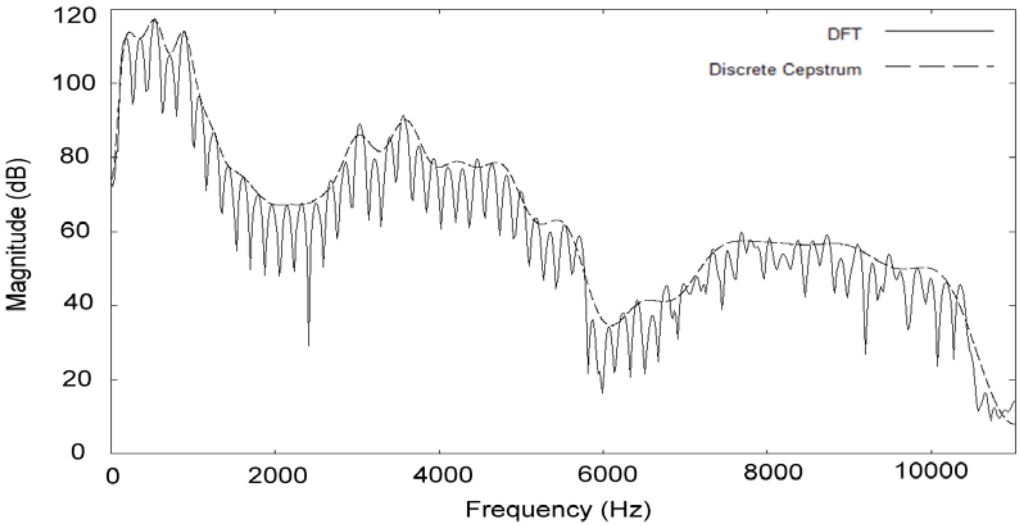


Figure 4. A spectral envelope computed with 40 regularized discrete cepstrum coefficients.

3. Selection of Spectral Peaks

Note that the discrete cepstrum coefficients are obtained by minimizing the squared errors between the selected spectral peaks, a_k, k=1, 2, …, L, and S(f). Therefore, selecting appropriate spectral peaks from a DFT spectrum is an important preprocessing step. The simplest selection method, *i.e.* locating and selecting all the spectral peaks on the spectrum as the final selected peaks, leads to the approximated spectral envelope being very bad and having a large approximation error. When such bad spectral envelopes are used to transform voice signals, the output obtained will suffer significant voice-quality degradation.

Therefore, we studied this problem and found that the concept of MVF (maximum voiced frequency) proposed in HNM (harmonic-plus-noise model) (Stylianou, 1996; Stylianou, 2005) is utilizable. The MVF of a DFT spectrum is searched by testing the sharpness of the spectral peaks one after another. After some low-frequency spectral peaks pass the test, eventually no more spectral peak can pass the test. Then, the frequency of the last spectral peak passing the test is defined to be the MVF. In this paper, we first detect if a signal frame is voiced or unvoiced. If it is detected to be voiced, the frame is further searched for the MVF value, *fv*, through the searching method proposed by Stylianou (1996). According to *fv*, the DFT spectrum of the frame is split into the lower-frequency harmonic part and the higher-frequency noise part. Then, for the harmonic part, the first spectral peak of a frequency within the range ($0.5\times F_0$, $1.5\times F_0$), where F_0 is the detected fundamental frequency, is searched for. Let the obtained frequency and amplitude be f_1 and a_1. Next, the second spectral peak of a frequency within the range ($f_1+0.5\times F_0$, $f_1+1.5\times F_0$) is searched for, and the results will be the frequency f_2 and amplitude a_2. When going on in this manner, we can find the frequencies and amplitudes of the other spectral peaks within the harmonic part. Sometimes, it may occur that no spectral peak is found within a designated frequency range. In this situation, we will right shift the frequency range, *i.e.* adding $0.5\times F_0$, and try to find again.

For the noise part of a voiced frame, we think the searching method explained above for the harmonic part cannot be adopted. Note that the harmonic structure becomes obscure in the noise part, and the frequency gaps between adjacent peaks become randomly varied. As an example, inspect the DFT spectrum curve beyond 5,800Hz in Figure 4. Therefore, we adopt another method to find the spectral peaks for the noise part. In this method, a smooth spectral curve is obtained first by truncating the real-cepstrum coefficients outside the leading 30 ones, and transforming (via DFT) the resulting real-cepstrum sequence back to the spectrum domain. Then, each spectral speak within the noise part is located and its amplitude is checked. It will be selected if its amplitude is higher than the height of the smooth spectral curve at the peak's frequency. As for an unvoiced frame, the method just explained can still be applied. This is because such a frame's MVF can be directly set to 0Hz and its spectrum can be viewed as all in the noise part. When applying the spectral peak selection method explained above, we may obtain a typical result shown in Figure 5. In this figure, each occurrence of plus-sign, +, represents a selected spectral peak.

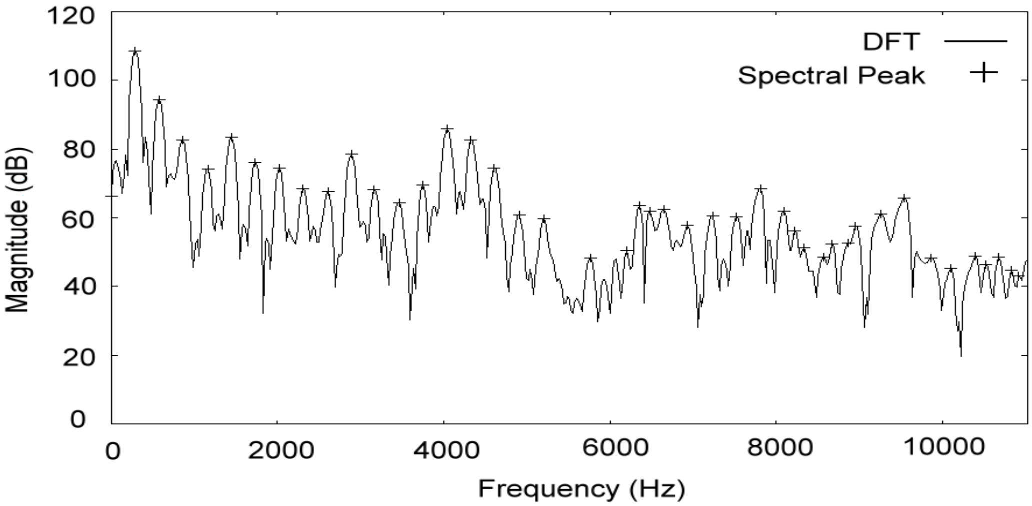


Figure 5. A typical result for spectral peak selection.

4. Order of Discrete Cepstrum and Frequency Axis Scaling

4.1 Order of Discrete Cepstrum

What value should be set for the parameter, p, for the order of a discrete cepstrum? If a smaller value (*e.g.* 10) is set, the approximated spectral envelope curve will be overly smooth, and the approximation error will become considerably large. On the other hand, when a larger value is set for p, the number of computation operations needed to solve Equation (10) will rise at a cubic rate. Nevertheless, sufficient accuracy in spectral-envelope approximation is very important to prevent quality-degradation and inconsistent-timbre from occurring. For setting the value of p, Shiga and King (2004) argue that p should have a value from the range (48, 64) to obtain a sufficiently accurate approximation of spectral envelope.

Here, we study the correlation between approximation errors and order numbers, p, experimentally. The approximation error is computed as

$$Es = \frac{1}{Nr}\sum_{t=0}^{Nr\text{-}1}\left[\frac{1}{L(t)}\sum_{k=1}^{L(t)}\left|20\log_{10} a_k^t - 20\log_{10} S(t, f_k)\right|\right] \tag{11}$$

where Nr is the total number of frames, a_k^t denotes the amplitude of the k-th spectral peak in the t-th frame, and S(t, f) represents the approximated spectral envelope for the t-th fame. Here, 375 Mandarin sentences, consisting of 2,925 syllables, were recorded from a male speaker and used as the testing data. According to our measurement results, the approximation error, Es, will decrease considerably as the order number p is increased from 5 to 30. Thereafter, Es will decrease only slightly as p is further increased to 50. This trend is illustrated by the curve in Figure 6. Therefore, we decide here to adopt 40 as the order number for p in order to obtain sufficiently accurate spectral-envelop approximation.

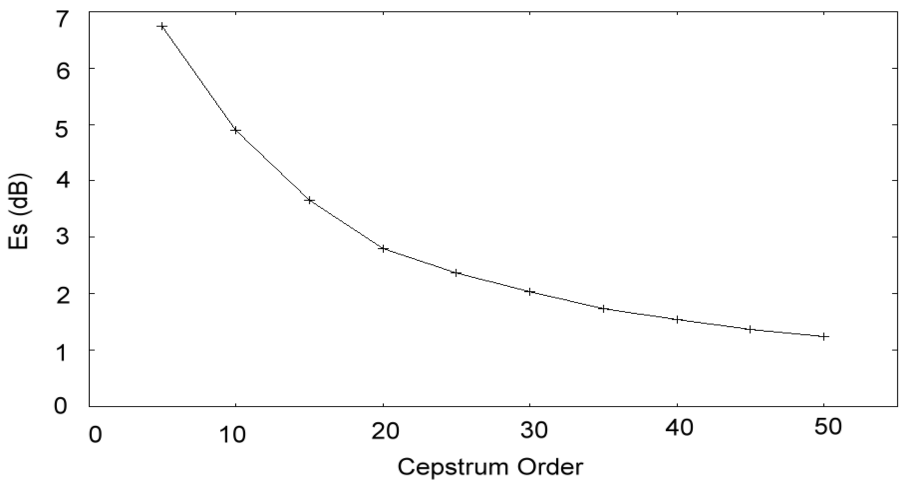
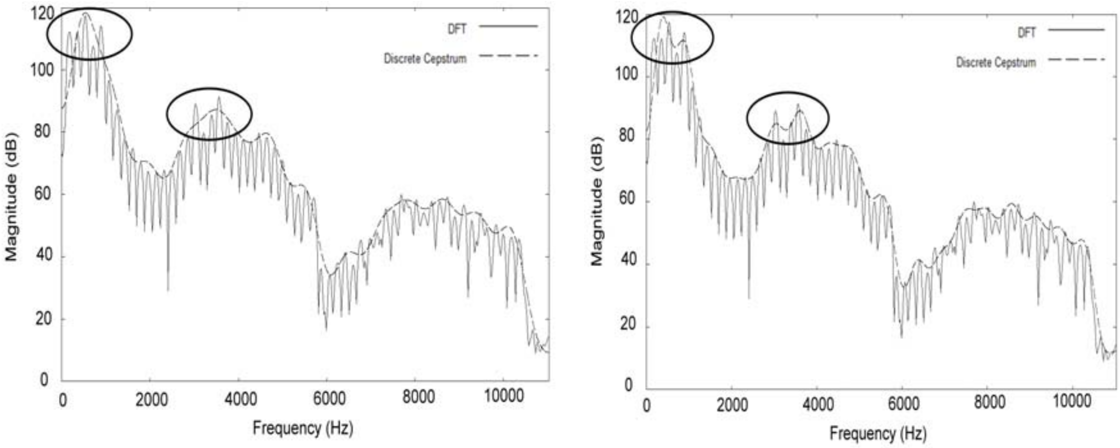


Figure 6. Approximation errors versus discrete cepstrum orders.

4.2 Frequency Axis Scaling

Through use of a larger order number, *e.g.* 40, the global approximation error of a frame's spectral envelope can now be controlled. Nevertheless, local approximation errors that are large and cannot be ignored can still be found. For example, the spectral envelope in Figure 7(a) is obtained by approximating with 30 discrete cepstrum coefficients, and it has two significant local approximation errors as circled. If the value of the order number p is increased to 40, the spectral envelope will become the one shown in Figure 7(b). Although the local approximation errors are somewhat reduced, they are still significant and cannot be ignored.



(a) Cepstrum order set to 30 (b) Cepstrum order set to 40

Figure 7. Spectral envelopes approximated in the linear frequency scale.

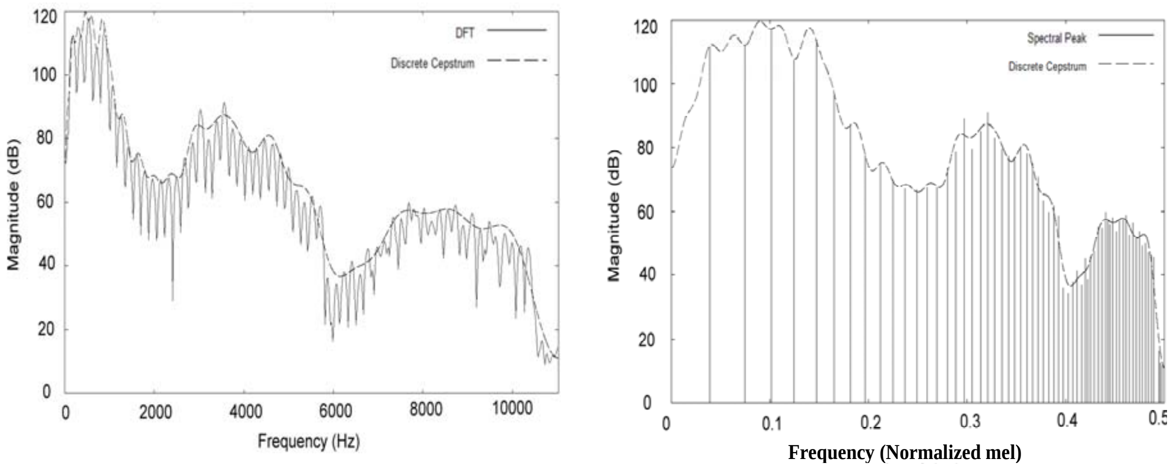
Consider the local approximation error around 1,000Hz, as seen in Figures 7(a) and 7(b). From the DFT spectrum, it is known that the pitch is low (*i.e.* it is a male's pitch) and the frequency gap between two adjacent spectral peaks is small. Under such a situation, the amplitudes of the third and fifth spectral peaks show rapid growth that is higher than the nearby peaks. Such rapid change of spectral envelope is very hard to approximate. To solve this situation, a conventional idea is to nonlinearly scale the frequency axis to enlarge the frequency gaps between low-frequency spectral peaks. Therefore, when a spectral peak of a frequency f_k is detected, its frequency value will be scaled to $\hat{f}_k$ according to the formula,

$$\hat{f}_k = \frac{1}{2} \cdot \frac{scl(f_k \times F_s)}{scl(0.5 \times F_s)} , \tag{12}$$

where $scl(\cdot)$ represents a frequency-scale conversion function and F_s is the sampling frequency. After frequency scaling, the spectral peaks' frequencies and amplitudes are then taken into Equation (10) to compute the optimal discrete cepstrum coefficients. The step of replacing f_k

with $\hat{f}_k$ also implies that the computed discrete cepstrum coefficients must be used in the scaled frequency axis instead of the original axis. That is, for a linear and normalized frequency f, its envelope magnitude, $S(f)$, should be computed by scaling f into $\hat{f}$ first then taking $\hat{f}$ into Equation (3).

For nonlinear frequency-axis scaling, mel and Bark frequency scales are the most famous (O'Shaughnessy, 2000). If the frequency conversion function, $scl(\cdot)$, adopted is a mel-frequency conversion, the spectral envelope shown in Figure 7(b) will be changed to the one shown in Figure 8(a). The major difference is that the lower-frequency spectral peaks in Figure 8(a) are now all passed by the approximated spectral envelope curve. Nevertheless, the local approximation error around 3,000Hz is still noticeable. In addition, the spectral envelope curve in Figure 8(a) shows much stronger vibration near the lower-frequency end than the curve in Figure 7(b). This stronger vibration can be seen in more detail when we redraw the approximated spectral envelope curve with a mel-frequency horizontal axis as shown in Figure 8(b). This phenomenon of over-vibration is thought to be due to the mel-frequency conversion that widens the frequency-scale at the low frequency end. According to the observed stronger vibration for mel-frequency conversion, we think much stronger vibration will occur if we adopt the Bark-frequency conversion for $scl(\cdot)$. This is because Bark-frequency conversion will have the frequency-scale at the low frequency end being widened more than that widened by the mel-frequency conversion.



(a) linear-frequency horizontal axis (b) mel-frequency horizontal axis

Figure 8. Spectral envelopes approximated in mel-frequency scale.

Therefore, we were motivated to design a frequency conversion function in the hope of eliminating the phenomenon of over-vibration at the low frequency end and of reducing the local approximation error around 3,000Hz. After several attempts at trying function-design and inspecting the approximated spectral envelope curves, we finally found a better frequency conversion function:

$$scl(f) = \log(1 + \frac{f}{1{,}750}) \qquad (13)$$

where f is in the unit Hz. This conversion function will have the scaled frequency value, $\hat{f}_k$, growing more slowly with f_k at the low frequency end when it is used as the $scl(\cdot)$ function for Equation (12). The three curves shown in Figure 9 are obtained by taking Bark, mel, and our frequency conversions, respectively, as the $scl(\cdot)$ function for Equation (12). From Figure 9, it can be seen that our frequency conversion, as given in Equation (13), can indeed grow the scaled frequency $\hat{f}$ more slowly with the linear frequency f. Via the frequency conversion function of Equation (13), the approximated spectral envelope in Figure 8(a) will become the one drawn in Figure 4. According to the spectral envelope obtained in Figure 4, it can be said that the frequency conversion function proposed can indeed eliminate the over-vibration phenomenon at the low frequency end, and reduce the local approximation error around 3,000Hz. The reducing of the local approximation error we think is due to the increased vibrating capability around 3,000Hz by using the proposed frequency conversion instead of the mel-frequency conversion.

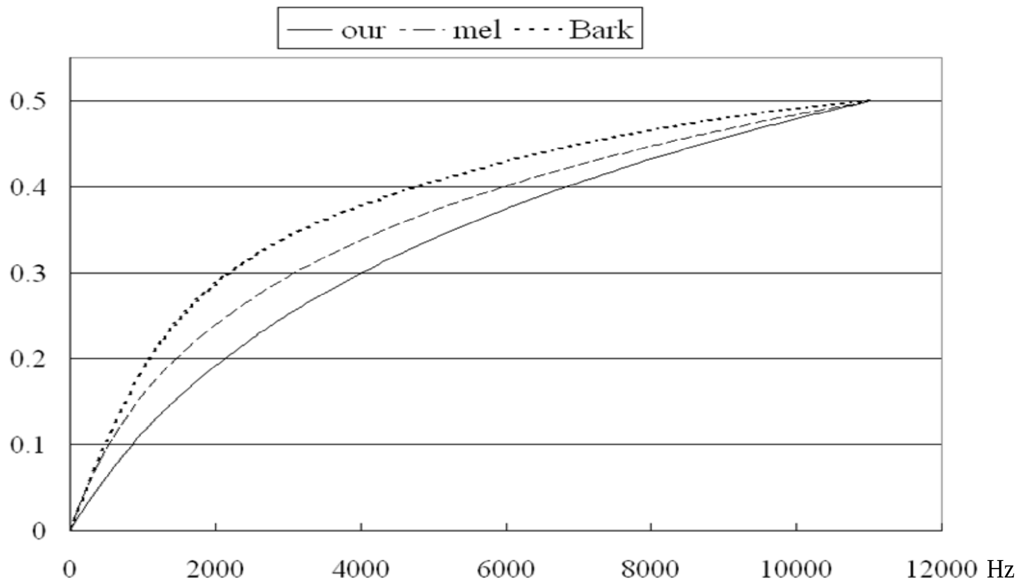


Figure 9. Three curves of scaled frequencies by using Bark, mel, and our frequency conversion functions, respectively.

4.3 Approximation Error Comparison

One may ask if our frequency conversion function is only better than mel-frequency conversion for certain signal frames. Therefore, we decided to compare the approximation errors of the two frequency conversions in the four frequency ranges, *i.e.* 0 ~ 2,000Hz, 0 ~ 4,000Hz, 0 ~ 6,000Hz, and 0 ~ 11,025Hz. Here, the approximation error is still measured by the formula in Equation (11). Nevertheless, the number of spectral peaks, L, is dynamically checked for each frame to ensure that only the spectral peaks of frequencies within the currently concerned frequency range are counted. Here, 375 Mandarin sentences, consisting of 2,925 syllables, as mentioned in Section 4.1 were used as the testing data. After all of the frames of the data are processed, the approximation errors measured in different frequency ranges and different discrete cepstrum orders are illustrated in Figure 10.

Inspecting the error curve in Figure 10, it can be seen that across the cepstrum order-numbers from 30 to 50, our frequency conversion and the mel-frequency conversion have almost same approximation errors in the frequency range, 0 ~ 2,000Hz. Nevertheless, in the other three frequency ranges, our conversion function will apparently obtain smaller approximation errors for different cepstrum-order numbers. This decreasing of approximation error becomes more apparent as the frequency range becomes wider.

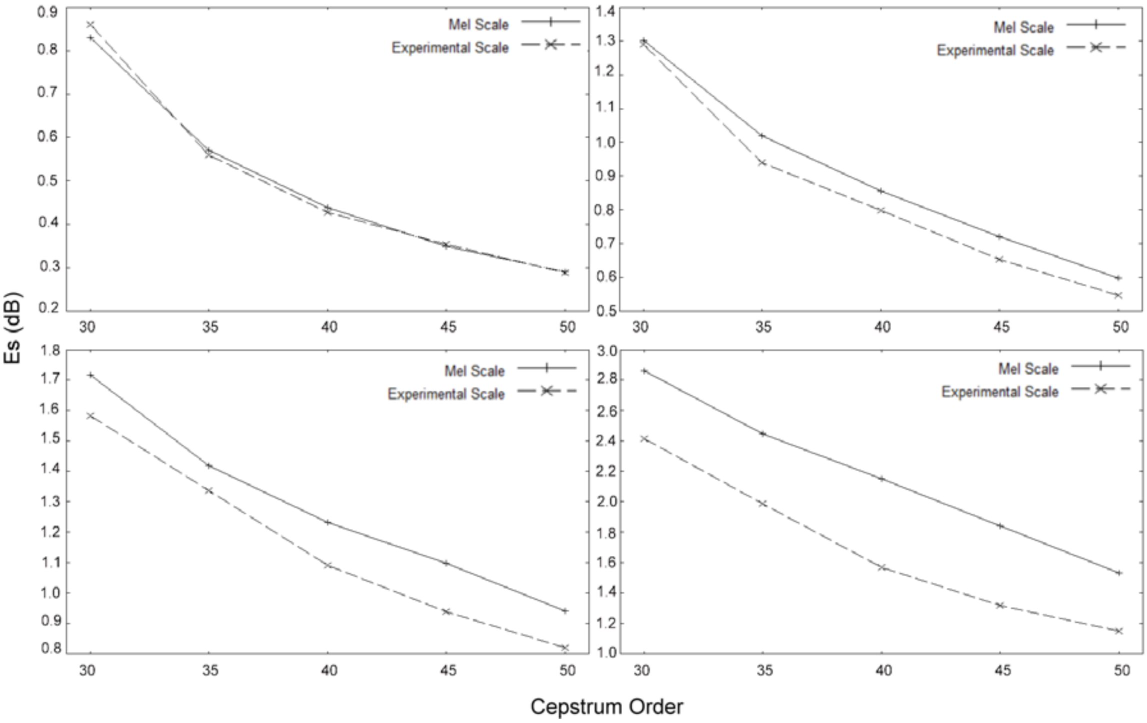


Figure 10. Approximation errors measured for our frequency conversion and mel-frequency conversion in the four frequency ranges: 0 ~ 2,000Hz (upper left), 0 ~ 4,000Hz (upper right), 0 ~ 6,000Hz (bottom left), and 0 ~ 11,025Hz (bottom right).

5. An Example Application: Voice Transformation

Here, voice transformation means to change the timbre of an input voice to a different timbre. For example, it could be changing the timbre of a female adult into the timbre of a male adult or a child. In the past, phase vocoder was a frequently used technique to transform voice timbre (Moore, 1990; Dolson, 1986). Nevertheless, the basic transformation method of phase vocoder cannot support independent control of spectral-envelope scaling and pitch shifting. Therefore, we decided to apply the technique of additive synthesis developed for computer music synthesis (Moore, 1990) and the signal model of HNM (harmonic-plus-noise model) (Stylianou, 1996; Stylianou, 2005). In other words, we will use the estimated spectral envelope to do spectral-envelope scaling. Then, we will place harmonic partials and noise sinusoids under the scaled spectral envelope according to the pitch shifting requirement. In this manner, spectral envelope scaling and pitch shifting can be performed independently.

We have practically implemented a voice transformation system. Its main processing flow is as shown in Figure 11. In this system, the input voice is first sliced into a sequence of frames. The frame width is 512 sample points (23.2ms) and the frame shift is 256 points (11.6ms) under the sampling frequency, 22,050Hz. For each frame, the processing flow shown in Figure 2 is executed to estimate its spectral envelope with 40 discrete-cepstrum coefficients. The other three blocks, “spectral envelope scaling,” “pitch shifting,” and “signal re-synthesizing,” will be explained in the following subsections. We tested the processing speed of this system on a notebook computer with an Intel T5600 1.83GHz CPU, and found that it will consume 0.75 sec. of CPU time on average to transform 1 sec. of voice signal.

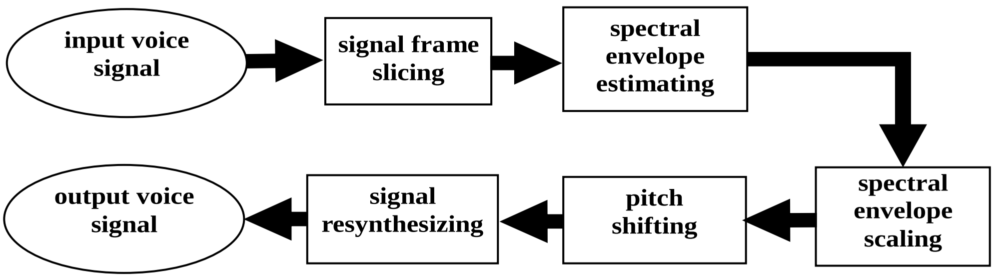
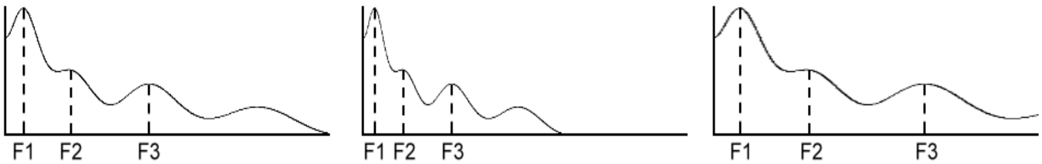


Figure 11. Main processing flow of the voice transformation system.

5.1 Spectral Envelope Scaling

Scaling of a spectral envelope can be performed in two possible directions. One direction is to shrink the spectral envelope to lower formant frequencies in order to obtain a male adult’s timbre. The other direction is to extend the spectral envelope to raise formant frequencies in order to obtain a child’s timbre. For example, inspect the spectral envelopes drawn in Figure 12. The curve drawn in Figure 12(a) represents the originally estimated spectral envelope, *Vo*(*f*). If this spectral envelope is shrunk and the shrinking rate is 0.7, the resulting envelope will be the one drawn in Figure 12(b). Apparently, the formant frequencies, F1, F2, and F3, are all lowered. Let the spectral envelope in Figure 12(b) be denoted by *Vs*(*f*). Then, it is simple to derive that $Vs(f) = Vo(\frac{10}{7}f)$. On the other hand, if the spectral envelope in Figure 12(a) is extended and the extending rate is 10/7, the resulting envelope will be the one drawn in Figure 12(c). Apparently, the formant frequencies, F1, F2, and F3, are all raised. Let the spectral envelope in Figure 12(c) be denoted with *Ve*(*f*). Then, it can be derived that $Ve(f) = Vo(\frac{7}{10}f)$.



(a) original envelope (b) shrunk envelope (c) extended envelope

Figure 12. The scaling of an example spectral envelope.

5.2 Pitch Shifting

After a frame's spectral envelope is shrunk (or extended), we can use *Vs*(*f*) (or *Ve*(*f*)) to determine a new set of harmonic partials and noise sinusoids. Suppose that the original pitch frequency of the *i*-th frame is 180Hz and that we intend to tune its pitch to 250Hz. Although the original pitch must be used in the block "spectral envelope estimating" of Figure 11, it is not used for pitch shifting and signal re-synthesizing. This is because we just need *Vs*(*f*) (or *Ve*(*f*)) to determine the amplitudes of the new harmonic partials and noise sinusoids. For example, the new harmonic structure of a voiced frame may look like the one shown in Figure 13. According to a given MVF, we can place new harmonic partials into the frequency range below MVF, and place new noise sinusoids into the frequency range above MVF.

In detail, the frequencies of the new harmonic partials are set as f_1^i=250, f_2^i=500, f_3^i=750, *etc*. As to their amplitudes, the spectral envelope, *Vs*(*f*) (or *Ve*(*f*)), is evaluated at the targeted frequencies. That is, their amplitudes are set to a_1^i = *Vs*(250), a_2^i = *Vs*(500), a_3^i = *Vs*(750), *etc*. Besides frequency and amplitude, the other parameter of a harmonic partial is phase. Nevertheless, the phase values of the harmonic partials are not a concern here because they will not be used in the signal re-synthesizing step. For the noise sinusoids, any two adjacent ones are placed 100Hz apart as shown in Figure 13. After being placed, each noise sinusoid's amplitude can be determined according its frequency position.

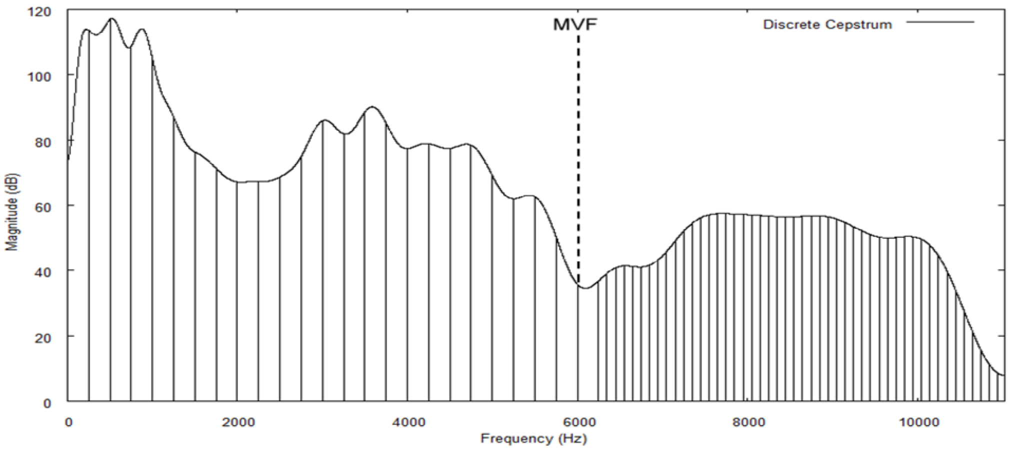


Figure 13. An example harmonic structure for a frame tuned to have the pitch, 250Hz.

5.3 Signal Re-synthesizing

Here, the signal model, HNM, proposed by Stylianou (1996) is adopted for signal re-synthesis. In HNM, the spectrum of a voice frame is divided into the lower-frequency harmonic part and the higher-frequency noise part. The frequency that the two parts are divided according to is called the MVF. In the original work (Stylianou, 1996), a method is provided to dynamically detect each frame's MVF. Here, to simplify the synthesis processing, we just use the static MVF value, 6,000Hz, across all voiced frames. As an example, the spectral envelope in Figure 13 is divided into the harmonic and noise parts according to the MVF, 6,000Hz.

Suppose the *i*-th and (*i*+1)-th frames are both voiced and have L^i and L^{i+1} harmonic partials, respectively, after pitch shifting. To synthesize a signal sample for the *t*-th sampling point between the *i*-th and (*i*+1)-th frames, we first derive the frequencies and amplitudes of the harmonic partials for this sampling point with linear interpolation. That is,

$$f_k(t) = f_k^i + \frac{f_k^{i+1} - f_k^i}{N} t, \quad k = 1, 2, ..., L ,$$
$$a_k(t) = a_k^i + \frac{a_k^{i+1} - a_k^i}{N} t, \quad k = 1, 2, ..., L \qquad (14)$$

where N is the number of sampling points between two adjacent frames, and L is the larger one of L^i and L^{i+1}. Here, we directly set $a_k^i = 0$, $k = L^i + 1, \ldots, L^{i+1}$, if L^i is less than L^{i+1}. Then, the harmonic signal, $h(t)$, for the *t*-th sampling point is computed as

$$h(t) = \sum_{k=1}^{L} a_k(t) \cdot \cos(\phi_k(t)), \quad 0 \le t < N ,$$
$$\phi_k(t) = \phi_k(t-1) + 2\pi \cdot f_k(t) / 22{,}050 \qquad (15)$$

where $\phi_k(t)$ denotes the accumulated phase on time t for the k-th harmonic partial and 22,050 is the sampling frequency. $\phi_k(-1)$ is equal to $\phi_k(N\text{-}1)$ of the last frame to keep continuity of phase. If i = 0, *i.e.* there is no last frame, the value of $\phi_k(-1)$ is set randomly.

To synthesize the noise signal for the t-th sampling point, we apply a method mentioned in Stylianou (1996). That is, synthesize the noise signal as the summation of the sinusoids whose frequencies are larger than MVF, fixed (not affected by the pitch frequency) and are 100Hz apart. The amplitudes of the sinusoids are, however, varied with time. Let KL = MVF / 100 and KU = 22,050 / 100. Then, the noise signal, $g(t)$, is synthesized as:

$$\begin{aligned} g(t) &= \sum_{k=KL}^{KU} b_k(t) \cdot \cos(\psi_k(t)), \quad 0 \le t < N \ , \\ \psi_k(t) &= \psi_k(t-1) + 2\pi \cdot k \cdot 100 / 22,050 \end{aligned} \tag{16}$$

where $b_k(t)$ and $\psi_k(t)$ denote the amplitude and accumulated phase, respectively, of the k-th sinusoid at the time point t. The value of $b_k(t)$ is obtained, with linear interpolation, with a formula similar to the one in Equation (15). Finally, the signal sample for the t-th sampling point is synthesized as $h(t)$ plus $g(t)$.

5.4 Perception Testing

To evaluate the performance of our voice transformation system, we first recorded three sentences each from a female adult and a male adult. For the female source voice, we set the envelope shrinking rate to 0.8 and set the pitch shifting rate to 0.6 in order to transform into a male timbre. For the male source voice, we set the envelope extending rate to 1.2 and set the pitch shifting rate to 2.1 in order to transform into a female timbre. The source voices and their transformed voices can be accessed at http://guhy.csie.ntust.edu.tw/dcc/vt.html.

We invited thirteen persons to participate in the perception tests. The first type of perception test conducted was for evaluating the timbre recognizability of the source and transformed voices. That is, each participant was asked how similar the timbre of a played voice was to a female (or a male). Each participant was requested to give a score between 1 and 5 to indicate how similar the heard timbre seemed. As a result, we obtained the averaged scores and standard deviations shown in Table 1. According to the average scores of the transformed timbres, *i.e.* 4.85 and 4.36, it can be said that the transformed voice from our system will have sufficiently high timbre recognizability. In addition, when comparing the score differences between the original and transformed voices, *i.e.* 0.10 (4.95-4.85) vs. 0.37 (4.73-4.36), we find that the female source voice will induce less recognizability degradation than the male source voice.

Table 1. Perception test results for timbre recognizability.

Source voice \ Timbre		Original voice	Transformed voice
Female	Avg. score	4.95	4.85
	Std. deviation	0.15	0.23
Male	Avg. score	4.73	4.36
	Std. deviation	0.45	0.48

The second type of perception test is for evaluating the voice qualities of the source and transformed voices. The same participants were asked what level the quality of a played voice was at. Each participant was requested to give a score between 1 and 5 to indicate the quality level of the played voice. After this type of perception test was conducted, we averaged the scores collected and computed their standard deviation. The results are shown in Table 2. According to the score differences, 0.67 (4.38 – 3.71) and 0.82 (4.00 – 3.18), it can be said that our system will inevitably induce a perceivable degradation of voice quality for the transformed voices no matter whether the source voice is uttered by a female or male. One of the possible reasons is that the pitch frequencies of some frames are wrongly detected, which causes their spectral envelopes to be approximated with noticeable errors.

Table 2. Perception test results for voice quality.

Source voice \ Timbre		Original voice	Transformed voice
Female	Avg. score	4.38	3.71
	Std. deviation	0.39	0.53
Male	Avg. score	4.00	3.18
	Std. deviation	0.74	0.72

6. Concluding Remarks

The concept of approximating spectral envelope with discrete cepstrum was proposed several years ago. There are, however, three problems that must be solved for practical implementation. The first problem is the regularization of the discrete cepstrum coefficients to prevent a radical vibrating envelope curve from occurring. This problem has been solved already by previous researchers. In this paper, we tried to solve the other two problems, *i.e.* selecting appropriate spectral speaks and finding a better frequency axis scale. For selecting spectral peaks, we apply the concept of HNM to divide a spectrum into the lower frequency harmonic part and the higher frequency noise part. Then, we find the spectral peaks in the harmonic part according to the detected pitch frequency and screen the spectral peaks in the noise part according to a cepstrum smoothed spectral curve. As to the problem of frequency axis scaling, we found that the spectral envelope approximated via the mel or Bark-frequency conversion still has noticeable local approximation errors. Therefore, after some attempts at scaling-function design, we propose a better frequency conversion function that can reduce the local approximation errors significantly. Then, applying the solutions to the three problems, we construct a spectral envelope estimation scheme.

In addition, we built a voice transformation system on the proposed spectral envelope estimation scheme as an example application. This system follows the steps, spectral envelope estimating, spectral envelope scaling, pitch shifting, and signal re-synthesizing, to transform an input voice into an output voice that is of a very different timbre, *i.e.* the perceived gender and age of the voice can both be changed. To evaluate the performance of this system, we conducted perception tests. The averaged scores from 13 participants show that our system can indeed achieve the function of timbre transformation. In the future, we will apply the proposed spectral envelope estimation scheme to study another kind of voice transformation problem. That is, we will convert the voice of a specific person into the voice of another specific person.

Acknowledgments

This study is supported by National Science Council under the contract number, NSC 98-2221-E-011-116.

References

Cappé, O., & Moulines, E. (1996). Regularization techniques for discrete cepstrum estimation. *IEEE Signal Processing Letters*, 3(4), 100-102.

Dolson, M. (1986). The phase vocoder: A tutorial. *Computer Music Journal*, 10(4), 14-27.

Galas, T., & Rodet, X. (1990). An improved cepstral method for deconvolution of source filter systems with discrete spectra: Application to musical sound signals. *Proceedings of International Computer Music Conference*, Glasgow, Scotland.

Gu, H. Y., Chang, H. F., & Wu, J. H. (2004). A pitch-contour normalization method following Zhao’s pitch scale and its application. *Proceedings of ROCLING XVI*, Taipei, Taiwan. 325-334. (in Chinese)

Imai, S., & Abe, Y. (1979). Spectral envelope extraction by improved cepstral method. *Electron. and Commun. in Japan*, 62-A(4), 10-17. (in Japanese)

Kawahara, H., Masuda-katsuse, I., & Cheveign, A. De. (1999). Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: possible role of a repetitive structure in sounds. *Speech Communication*, 27, 187-207.

Kim, H. Y., *et al.* (1998). Pitch detection with average magnitude difference function using adaptive threshold algorithm for estimating shimmer and jitter. *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Hong Kong, China.

Moore, F. R. (1990). *Elements of computer music*. Prentice-Hall, Englewood Cliffs, NJ, U.S.A.

O'Shaughnessy, D. (2000). *Speech communications: human and machine*, IEEE Press, Piscataway, NJ, U.S.A.

Robel, A., & Rodet, X. (2005). Efficient spectral envelope estimation and its application to pitch shifting and envelope preservation. *Proceedings of International Conference on Digital Audio Effects*, Madrid, Spain.

Schwarz, D., & Rodet, X. (1999). Spectral envelope estimation and representation for sound analysis-synthesis. *Proceedings of International Computer Music Conference*, Beijing, China.

Shiga, Y., & King, S. (2004). Estimating detailed spectral envelopes using articulatory clustering. *Proceedings of International Conference on Spoken Language Processing* (ICSLP2004), Jeju, Korea.

Stylianou, Y. (1996). *Harmonic plus noise models for speech, combined with statistical methods, for speech and speaker modification*, Ph.D. thesis, Ecole Nationale Supèrieure des Télécommunications, Paris, France.

Stylianou, Y. (2005). Modeling speech based on harmonic plus noise models. in *Nonlinear Speech Modeling and Applications*, eds. Chollet, G., *et al.*, Springer-Verlag, Berlin, Germany.

Computational Linguistics and Chinese Language Processing
Vol. 14, No. 4, December 2009, pp. 383-402

Identification of Opinion Holders

Lun-Wei Ku[*], Chia-Ying Lee[*] and Hsin-Hsi Chen[*]

Abstract

Opinion holder identification aims to extract entities that express opinions in sentences. In this paper, opinion holder identification is divided into two subtasks: author's opinion recognition and opinion holder labeling. Support vector machine (SVM) is adopted to recognize author's opinions, and conditional random field algorithm (CRF) is utilized to label opinion holders. New features are proposed for both methods. Our method achieves an f-score of 0.734 in the NTCIR7 MOAT task on the Traditional Chinese side, which is the best performance among results of machine learning methods proposed by participants, and also it is close to the best performance of this task. In addition, inconsistent annotations of opinion holders are analyzed, along with the best way to utilize the training instances with inconsistent annotations being proposed.

Keywords: Opinion Holder Identification, Opinion Mining, Conditional Random Field, Support Vector Machine.

1. Introduction

Opinions describe subjective thinking of people. With the blooming of Web 2.0, a large number of free and online articles have become easily accessible. Although people are interested in the shifting of opinions, they cannot read such a large quantity of articles in a short time. Opinion mining can analyze opinions from many information sources automatically and helps extract opinions, along with determining their polarities, strength, holders, and targets. Opinion polarities tell us whether the current opinions are positive, neutral, or negative. The opinion strength then tells us the degree of their attitude, *i.e.*, strong, medium, or weak. Opinion holders are the people who express opinions, while opinion targets are the objects of those opinions. Let us take "Mr. Wang loves to play baseball" as an example. In this opinion sentence, its polarity is positive, its strength is strong, the opinion holder is Mr. Wang, and the opinion target is playing baseball. It is an opinion from Mr. Wang that indicates

---

[*] Department of Computer Science and Information Engineering, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei, 10617 Taiwan, TEL: +886-2-33664888*311, FAX: +886-2-23628167
E-mail: {lwku, cylee}@nlg.csie.ntu.edu.tw; hhchen@csie.ntu.edu.tw

he has a positive attitude towards playing baseball.

Opinion holder identification is useful in knowing who has the same attitude, what kind of issues a specific person cares about, and whether there are different opinions from some specific persons. This technique can also be applied to social network analysis to discover who the opinion leader is. It is also important in an opinion question answering system as it can provide the owner of opinions.

There are three major challenges in opinion holder identification: co-reference resolution, parsing nested opinions, and inconsistent annotation utilization. Like the conventional question answering problem, pronoun-antecedent and zero anaphor problems have to be solved before identifying opinion holders. Nested opinions are common in long sentences. People like to quote opinions of other people to show that they are impartial, but this behavior also implies that they agree with their quotes. In this case, we need to identify both the quoting and the quoted holders for further analysis.

It is sometimes difficult to determine the holder of an opinion. For example, even though the holder is obvious in the opinion sentence, we may find that he represents some organization and is presenting the organization’s opinion. The following sentence is an example: “According to the media, [the] U.S. and China are discussing the agreement of terminating the usage of nuclear weapons; Becon said they have discussed this issue before.” In this sentence, “they” refers to the U.S. and China. Becon quoted words from the U.S. and China, so this is a nested structure. In addition, although this expression is said by the media and Becon, the holder should be the U.S. and China. These challenges all complicate the annotation process, and a double check and a selection process are necessary when generating the gold standard.

2. Related Work

Pang and Lee (2008) have mentioned some important research projects in the domain of opinion mining. Kim and Hovy (2004) proposed four elements in opinion mining, including the opinion polarity, the opinion strength, the opinion holder, and the opinion target. Among them, the research for opinion holder identification is new. Previous researchers mainly have proposed two kinds of methods: heuristic rule based and machine learning based methods.

2.1 Heuristic Rule based Methods

For heuristic rule based methods, Seki *et al.* (2009) utilized noun phrases and linguistic features, and adopted SVM to classify opinion holders into authors and non-authors in English and Japanese materials. Xu and Wang (2008) first solved the co-reference resolution, and extracted opinion holders by rules involving punctuation marks, conjunctions, prefixes, suffixes, and opinion operators. They achieved an f-score of 0.825 in the NTCIR7 MOAT task on the Traditional Chinese side, which is the state of the art.

2.2 Machine Learning based Methods

For the machine learning based methods, many researchers have adopted maximum entropy algorithms, SVM, or the conditional random field model. Kim and Hovy (2006) utilized the maximum entropy model to extract opinion holders and targets from news articles. They first found opinion words and labeled semantic roles, then identified the semantic roles that are opinion holders and targets. Kim (2007, 2008) classified opinion holders into authors, simple holders and co-referenced holders, then extracted lexical and syntactic features for SVM to select the best opinion holder. So far, this is the best method for English materials, and it achieved an f-score of 0.346. Wu (2008) used words and parts of speech as features in L2-norm linear SVM to solve this research problem as a similar method for named entity identification. Breck and Choi (2007, 2005) utilized lexical features, syntactic features, dictionary-based features, and dependency features by CRF to identify opinion holders. Meng and Wang (2008) used words, parts of speech, and opinion operators, while Liu and Zhao (2008) extracted parts of speech, semantic features, contextual features, dependency features, and position features by CRF.

2.3 Proposed Methods

We propose a unique approach that divides opinion holder identification into two tasks: author's opinion recognition and opinion holder labeling. We then find better strategies to perform these two tasks. For author's opinion recognition, we adopt SVM by features such as words and their parts of speech, named entities, punctuation marks, the context, and opinion related information in the current sentence. Among them, some context features (the roles of verbs) and opinion related features (information of positive words, neutral words, negative words, and opinion operators) have not been utilized in opinion holder identification before. Detailed features will be described in Sections 3.2 and 3.3.

3. Opinion Holder Identification

Five procedures of opinion holder identification are proposed in this paper, including text pre-processing, author's opinion recognition, opinion holder labeling, post-processing, and result generation. Chinese word segmentation, parts of speech tagging, and named entity recognition are performed in the text pre-processing stage. Then, author's opinions are recognized and opinion holder labeling determines the text segment referring to the holder. We have two strategies for applying the proposed methods of author's opinion recognition and opinion holder labeling. These two strategies are described in Section 3.5. After that, this labeled text segment is processed by the post-processing procedure to generate the final opinion holder. The flowchart describing these five procedures is shown in Figure 1.

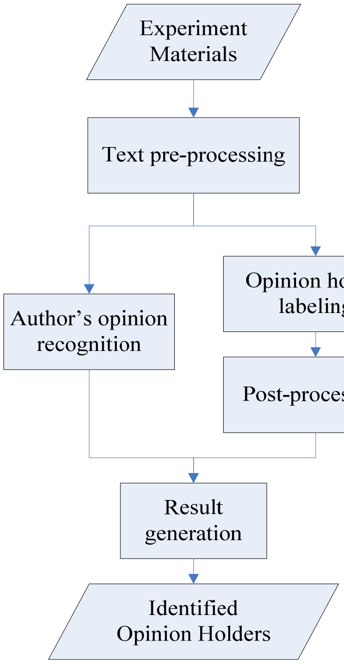


Figure 1. Flowchart of opinion holder identification

3.1 Text Pre-processing Stage

In the text pre-processing stage, we utilize the Chinese word segmentation system developed by Lo (2008). We, however, modify its segmentation module and add additional name dictionaries to it. The length limit of the modified Chinese name module is set looser and Japanese family names are added so that the segmentation system can recognize Japanese names, which are usually longer than Chinese names. Occupations, titles, and company names are also added to the dictionary of the segmentation system to provide useful holder relevant information.

3.2 Author's Opinion Recognition Stage

Author's opinion recognition finds out whether the opinion holder of the current sentence is the author. In this paper, this task is viewed as a binary classification task and LIBSVM (Chang & Lin, 2001) is adopted for classification. The main features extracted for this task are words, parts of speech, named entities, punctuation marks, sentence components, and opinion operators. Table 1 shows all of the features utilized in this task. The lexicon features include first person pronouns, which are often utilized by the authors to refer to themselves. The part-of-speech features include general pronouns and personal pronouns, because pronouns usually refer to persons and organizations and they can express opinions. The named-entity features are considered because they can either be the opinion holders or the opinion targets.

Table 1. Features for author's opinion recognition

Feature Type	Feature Name	Feature Description
Lexicon	FHasI	Does the word "I" (我) appear?
	FHasWe	Does the word "we" (我們) appear?
	fNumI	The number of the word "I" (我)
	fNumWe	The number of the word "we" (我們)
Part of speech	fHasPronoun	Are there any pronouns?
	fHasManPronoun	Are there any personal pronouns?
	fNumPronoun	The number of pronouns
	fNumManPronoun	The number of personal pronouns
Named entity	fHasPer	Is there a person name (named entity)?
	fHasLoc	Is there a location name (named entity)?
	fHasOrg	Is there an organization name (named entity)?
	fHasNa	Are there any common nouns?
	fHasNb	Are there any proper nouns?
	fHasNc	Are there any common location nouns?
	fNumLoc	The number of location names (named entity)
	fNumOrg	The number of organization names (named entity)
	fNumPer	The number of personal names (named entity)
	fNumNa	The number of common names
	fNumNb	The number of proper names
	fNumNc	The number of common location names
Punctuation mark	fHasExclamation	Is there an exclamation mark ("！" or "!") ?
	fHasQuestion	Is there a question mark ("？"or "?") ?
	fHasColon	Is there a colon ("："or ":") ?
	fHasLeftQuotation	Are there any quotation marks ("「" or "【") ?
	fHasRightQuotation	Are there any quotation marks ("」" or "】") ?
Sentential	fNumChar	The number of Chinese characters
	fNumWord	The number of Chinese words
	fNumSubsen	The number of clauses
Opinion	fOperator1 to 203	Is there an opinion operator?

The punctuation features include punctuation marks that possibly co-occur with opinions. For example, exclamation points and question marks often appear in sentences expressed by people because they can bear sentiment, whereas colons and quotations are usually used to quote expressed words. The sentential features tell the length of the sentences by their composite characters, words, and clauses. We consider these features because we think that authors may need a sentence of a suitable length to express opinions. As to the opinion features, a total of 203 opinion operators, such as 說 (say), 指出 (point out), and 認爲 (think), are collected manually from the earlier NTCIR corpus (Seki *et al*., 2008) for this task.

3.3 Opinion Holder Labeling Stage

Opinion holder labeling finds the text segment that represents the opinion holder. In the beginning, this task is also viewed as a binary classification problem for all words of a sentence, where the decision tree determines whether the current word is part of the opinion holder or not. CHAID decision tree algorithm provided by RapidMiner (Mierswa, Wurst, Klinkenberg, Scholz, & Euler, 2006) is adopted. It is a pruned decision tree using the chi-square test.

As the other alternative, we view the opinion holder labeling problem as a sequential labeling problem. Therefore, the CRF algorithm (Lafferty, McCallum, & Pereira, 2001) is selected to label whether each composite word is a portion of the opinion holder and CRF++ (Kudo, 2003) is adopted for implementation. Features for experiments are listed in Table 2.

Table 2. Features for opinion holder labeling

Feature Type	Feature Name	Feature Description
Lexicon	fWord	The current Word
Part of speech	fPOS	Part of speech of the current word
	fIsPronoun	Is the current word a pronoun?
	fIsNoun	Is the current word a noun?
Named entity	fIsPer	Is the current word a person name?
	fIsLoc	Is the current word a location name?
	fIsOrg	Is the current word an organization name?
Punctuation mark	fAfterParen	Does the current word appear one word after a parenthesis "」" or "】"?
	fBeforeColon	Does the current word appear one word before a colon "：" or ":"?
Sentential	fNearSenStart	Is the current word close to the sentence head?
	fSenLen	The number of words in the current sentence
	fWordOrder	The absolute position of the current word in the sentence
	fWordPerc	The absolute position (in percentage) of the current word in the sentence

Context	fNearVerb	The nearest verb in the current sentence
	fNearVerbPOS	The type (POS) of the nearest verb[1], *e.g.*, VA (transitive verb), VB (intransitive verb), *etc*.
	fDistNearVerb	The distance between the current word and its nearest verb
Opinion	fHasOpKW	Is there an opinion operator in this sentence?
	fHasPosKW	Are there any positive opinion words in this sentence, *e.g.*, 成功 (success), 同意 (agree)?
	fHasNegKW	Are there any negative opinion words in this sentence, *e.g.*, 錯誤 (wrong), 失敗 (fail)?
	fHasNeuKW	Are there any neutral opinion words in this sentence, *e.g.*, 不予置評 (no comment), 兩難 (a difficult choice)?
	fNearOpKW	Nearest opinion word in this sentence
	fNearPosKW	Nearest positive opinion word in this sentence
	fNearNegKW	Nearest negative opinion word in this sentence
	fNearNeuKW	Nearest neutral opinion word in this sentence
	fNearOpKWPOS	POS of the nearest opinion operator
	fNearPosKWPOS	POS of the nearest positive opinion word
	fNearNegKWPOS	POS of the nearest negative opinion word
	fNearNeuKWPOS	POS of the nearest neutral opinion word
	fDistOpKW	The distance to the nearest opinion operator in this sentence
	fDistPosKW	POS of the nearest NTUSD positive opinion word in this sentence
	fDistNegKW	POS of the nearest NTUSD negative opinion word in this sentence
	fDistNeuKW	POS of the nearest NTUSD neutral opinion word in this sentence

Features for opinion holder labeling include words, parts of speech, named entities, punctuation marks, sentential information, contextual information, and opinion related information. Some of the features are the same as those we have selected for the author's opinion recognition. The lexicon feature is the current word to be determined. The part-of-speech features of the current word include its part of speech, and whether it is a noun or a pronoun. The binary properties of being a noun or a pronoun are emphasized here because they are the most commonly seen parts of speech in opinion holders. Punctuation marks also are considered as features here. Sentential features tell the position of the current word in the current sentence. They are included in the feature set because, according to our observations,

[1] The part of speech tagging set is listed in Technical Report no. 95-02/98-04, Chinese Knowledge Information Processing Group, Academia Sinica.

holders often appear in the beginning or at the end of the sentence. The context features include the information of the nearest verbs with respect to the current word. If the current word is a part of the opinion holder, its nearest verb could be an opinion operator. The collocation of the current word and the nearest verb are considered. For the opinion information, the appearance of opinion operators, positive words, neutral words, and negative words are considered. Here, positive words are words used to express a supportive attitude, such as success, good, *etc*.; neutral words express an impartial attitude, such as no comment, difficult to say, *etc*.; negative words express opposite attitude, such as objection, accusation, *etc*. The occurrence of opinion words may indicate the existence of opinions, and opinions may further indicate the existence of their holders. The method of utilizing contextual and opinion information, along with the features of nouns and pronouns, are first proposed in this paper.

The training set for the NTCIR-7 MOAT Task, introduced in Section 4.1, is adopted for extracting training features and for building models. As the size of this set is not large, the co-training method is adopted to improve the performance (Blum & Mitchell, 1998). Co-training is a semi-supervised learning method that trains models together with labeled and un-labeled materials. In co-training, sentences with high CRF confidence scores are selected, and sentences among them without words that are portions of the opinion holder are dropped. These selected sentences, along with their predicted labels, are fed back to the CRF system as the training sentences in the next iteration. The co-training process is shown in Figure 2.

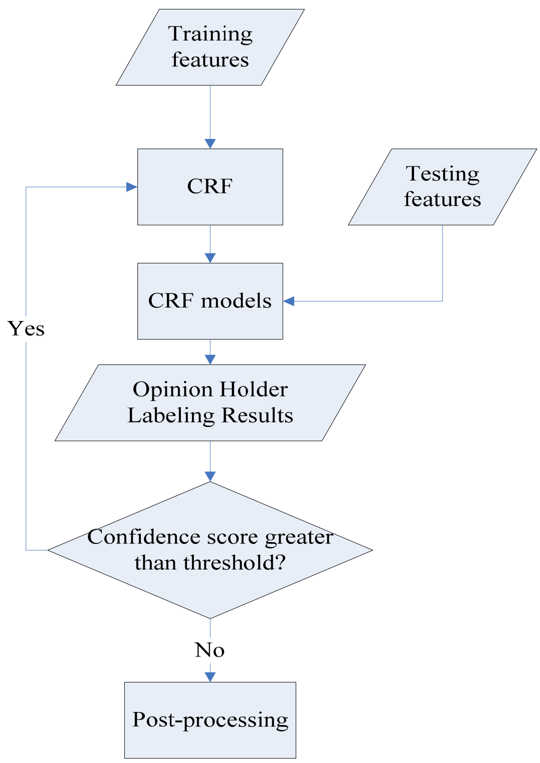


Figure 2. Co-training process

3.4 Post-processing Stage

Post-processing includes two steps: processing phrasal opinion holders and recovering named entities. As mentioned above, opinion holder labeling tells whether the current word is a portion of the opinion holder. Nevertheless, opinion holders often are composed of multiple words, and we need to combine these words to propose the final result if the holder is longer than one word.

3.4.1 Processing Phrasal Opinion Holders

In our experiments for processing phrasal opinion holders, five labels are used to label words in CRF opinion holder labeling: H (head word), I (middle word), T (tail word), S (single word), and O (not opinion portion), abbreviated as HITSO hereafter. Instead, when working with CHAID, the decision tree, it can only generate two labels "YES" and "NO" to tell whether the current word is a portion of the opinion holder. In this phase, comparing the performances of using different label sets is not the focus, so we just use the label set HITSO. The effect of different labels will be tested later.

Different post-processing rules are applied according to the tagging sets. If words are labeled by the tagging set H, I, T, S, and O by CRF, we use the following rules to combine words to a phrase if necessary:

(1) Find the H labeled word with the highest confidence score in the current sentence.

(2) Combine the sequent I labeled words until a T labeled word is found.

(3) The final opinion holder includes words that begin from the word with label H found in (1) to the word with label T found in (2).

(4) If all words in the current sentence are all labeled "O", the opinion holder will be set to the author.

If words are labeled with the tagging set {YES, NO} by CHAID, we use the following three rules to combine words to a phrase if necessary:

(1) Combine consecutive nouns. For example, "印度 (India, Nc) 總統 (president, Na) 瓦希德 (Abdurrahman Wahid, Nb)" are combined into one opinion holder phrase.

(2) Use conjunctions and "、" (a mark in Chinese punctuations to list items in a series) to combine nouns into one holder group. For example, "瓦希德 (Abdurrahman Wahid, Nb) 、 (PAUSECATEGORY) 柯林頓 (Clinton, Nb) 與 (and, Caa) 小淵惠三 (Keizo Obuchi, Nb)" are combined into one opinion holder group. If people express the same opinion together, they are usually grouped as an opinion holder group in sentences.

(3) If all words in the current sentence are labeled "NO," the opinion holder will be set to the author.

3.4.2 Recovering Named Entities

Translated named entities tend to be segmented incorrectly in text pre-processing. These segmentation errors will make opinion holders incomplete, so recovering wrongly segmented named entities is necessary. In this paper, we postulate that the number of occurrences of a complete named entity should be the same as its partial sequences. Therefore, we will compare the occurrences of the current holder sequence with the current holder sequence plus its previous/following word to decide whether we should combine them to generate a more complete opinion holder. For example, in the sentence “Indonesian big man Suharto ruled Indonesia for 32 years,” the name “Suharto” is translated into “蘇哈托”. Nevertheless, this name is wrongly segmented into two words ”蘇哈” and “托”. In this case, we will check whether the number of appearance of “蘇哈托” is the same as ”蘇哈”. If it is, we will combine ”蘇哈” and “托” into one word “蘇哈托”. This process is done iteratively until the numbers of appearance of the current word and that plus the previous/following word are not the same. In this example, we further check the next word “統”. In doing so, we will find that the numbers of appearance of “蘇哈托” and “蘇哈托統” are not the same. Therefore, the recovery process stops and we propose “蘇哈托” as the opinion holder. This process is shown as follows.

For each sentence $S = \{w_1, w_{i+1,} ..., w_n\}$, w is a word in S
where the identified opinion holder $OP = \{w_i, w_{i+1,} ..., w_j\}, i \geq 1, j \leq n$:

while $i > 1$

if the number of occurrence of *OP* equals that of the string $\{w_{i-1}, OP\}$, i--;

while $j < n$

if the number of occurrence of *OP* equals that of the string $\{OP, w_{j+1}\}$, $j++$;

3.5 Result Generation Stage

The final step is result generation, which considers the results of the author’s opinion recognition and opinion holder labeling. Author’s opinion recognition classifies all opinion sentences into author’s opinions and non-author’s opinions, while opinion holder labeling labels the opinion holder. We propose two result generation strategies, and their flowcharts are shown in Figures 3 and 4.

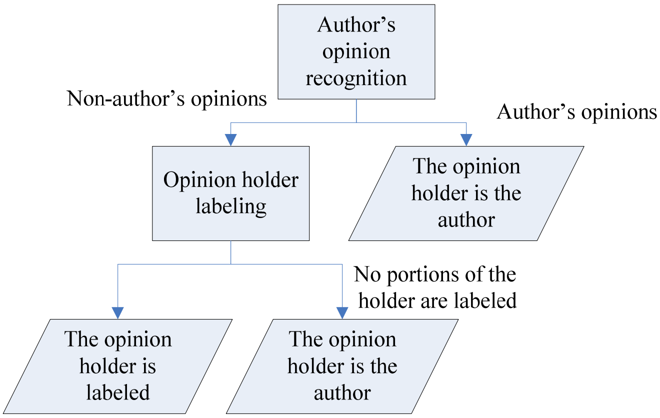


Figure 3. Result generation strategy A

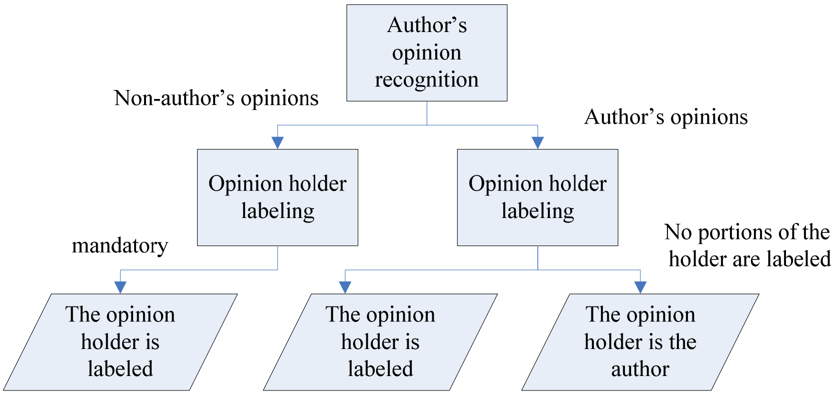


Figure 4. Result generation strategy B

In Result generation strategy A, we have more confidence in the author's opinions recognized by the author's opinion recognition module. Non-author's opinions are passed to the opinion holder labeling module. If there is an opinion holder labeled by this module, we propose it; otherwise, we propose the author as the opinion holder. In Result generation strategy B, we have more confidence in the non-author's opinions recognized by the author's opinion recognition module. Both kinds of opinions are then passed to the opinion holder labeling module. For the non-author's opinions, we force the opinion holder labeling module to propose an opinion holder by considering the most possible opinion holder among words in the current sentence. For the author's opinions, if there is an opinion holder labeled by this module, we propose it; otherwise, we propose the author as the opinion holder. After the text pre-processing, author's opinion recognition, opinion holder labeling, post-processing, and the result generation, the opinion holder is determined.

4. Experiment and Discussion

In this section, the experimental corpus and resources are introduced. Results of author's opinion recognition, opinion holder labeling, and the complete opinion holder identification are shown and discussed.

4.1 Corpus and Evaluation Tasks

The adopted experimental corpus is the NTCIR-6 Pilot Task & NTCIR-7 MOAT Task Traditional Chinese corpora. NTCIR[2] is one of the three important international evaluative forums. MOAT (Multilingual Opinion Analysis Task) is one of its evaluative tasks (Seki *et al.* 2008). The MOAT task provides English, Japanese, Traditional Chinese, and Simplified Chinese materials, which include news articles collected from 1998 to 2001. Labels for relevance, the opinion sentence, the opinion polarity, and the opinion holder are provided by the NTCIR-6 Pilot Task. In addition to these labels, labels for the opinion target are provided later by the NTCIR-7 MOAT Task.

Each MOAT corpus is classified into training and testing sets. The original purpose of the corpus released prior to the formal run (the training corpus mentioned here) is to provide samples to the participants, so its quantity is comparably small. The NTCIR-7 MOAT training set includes documents of 3 topics, consisting of 1,509 sentences, with 944 of them being opinion sentences; the NTCIR-7 MOAT testing set includes documents of 14 topics, consisting of 4,665 sentences, with 2,174 of them being opinion sentences. The opinion labels are sentence-based. As the size of the NTCIR-7 MOAT training set is small, the NTCIR-6 Pilot Task testing set is added for training in this paper. This set includes documents of 29 topics, consisting of 9,240 sentences, with 5,453 of them being opinion sentences. Labels for opinion sentences and opinion holders in these training and testing sets are utilized for opinion holder identification in this paper.

4.2 Experimental Resources

The opinion dictionary and named entity dictionaries are adopted in this paper. The opinion dictionary includes opinion operators, positive, neutral, and negative opinion words extracted from the NTCIR7 MOAT training set, and NTUSD (Ku & Chen, 2007). It is utilized for feature extraction. The person name, location name, and organization name dictionaries are collected for named entity recognition here, including the Million person name dictionary, Sinica corpus[3], CNA translated name dictionary, The Revised Chinese Dictionary[4], Japanese

[2] http://research.nii.ac.jp/ntcir/

[3] http://dbo.sinica.edu.tw/ftms-bin/kiwi1/mkiwi.sh?language=1

[4] http://dict.revised.moe.edu.tw/

common family name dictionary, MOE translated location name dictionary, translated foreign location name list, and Taiwan national industry list.

4.3 Author's Opinion Recognition

In this experiment, the NTCIR-6 testing set and the NTCIR-7 training set were utilized for training, while the NTCIR-7 testing set was used for testing. NTCIR generates the gold standard under two metrics: the strict metric and the lenient metric. We selected opinion sentences by the lenient metric as the gold standard for testing, *i.e.* for these sentences, at least two out of three annotators label them as opinions. Precision, recall, f-score, and accuracy were adopted for evaluation.

The experimental corpus was annotated by three annotators. Therefore, there are inconsistent annotations of opinion holders in some opinion sentences. For example, in the sentence "Because Taiwan understands, using nuclear weapons will destroy the relationship with the U.S." one annotator labeled "Taiwan" as the holder, while the other selected "the author". The example sentence is an implicit nested opinion, and inconsistent annotations are found often in nested opinions. We checked all of the sentences to see if there are many inconsistently labeled opinion holders. The results are shown in Figure 5.

For the NTCIR-6 corpus, if the annotators could find any opinion holder, it was reported; if they could not, the opinion holder was automatically set to the author. Therefore, we cannot know how many opinion holders are inconsistently labeled (see the "?%" in Figure 5). Instead, for the NTCIR-7 corpus, the opinion holder "the author" was explicitly annotated. Therefore, we are able to find the percentage of inconsistently labeled opinion holders. Figure 5 shows that the opinion holder in 19% of sentences is consistently labeled as the author, while that in 15% of sentences is inconsistently labeled as the author and the other named entity. These two percentages are close to each other.

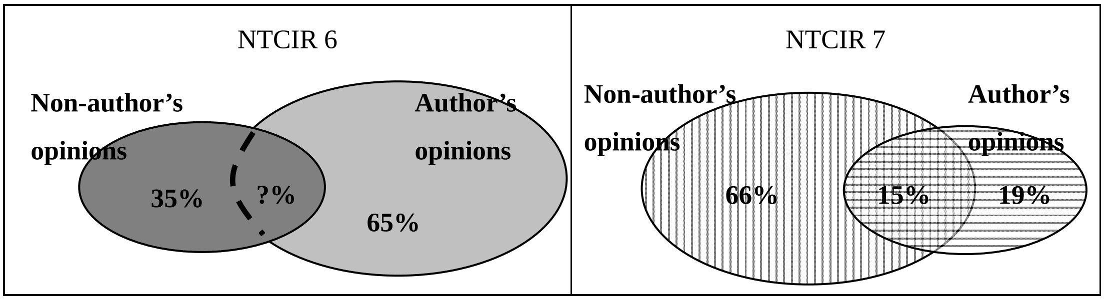


Figure 5. The percentages of the author's opinions in NTCIR-6 and NTCIR-7 corpora

From experiments, we have found that using opinion sentences as the training materials performs better than using all sentences and using both NTCIR-6 (all) and NTCIR-7 (testing) sentences for training also performs better than only using one of them. Therefore, in the experiments for author's opinion recognition, opinion sentences of NTCIR-6 and NTICR-7 were both used for training.

We have discussed the inconsistency in the gold standard for the author's opinion recognition. Therefore, three settings are tested: treating inconsistently labeled opinions as the author's opinions, treating inconsistently labeled opinions as non-author's opinions, and expelling inconsistent labeled opinions in the training set. Table 3 shows the testing performances of these three settings. From Table 3, we find that the first setting, treating inconsistently labeled opinions as the author's opinions, performs the best. It achieves an f-score of 79.98%, which outperforms the setting of treating them as non-author's opinions. The f-score, 65.0%, is even worse when the inconsistently labeled data is not considered, compared to the former two settings. Therefore, we conclude that inconsistently labeled data is useful and we should treat them as the author's opinions.

Table 3. Performances of three settings for inconsistently labeled author's opinions

Settings	Precision	Recall	f-score	Accuracy
Author's opinions	69.68%	93.85%	***79.98%***	83.49%
Non-author's opinions	64.87%	95.94%	77.40%	80.31%
Expelling inconsistency	50.52%	91.53%	65.10%	77.28%

4.4 Opinion Holder Labeling

For opinion holder labeling, the NTCIR-7 training set is utilized. Here, we compare the performances of using strict opinion sentences (agreed on by three annotators) and lenient opinion sentences for training. After labeling words in sentences, the results of author's opinion recognition and opinion holder labeling are considered together to generate the proposed opinion holder. In this experiment, only opinion sentences correctly proposed by the system are evaluated, so the real performance of the opinion holder labeling can be calculated without the propagation errors from opinion extraction. As the number of sentences is the same as the number of opinion holders, the precision, recall, and f-score are not used as evaluation metrics because they will be equal. Instead, the number of correct holders and wrong holders, and the set f-score is adopted. The formula for calculating the set f-score is shown below.

$$set-f-score=\frac{system\_correct(holder)}{system\_correct(opinion=yes)} \quad (1)$$

As mentioned in the previous section, CHAID and CRF are both tested for their performances in opinion holder labeling. Their performances are shown in Table 4.

Table 4. Performances of opinion holder labeling using CHAID and CRF

	Method	Correct #	Error #	Set f-score
Strict	CHAID	564	605	48.16%
	CRF	818	351	69.89%
	CHAID+CRF	825	344	70.57%
Lenient	CHAID	981	967	50.31%
	CRF	1317	631	67.57%
	CHAID+CRF	1322	626	67.83%

Table 4 shows that CRF performs much better than CHAID in opinion holder labeling for both strict and lenient opinion sentences. In the setting CHAID+CRF, we first use CHAID to get the predicted labels, and use these labels together with other features as the input features for CRF. Results show that this setting can slightly improve the performance. One reason could be that the tagging set for CRF is larger than CHAID. Another reason could be that CRF has better performance in combining words with labels, while CHAID needs to apply rules on the proposed labels to find the complete opinion holder. The best performances achieved are a set f-score of 70.57% for strict opinion sentences and 67.83% for lenient opinion sentences.

Furthermore, we classify labeling errors into six types according to the position of the proposed opinion holder and the correct opinion holder:

(1) The proposed opinion holder is not related to the correct holder in the aspect of position. 29.1% errors are of this type.

(2) The proposed opinion holder has one additional word in the front or rear, compared to the correct opinion holder. For example, “魯斯曼日前” (Russman the other day, proposed) and “魯斯曼” (Russman, correct). 18.1% errors are of this type.

(3) The system only proposes the title of the opinion holder, but not the name. For example, “科索伏著名塞裔領袖” (The famous Serbian leader of Kosovo, proposed) and “科索伏著名塞裔領袖特拉伊科維契” (The famous Serbian leader of Kosovo Trajkovic, correct). 8.3% errors are of this type.

(4) The system only proposes the modifier of the correct opinion holder. For example, “該” (That, proposed) and “該裁決” (That decide, correct). 7.5% errors are of this type.

(5) The proposed opinion holder has two or more additional words compared to the correct opinion holder. For example, “狄蘭在記者會” (Dylan proposed in the press conference) and “狄蘭” (Dylan, correct). 5.5% errors are of this type.

(6) The proposed opinion holder is incomplete. 4.7% errors are of this type.

From these errors, we find that most errors occur because the system cannot determine the first word and the last word of the opinion holder properly. Therefore, different tagging sets were tested (IO, ISO, HTO, HISO, HITSO, *etc*.), and we have found that using the tagging set {H, I, O} can achieve the best performance, which is the set f-score of 70.57% for strict opinion sentences. We also propose the name entity recovery method to deal with errors of the sixth type. Experiments show that with co-training, the best confidence score threshold 0.7 and the named entity recovery, our system achieves the best set f-score of 72.03%.

4.5 Experiments for Opinion Holder Identification

After author's opinion recognition and opinion holder labeling, we need to propose the opinion holder according to these results. Table 5 shows the performances of applying different result generation strategies.

Table 5. The performance of opinion holder identification: applying different result generation strategies

	Strategy	Correct #	Wrong #	Set f-score
strict	A	829	340	70.92%
	B	858	310	73.40%
lenient	A	1338	611	68.65%
	B	1372	576	70.40%

Table 5 shows that the performance of applying Result generation strategy B is better than applying Result generation strategy A. It indicates that the proposed method for author's opinion recognition works better in determining non-author's opinion. That is, we can be more sure when the system tells that the current opinion is not expressed by the author. Using fewer author relevant features may be the reason for this phenomenon.

4.6 Performances of NTCIR-7 Participants

NTCIR-7 evaluates the system performance in two ways. One is to evaluate sentences correctly proposed by the system, which is also used in the previous section; the other is to evaluate all opinion sentences in the testing set.

Table 6 shows the performances of all participants' systems together with the performance of our system. WIA's system performs the best in both evaluation metrics. WIA adopts heuristic rules to design their systems. Therefore, our system performs the best among all systems using machine learning methods. Moreover, our system performs better for strict opinion sentences, which is different from other systems. In other words, our system is good at identifying the holder of opinions that were consistently annotated by annotators.

Table 6. Performances of NTCIR-7 participants

	Participants	Number of correctly proposed opinion sentences	Evaluate correctly proposed opinion sentences	Evaluate all opinion sentences		
			Set f-Score	Precision	Recall	f-score
Strict	WIA	757	***82.30%***	***19.88%***	***49.52%***	***28.38%***
	iclpku-1	880	57.84%	13.03%	40.53%	19.72%
	iclpku -2	989	58.04%	10.35%	45.70%	16.88%
	TTRD-1	1213	54.91%	8.22%	52.95%	14.23%
	TTRD-2	866	58.31%	9.72%	40.13%	15.65%
	NTU-1	1169	48.16%	8.14%	44.90%	13.78%
	Our System	1169	**73.40%**	**12.38%**	**68.31%**	**20.97%**
Lenient	WIA	1134	***82.54%***	***29.92%***	***43.05%***	***35.31%***
	iclpku-1	1364	58.72%	20.51%	36.84%	26.35%
	iclpku -2	1606	59.90%	17.33%	44.20%	24.90%
	TTRD-1	2070	56.47%	16.78%	40.02%	23.65%
	TTRD-2	1464	59.49%	14.43%	53.73%	22.75%
	NTU-1	1948	50.31%	14.43%	53.73%	22.75%
	Our System	1948	**70.40%**	**19.80%**	**63.11%**	**30.15%**

5. Conclusion and Future Work

We proposed a machine learning based method for opinion holder identification. We classified this task into two subtasks: author's opinion recognition and opinion holder labeling. SVM was adopted for author's opinion recognition, and CRF was adopted for opinion holder labeling. We proposed lexical, syntactic, contextual, and opinion features. Named entities and punctuation marks were also utilized as features. We tested different tagging sets to find the best set {H, I, O}. Co-training was proposed to solve the problem of insufficient training materials, and results merging strategies were proposed to improve the performance. We also mentioned the methods of utilizing inconsistent annotated materials and analyzed system errors to find solutions for improving the performance.

The proposed system for the opinion holder identification achieved an f-score of 0.734, which is the best among machine learning based systems and is close to the state of the art. The state of the art system adopts heuristic rules. Nevertheless, heuristic rule based systems like it are difficult to rebuild because the rules are usually not described in detail in the previous literature.

In the future, we hope to solve the co-reference resolution problem, which is important in named entity extraction and also in opinion holder extraction. In addition, we plan to add parsing information to improve the performance. Finding a good named entity recovery algorithm is also one of our next attempts. In summary, utilizing techniques of opinion holder identification in the opinion analysis system to compare opinions of different people is our future goal.

References

Blum, A., & Mitchell, T. (1998). Combining labeled and unlabeled data with co-training. *Conference on Computational Learning Theory*, 92-100.

Breck, E., Choi, Y., & Cardie, C. (2007). Identifying expressions of opinion in context. *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, 2683-2688.

Chang, C. C., & Lin, C. J. (2001). LIBSVM: a library for support vector machines.

Choi, Y., Cardie, C., Riloff, E., & Patwardhan, S. (2005). Identifying sources of opinions with conditional random fields and extraction patterns. *Proceedings of EMNLP conference*, 355-362.

Kim, S. M., & Hovy, E. (2004). Determining the sentiment of opinions. *Proceedings of the COLING conference*, 1367-1374.

Kim, S. M., & Hovy, E. (2006). Extracting opinions, opinion holders, and topics expressed in online news media text. *Proceedings of the Workshop on Sentiment and Subjectivity in Text at the joint COLING-ACL conference*, 1-8.

Kim, Y., Jung, Y., & Myaeng, S. H. (2007). Identifying opinion holders in opinion text from online newspapers. *International Conference on Granular Computing*, 699-702.

Kim, Y., Kim, S., & Myaeng, S. H. (2008). Extracting topic-related opinions and their targets in NTCIR-7. *Proceedings of the Seventh NTCIR Workshop*, 247-254.

Ku, L. W., & Chen, H. H. (2007). Mining opinions from the web: beyond relevance retrieval. *Journal of American Society for Information Science and Technology*, 1838-1850.

Kudo, T. (2003). CRF++: yet another CRF toolkit. http://crfpp.sourceforge.net/.

Lafferty, J., McCallum, A., & Pereira, F. (2001). Conditional random fields: probabilistic models for segmenting and labeling sequence data. *Proceedings of International Conference on Machine Learning*, 282-289.

Liu, K., & Zhao, J. (2008). NLPR at Multilingual Opinion Analysis Task in NTCIR7. *Proceedings of the Seventh NTCIR Workshop*, 226-231.

Lo, Y. S. (2008). *An Approach of Using Multiple Dictionaries and Conditional Random Field in Chinese Segmentation and Part of Speech Tagging*. Master Thesis, National Taiwan University.

Meng, X., & Wang, H. (2008). Detecting opinionated sentences by extracting context information. *Proceedings of the Seventh NTCIR Workshop*, 268-271.

Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M., & Euler, T. (2006). YALE: rapid prototyping for complex data mining tasks. *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 935-940.

Pang, B., & Lee, L. (2008). *Opinion mining and sentiment analysis*. Foundations and Trends in Information Retrieval, 2(1-2), 1-135.

Seki, Y., Evans, D. K., Ku, L. W., Sun, L., Chen, H. H., & Kando, N. (2008). Overview of multilingual opinion analysis task at NTCIR-7. *Proceedings of the Seventh NTCIR Workshop*, 185-203.

Seki, Y., Kando, N., & Aono, M. (2009). Multilingual opinion holder identification using author and authority viewpoints. *Journal of Information Processing and Management*, 189-199.

Wu, Y. C., Yang, L. W., Shen, J. Y., Chen, L. Y., & Wu, S. T. (2008). Tornado in multilingual opinion analysis: a transductive learning approach for Chinese sentimental polarity recognition. *Proceedings of the Seventh NTCIR Workshop*, 301-306.

Xu, R., & Wong, K. F. (2008). Coarse-Fine opinion mining - WIA in NTCIR-7 MOAT task. *Proceedings of the Seventh NTCIR Workshop*, 307-313.

Computational Linguistics and Chinese Language Processing
Vol. 14, No. 4, December 2009, pp. 403-422 403

以學習者平行語料庫爲本之西班牙語連接詞研究[1]

Parallel Corpus-based Study of Conjunctions

盧慧娟[*]、呂羅雪[+]

Hui-Chuan Lu and Lo Hsueh Lu

摘要

本論文旨在結合語料庫技術建構「西語學習者英西平行語料庫」，並以之爲本、輔以應用工具，進行學習者連接詞使用之分析研究。

「西語學習者英西平行語料庫」現階段所收集的語料來源是靜宜大學西文系和成功大學外文系西語組兩不同族群的學習者在課堂內完成的西文作文及與之對應的英文翻譯。上述語料皆經母語者修正與詞類註記之程序，以達有效率系統化分析之目的。以過去相關研究爲基礎，我們使用語料庫應用工具對比分析連接詞的使用情形，以探討學習者在第三語連接詞不同語意類型的多用、誤用和少用的情形，及其可能的影響因素。研究結果顯示：(1)學習者在西文和英文整體連接詞的使用比率低於在西語及英語母語者修正文中的連接詞使用率，推論連接詞的少用可能和跟學習者的母語有關。(2)錯誤類型少用的情形集中於「聯繫、補充」類型的連接詞；多用的情形則多屬於「反義、因果」類型的連接詞。(3)在可能的影響變數的檢定中，連接詞的使用錯誤與西語程度間，未呈現顯著的關係性，和第二語（英語）之相關性也未達顯著差異的水準。最後，我們根據分析結果，利用現有之工具，提出連接詞的教學範例以供參考。

關鍵字：語料庫爲本、學習者語料庫、平行語料庫、連接詞、第二語、第三語

[1]本論文是由以「以平行學習者語料庫爲本之西班牙語連接詞研究」爲題口頭發表於「當代歐洲、文化暨語言發展新趨勢國際學術研討會」(政治大學於 99 年 5 月 15 日舉辦)並被編印於該研討會會前論文集(93-110 頁)的會議論文所修改而成。

[*]成功大學外文系，台灣台南市大學路 1 號，Tel: (06)275-7575*52227 Fax: (06)238-7730
Department of Foreign Languages & Literature, National Cheng Kung University, Tainan, Taiwan
E-mail: huichuanlu1@gmail.com

[+]靜宜大學西文系，台灣台中縣沙鹿鎮中棲路 200 號，Tel: (04)2632-8001*12303 Fax: (04)2462-1281
Department of Spanish Language and Literature, Providence University, Taichung, Taiwan
E-mail: slo@pu.edu.tw

Abstract

This paper aims to introduce the process used to create a "Parallel Learners' Corpus of Spanish and English" (CPATEI) by integrating the theories and techniques of Corpus Linguistics, and to investigate the conjunctions used by the Spanish learners in Taiwan.

The collected data in CPATEI are contributed by learners of Spanish from two universities in Taiwan, including Spanish compositions and their corresponding English translations written by two different levels of Spanish learners. Besides, we have collected texts revised by native speakers and POS-tagged.

Based on the collected data, we will examine the tendency of learners' overuse, underuse and misuse of the conjunctions. At the same time, we will discuss the possible factors leading to incorrect usage of conjunctions. The results demonstrate that the frequency of conjunctions used by the learners in their Spanish compositions and English translations are lower than that of the Spanish and English native speakers in their corrections. Among the types of errors of conjunctions, underuse of the copulative and additive conjunctions and overuse of the adversative and cause-consecutive conjunctions are the two types of errors that have been noted in the writing of the Taiwanese learners. Furthermore, there are no relationships found between the wrong use of the Spanish conjunctions and their language level as well as the second language as English and the third language as Spanish. Finally, we hope the result of this research will bring benefits to the curriculum designs for the teaching of Spanish conjunctions.

KeyWords: Corpus-based, Learners' Corpus, Parallel Corpus, Conjunctions, Second Language, Third Language

1. 語料庫建構與連接詞分析之目的

在過去幾十年，有關學習者語料庫的研究已引起越來越多的關注，但不可諱言的，其重心仍多集中於英語或單語的建構與開發；而在平行學習者語料庫的建置方面，其中大多數仍聚焦於第二語言(L2)和母語(L1)的平行影響上；以西語學習者爲對象，且擴及第二語對第三語影響因素考量的平行語料庫則未見。有鑑於研究的特殊性與實際需求之缺乏，本研究的目標之一即建構「平行雙語學習者語料庫」(CPATEI)。

本論文的語料庫應用部份，則是以連接詞爲研究議題。連接詞的主要功能在於連接句中的詞與詞、片語與片語、子句與子句或句與句，扮演著聯繫語意、使表達的語氣通順，以幫助讀者理解文本的關鍵角色。語言分析主題的選定主要來自西班牙語教學的實際觀察，我們發現：(1)台灣西語學習者有寫出具連貫性文章的困難。學習者在書寫句子時，特別在連接詞使用方面，常有少用或誤用的情形，造成語意連貫上的不順暢。(2)他

們所書寫的文本內容過於口語化，對正式書寫文體風格的認知較缺乏。(3)目前台灣所使用西班牙語教科書多以溝通式教學法爲主導，內容偏重語言表達之功能，對連接詞的特性、分類及其在不同文體中的使用功能較少涉略，可能因此造成學習者對連接詞的掌握較不足。

本論文旨在結合語料庫技術建構「西語學習者英西平行語料庫」(Corpus Paralelo de Aprendices Taiwaneses de Español e Inglés, CPATEI)，並以之爲本、輔以應用工具，進行學習者連接詞使用之分析研究。以下內容主要涵蓋語料庫和連接詞兩大部分。我們將由語料庫建構出發，簡述相關工具的應用，呈現現階段建構成果及語料後續處理的過程。接著，分析連接詞的使用與錯誤傾向，並進而檢視可能的影響因素，最後參考母語者使用情形，提出教學範例以供參考。

2. 文獻回顧

2.1 學習者語料庫建構

2.1.1 單語學習者語料庫

在國外，僅有「以西語爲第二語之書寫語料庫」(CEDEL2: Corpus Escrito del Español L2) 和「西班牙語學習者口語語料庫」(SPLLOC: Spanish Learner Language Oral Corpus)，且兩者皆以英語爲母語之西語學習者爲語料收集的對象。「以西語爲第二語之書寫語料庫」(CEDEL2) 由西班牙馬德里自治大學(Universidad Autónoma de Madrid)暨格拉納達大學(Universidad de Granada)所建構。語料量約 60 萬字，語料來源主要涵蓋超過 1500 位以英語爲母語之西語學習者和約 500 位西語母語者。目前未開放檢索使用，因此可利用性有限。另一西語學習者語料庫「西班牙語學習者口語語料庫」(SPLLOC) 由英國南安普敦大學(Southampton University)、紐卡索大學(Newcastle University)及約克大學(York University) 三所大學暨經濟社會研究院(Economic and Social Research Council)合作所建構。目前共有 60 位以英語爲母語之西語學習者參與的錄音檔案約 240 筆。此外，也包含西語母語者的語料錄音檔案約 50 筆。其建構計畫包含兩階段：SPLLOC 1 和 SPLLOC 2。目前 SPLLOC 1 已開放提供研究使用，可選擇限定或不限定條件進行關鍵字的檢索。

在國內，針對「語料庫」一詞的定義，如將之詮釋爲「語料彙整」的廣義解釋，則在台灣西語教學界的研究不乏以之爲本的相關論著，如白保羅、林姿如等人皆以各自所彙整的書面語語料進行學習語言分析；在口語語料庫方面，有由 Javier Pérez & Miguel Rubio 所建構的台灣西班牙文口語語料庫(Corpus Oral del Español en Taiwán, COET)，建構於 2004 年，內容包含了 82 個以文藻外語學院西語學習者爲對象所收集到的會話錄音檔案。上述幾個語料庫的語言資料大部分皆已電子化或經過人工註記，比較可惜的是沒有進一步整合成可互換、再利用或結合科技使之成爲可公開檢索的語料共享資源。此外，政治大學建置了以顯示該校外語學院多語種（23 種語種）特色的「外語學習者語料庫」，目前未包含西班牙語。最後，建構中的「成功大學多國語語料庫」，目前包括台灣日語

學習者語料庫和台灣西語學習者語料庫。其中,「台灣西語學習者語料庫」(CATE: Corpus de Aprendices Taiwaneses de Español),可公開檢索的網址爲 http://corpora.flld.ncku.edu.tw/。目前計含 1,913 篇學習者書面文本,總字數量約 34 萬字,文本本身標註有錯誤修正、詞類與字根註記,檢索時可設定條件、進行字(串)、該字詞類、字根及下一字詞類的檢索。

2.1.2 平行學習者語料庫

在平行的學習者語料庫方面,第一、二語的平行學習者語料庫極爲有限,如日-英語(Mark, 2001)、中-日(Shimizu, Du & Dantsuji, 2004)等,而我們所著手建構的「台灣西語學習者英西平行語料庫」(CPATEI)則具有西漢語言組合以及第二和第三學習語平行的特色。

2.2 連接詞研究

從寫作表達的角度出發,如同閱讀理論中所強調的閱讀過程,一般而言,無論是說明體,還是敘事體,文章的基礎都是從迷你結構開始,也就是說,從字、句開始,逐步發展到子結構,再形成段落,最後朝向頂尖大結構(Lu & Lu, 2005),而具有連貫字句功能的連接詞在整體架構則扮演著特殊的角色。

2.2.1 連接詞分類

根據孫義禎(1993)、陳樹升(1993)、劉啓分(1993)、Montolío (2001)與 Gómez (1997)等對連接詞的屬性分類,整合出以下主要類別及所屬連接詞範例。首先,根據語法屬性,分(1)對等並列、(2)附屬和(3)相關三大類。(1)對等並列連接詞涵蓋以下類型:(a)聯繫連接詞“y/e, ni”。(b)選擇連接詞“o/u”。(c)對稱連接詞“ahora…ahora, bien…bien, ya…ya, ora…ora”。(d)補充連接詞“además”。(e)反義連接詞“pero, mas, sin embargo”。(f)解釋連接詞“es decir, o sea”。(2)附屬連接詞包括(a)一般連接詞“que”。(b)原因連接詞“porque, puesto que”。(c)結果連接詞“por eso, por consiguiente, por (lo) tanto”。(d)目的連接詞“para que, a fin de que”。(e)條件連接詞“si, con tal de que”。(f)讓步連接詞“aunque”。(g)比較連接詞“como”。(3)相關連接詞包括“apenas…cuando, ni…ni…, o…o…”。針對現有語料的文本類型,本文僅就學習者文本中所使用的連接詞屬性進行分析與討論。

2.2.2 以母語者語料爲研究對象

有關西班牙語連接詞方面,Martínez (1997), Martín & Portolés (1999), Montolío (1998, 2000, 2001), Sánchez (1993), Lu & Lu (2006)等均探討連接詞在西文文章中的功能。Lu & Lu (2006)根據「西班牙語語料庫」(Corpus del Español) 的查詢結果指出自然語中連接詞出現頻率的差異,如:“pero” (43330)高於“mientras que” (1502)[2]等。另外,Lu *et al.* (2009)

---

[2] 爲 2006 年時的查詢結果,因該語料庫語料之擴充,現查詢所得筆數分別爲 pero (131,191), mientras que (3,993),差距亦隨之擴大。

根據所建構的「西、英、中平行語料庫」(CPEIC)分析連接詞的使用情形。由於平行語料資源的缺乏，這個語料庫的多語平行語料主要來自聖經[3]。分析結果顯示在平行的三語語料中，西文連接詞使用的頻率最高，英文次之，中文最少；而根據連接詞的屬性分類，在各個語言中，並列連接詞的數量皆是最多的。

2.2.3 以學習語為研究對象

2.2.3.1 以英語為學習語

以英語為學習語的連接詞研究在國外如：Granger y Tyson (1996)透過對比分析學習者母語及所學之外語寫作中連接詞的使用情況，以瞭解他們中介語的發展。所分析之連接詞量包括以法語為母語(89,918 字)和以英語為母語者(77,723 字)的英文寫作分別為 916 和 976 個連接詞。研究結果發現：在學習者的英語外語作文中，出現有連接詞使用過多的現象，如：佐證連接詞"indeed, of course, in fact"、舉例連接詞"e.g., for instance, namely"，以及補充連接詞 "moreover"；有些連接詞卻使用不足，如：反義連接詞 "however, though, yet"和結果連接詞 "therefore, thus, then"。Field & Yip (1992)指出補充連接詞多用的現象也出現在香港學生的第二語英語寫作中。上述連接詞使用過多的情形，被推論可能和受母語影響有關。以中文為母語的學習者連接詞使用過多的現象，如要被推論為跟母語有關，則需根據漢英對比或平行對應語料分析的結果，同樣的語意，英文會表達出連接詞的次數應該要少於漢語才算合理。

在台灣，Wu (1996)研究連接詞與英語能力之間的關係，發現連接詞的使用與語言程度的高低有密切關係。他在台灣大學生對英語連接詞的理解與使用之研究中，指出英語能力愈強，使用連接詞的能力也愈強。再者，周玉玲(2002)探討短篇小說運用於英語連接詞教學之情形，研究以事實性文章作為測試，結果發現學生對時間性連接詞掌握較低，其次是補充和反義，對因果連接詞掌握得最好。余昇易(2006)和曾嬿竹(2004)探討連接詞教學與大學生寫作表現上的成效及影響，結果顯示連接詞的密集教學能有效改善語言程度能力較低學生之整體寫作表現。

[3] 雖然使用聖經作為平行語料分析的對象一直有相關的爭議性存在，然而不可否認的，它仍是語料庫語言學和自然語言處理平行對應和剖析應用中最常被利用的語言資源，原因不外乎在於它有豐富的多國語言文本、可合法取得的路徑與便利使用的電子化格式。我們使用聖經作為平行語資料，主要是西英漢三語平行電子檔案格式的語料資源不易取得。以其語料庫所建構的「西英漢平行語料庫」(CPEIC)彙整了 BibleGateway.com 網路資源中的三語平行對應文本，目的是為了便利分析被平行翻譯的三個語言資料。此外，其平行文本的版本採用的是西英文的新國際版(New International Version for English, la Nueva Versión Internacional)和中文的聯合版(Union Version for Chinese) ，文本的內容則是聖經的創世紀；其西、英、中語料量分別是 37,407、38,895 和 52,546 字。

2.2.3.2 以西班牙語爲學習語

以西班牙語爲學習語的連接詞研究在國外如：Romero (2005)指出，因爲有連接詞連結句子和段落，才能使文章的語意連貫，這對學生的閱讀理解是有助益的。他把 Montolío (2001)對連接詞之分類應用於研究中，分析了 32 篇大學生的作文（含一、四年級兩組），研究發現大學生對反義連接詞的掌握較好。其中，一年級學生在使用「介紹觀點」之連接詞的表現比四年級學生好。至於補充連接詞方面，四年級學生使用「開啓新資訊」之連接詞優於一年級的學生。此外，Poblete (1999)指出學習者所寫的敘事性文章中，具補充屬性之連接詞的使用頻率較高。

在台灣方面，Blanco (2007)對比分析以西班牙語爲母語的記者所寫之連續事件爲主題的新聞文章 300 篇和 30 位台灣西班牙語學習者的作文，發現同一文章類型中，不同研究對象之相同連接詞使用的頻率不同（如：母語者使用時間連接詞“antes”和“antes de”的頻率分別是 84.94%和 12.04%，而學生在寫作時使用的百分比分別是 64.42%和 34.58%）。研究結果也顯示：學習者在反義連接詞類型有多用的情形，而特定文章類型與特定連接詞之使用間有相關性。而根據我們在「西班牙語語料庫」(Corpus del Español) 的查詢結果顯示“antes”和“antes de”的出現頻率分別是 240417 次(96%)和 10027 次(4%)，如此的結果也說明連接詞在不同文本類型間的使用差異。

2.3 第三語相關理論與研究

Leung (2005, 2006)指出第三語言學習者語言知識的轉換來源有兩種：第一語言和第二語言的中介語文法。此外，De Angelis (2007)延續(2005)的研究指出有關跨語言影響或語言知識轉移之因素更可細分成以下幾個主題，包括語言的距離(兩種語言間的同異程度)、學習者目的語和原生語(或第一語)能力、最後使用該語言的時間距離（即：多久以前使用過該語言）、在非母語環境下居住和處於非母語環境的時間長度、習得語言順序以及語言正式度(語言使用於正式或非正式的情境) 等。在影響第三語或多國語言習得的因素中，Ringbom (2007)也討論到語言距離和語言能力程度兩個相關的因素。

綜合上述所言，我們瞭解到國內外相關語料庫和連接詞方面的開發與研究情形，由此我們發現在有關以西班牙語爲學習語的連接詞研究領域裡，尚缺乏語料庫爲本、以西語爲第三語或平行語料，探討學習者連接詞使用情形方面的研究。因此，本研究將使用語料庫應用工具，對比分析我們所收集之語料中連接詞使用的情形，並探討學習者多用、誤用和少用的第三語（西班牙語）連接詞的類型，及其可能的影響因素。

3. 學習者平行語料之建構與利用

「西語學習者英西平行語料庫」的前期建置工作是由國科會人文學研究中心經費補助之整合型前置規劃案。語料庫前置與現階建構成果有以下幾點差異：(1)過去對於參與的西語程度沒有一致的比較基準，只能以學習時數區分，現階段的參與者則皆統一接受西語程度分級的測試。(2)在前置的建構中，兩個學習者族群完成文本的地點不同，靜宜大學的學生是在課堂中完成，而成功大學是在家中完成；爲了達到比較對照的目的，在 2009

年的收集中，所有的文本皆在課堂中完成，以統一控制變數。(3)在 2006 年的建置中，初期使用人工進行 XML 的格式註記，爲達系統化與提昇便利性，目前則利用多項工具協助註記與分析的工作。「平行雙語學習者語料庫」(CPATEI) 經過前置階段的檢討、改進，現仍處語料庫的建構初期，因其學習者雙語語料在收集與註記工作之特性，語料量之累積速度較慢，待語料累積至一定數量，計畫將會架設可供公開檢索的語料庫介面。

3.1 語料收集

爲達到便利分析學習者連接詞使用的情形，我們先行結合語料庫技術，建構「西語學習者英西平行語料庫」(CPATEI：Corpus Paralelo de Aprendices Taiwaneses de Español e Inglés) 。現階段「西語學習者英西平行語料庫」所收集的語料對象是靜宜大學西文系二年級和成功大學外文系西語組二年級兩族群類型的西語學習者（分別是 33 和 22 人），中文是他們的母語、英文是第一外語，且西班牙文是第二外語。

在 2009 年 3 月執行正式測試前，我們先進行前置測試，由 2 位西語學習者參與，各寫一篇西文作文後翻成英文，西文寫作和英文翻譯的時間各爲 25 分鐘，西文作文題目是“Un día en la vida de Taiwan”，最後寫出測試的感想。我們透過前置測試的結果和參與者的感想瞭解正式測試時所該注意的事項。進行正式測試的前一週，參與者先閱讀研究說明書（研究目的與方式）後填寫同意書，同意他們的西文作文和英文翻譯被收錄於「台灣西班牙語學習者西英平行語料庫」，日後僅供學術研究之用途；同時，並收集參與者相關的個人資料。一週後於課堂時間執行正式測試，對象分別是靜宜大學西文系二年級學生和成功大學外文系西語組二年級學生，於課堂中以“Un día en la vida de Taiwan”爲題依西文作文、英文翻譯的順序，在不可以使用任何工具書的情況下，於 50 分鐘內完成測試。爲有效區分學習者西文程度，每位參與者於測試結束後皆接受威斯康辛西語能力分級測驗(Wisconsin Placement Test)。本研究後續分析所使用的語料來源即是上述兩組學習者在課堂內完成的西文作文（6,185 字）及與之對應的英文翻譯（6,057 字），其來源對象的特質分佈如表一所示。

表一、語料及參與者特質分佈情況

學校系級 (2009 年收集)	人數 (男、女)	學習時數[4]	西語能力檢定分數 (平均數)	語料庫字數 西：6,185；英：6,057
靜宜西文二	33 (1, 32)	1024	371~588 (445)	西：3,721；英：3,520
成功外文二	22 (9, 13)	160	355~548 (491)	西：2,464；英：2,537

[4] 學習時數的計算方式爲：α (每週西文上課時數) x 16(週) x 2(學期數) x β (學習年數)。

3.2 語料後續處理

3.2.1 錯誤修正

語料的後續處理包括：學習者語料之西文作文和英文翻譯分別經由西語母語者修改學生西文作文[5]，以及由英語母語者修改學生英文翻譯之程序完成修正文的語料彙整。該語料庫因此涵蓋學習者西文作文原始文、西語母語者西文修正文、學習者英文翻譯原始文和英語母語者英語修正文共四類語料。

3.2.2 可利用之學習者語料分析工具

爲達後續有效率系統化分析之目的，本研究所利用之語料庫工具和與之配合的數個小程式分述如下。

(1) Michael O'Donnell 所開發的「語料庫工具」(UAM Corpus Tool)：可便利外籍母語者針對學習者錯誤進行修正、註記者分類註記，結果輸出、檢索與統計的語言分析工具。

(2) Helmut Schmid 的「詞類註記器」(Tree Tagger of University of Stuttgart)：可依西文、英文不同語言別類，自動標註文本之詞類與字根。

(3) Mike Scott 的「WordSmith」詞彙分析工具(WordSmith Tool, 第五版)：可進行詞表高低頻排序、兩頻率表關係性檢定和搭配詞檢索。

(4)「西語搭配詞工具」(Spanish Collocation Tool)[6]：對西語語料進行前後鄰近搭配詞的統計檢定，提供搭配詞與其詞類統計及排序結果，並可進一步比較兩語料搭配詞表之差異性。不同於工具(3)所獲取的結果是詞頻表，「西語搭配詞工具」所提供的則是搭配詞表。

(5) Michael Barlow 的「平行檢索器」(ParaConc)：方便對不同語言的平行檢索。

在過去，就算沒有相關工具或小程式可協助或利用的前提下，研究分析的工作仍可以人力手動的方式一一被完成，但善用工具的優勢，則讓我們節省時間與精力，以投注於更多電腦科技尚無法取代的語言分析工作。

4. 連接詞分析應用

4.1 研究問題

在以學習者平行語料庫爲本之語言分析方面，我們根據 1.研究目的所陳述有關連接詞學習探討的重要性與必要性以及 2.相關文獻的論證，我們設定以下的研究問題：

---

[5] 爲避免母語修改者的風格影響分析結果，本論文所分析的西文原始文經過四位母語者(Rita Swade, Cindy Barquero, Jesús Robla 和 Saul Aviles)的修改，不同母語者間的差異度如下：在成功大學的學習者作文修改是 2.94%，而靜宜大學則是 4.17%，皆在 5%的區間範圍內。

[6] 由成功大學資工所王廷軒助理開發。

(1) 學習者作文和母語者的修正文中，連接詞使用的分佈情形有何差異？

(2) 連接詞主要的錯誤類型是什麼？

(3) 語言程度和第二語是否對第三語的連接詞錯誤使用有顯著影響？

4.2 研究對象

以所建構「台灣西語學習者英西平行語料庫」中現階段靜宜 33 名和成大 22 名兩組參與者在課堂內所完成的西文作文（6,185 字）、與之對應的英文翻譯（6,057 字）[7]及由母語者所修正共四類型語料進行研究分析。

4.3 研究方法

以語料庫法，充分利用語料庫分析工具的多樣化與便利性，透過詞類與錯誤註記、語料檢索、分類、計算與統計檢定變數間之相關性。與之配合的工具與執行步驟敘述如下。

(1) 透過「語料庫工具」，母語者修正學習者錯誤，註記者將錯誤分類（錯誤使用、多用、少用）的訂定，註記錯誤類型（根據母語者的修改結果分：選擇錯誤的連接詞（含標點符號）、該用連接詞而未用、不該使用連接詞而使用）和語意屬性（轉折、結果、補充等）。分析者再根據錯誤類型的檢索結果計算，並使用 SPSS (15 版)統計檢定變數間（西語能力檢定、英文連接詞錯誤數、西文連接詞錯誤數）的相關性。

(2) 使用「詞類註記器」對四類語料（西文作文和英文翻譯的原始文和修正文）依語言分別進行詞類自動註記，將檢索的層次由字詞本身提昇到詞類的範圍。採用互現訊息統計法，針對靜宜和成大兩校學習者的西文作文中連接詞的前後搭配成分進行分析，並以 KL (Kullback-Leibler divergence)法比較學習者原始作文和母語者修正文的差異。

(3) 採用「WordSmith 詞彙分析工具」協助我們從詞類註記過的西文和英文語料，擷取不同語料中連接詞的使用頻率，原始文和修正文間的相關性，以及連接詞所出現的位置等訊息。

(4) 透過「西語搭配詞工具」統計檢定，進一步瞭解連接詞前後緊鄰搭配成分的詞彙與詞性。

(5) 將處理完的語料載入「平行檢索器」，透過連接詞的詞類檢索，獲取西文所平行的英文句中是否有連接詞對應，並重複由英文檢索，查詢計算西文平行對應的結果。

---

[7] 針對論文審查曾提出有關語料量「少」可能影響語言分析結果之代表性的問題，我們以相對於本論文所分析的學習者程度為對象、透過檢索「台灣西語學習者語料庫」(CATE，其文本類型包含敘述、論說類型，以及藝術、歷史、休閒、健康等九大主題）的語料並逐一進行觀察、對照與檢視本論文所提出的分析結果與論點，藉此增加文本類型的多樣性並彌補現有平行學習者語料量的不足所做的推論結果。其檢索結果如「使用最多者仍是並列連接詞，補充連接詞“además”有少用的情形，轉折及反義連接詞有多用等的情形」和本論文的檢索、分析結果類似。

4.4 結果與討論

4.4.1 學習者原始文和母語者修正文的連接詞使用比較

表二顯示兩族群學習者之原始文和修正文中連接詞之使用率（使用出現次數/總字數）及其差異[8]。

表二、學習者原始文和修正文之連接詞使用情形

連接詞/總字數	靜宜大學	差距	成功大學	差距
西文原始文	5.03%	0.32%	4.63%	0.24%
西文修正文	5.35%		4.87%	
英文原始文	6.51%	0.51%	6.54%	0.04%
英文修正文	7.02%		6.58%	

從表二，透過比較原始文和修正文中連接詞的使用比率差異，我們發現在西文作文(靜宜：5.03%;成大：4.63%)和英文翻譯(靜宜：6.51%;成大：6.54%)的原始文中連接詞的使用比率皆低於在西語(靜宜：5.35%;成大：4.87%)及英語(靜宜：7.02%;成大：6.58%)母語者修正文中的連接詞使用率，也就是學習者低於母語修改者的使用率。此外，我們也觀察到：在成大西文(0.24%)的使用率差異比英文大(0.04%)，而在靜宜的使用率差異則是英文(0.51%)比西文大(0.32%)。整體而言，兩族群的學習者在連接詞的使用率都還要再提高，才能更接近母語者的使用情形。此外，主修的不同也反應在不同語言連接詞使用率的差異上。亦即：外文系學習者西文連接詞使用率差異大於英文；反之，西文系的學習者英文連接詞使用之差異大於西文連接詞。外文系的在英文原始文和修正文中連接詞使用率的差異小於西文系的表現，第二語（英語）的程度差異可以合理地解釋此一結果。而另一方面，西文系和外文系在西文原始文和修正文中的連接詞使用率的差異雖然不大，但西文系在西文連接詞使用率的差異表現上有些微高於外文系學習者的連接詞使用。我們透過進一步觀察學習者所使用的連接詞發現[9]：連接詞使用的傾向可能是導致此

[8] 由於我們所比較的兩個族群的作文篇數不同，故採以使用率（連接詞數/總字數）的計算方式較爲客觀公平。如和其它相關文獻在探討連接詞的使用率（以所有連接詞爲分母，去計算不同類型或個別的連接詞的使用比例）做比較，我們上述的計算方式所得出的平均值會顯得比較小。

[9] 學習者連接詞使用情形如下表所示。

連接詞類型(連接詞/總字數)	靜宜大學	成功大學
y （聯繫）	2.95%	3.08%
o （選擇）	0.86%	0.73%
cuando （時間）	0.24%	0
porque （因果）	0.27%	0.12%

一結果的因素之一，因爲被視爲較易使用的西文並列連接詞(y/o)，在靜宜西文系的學習者方面，所使用的比率佔所有連接詞的 75.94%，低於成大西文系學習者的使用率(82.46%)。亦即：學習時數較少的學習者（成功大學學生）在較簡單的聯繫和選擇連接詞上使用得比較頻繁。相對的，在其它連接詞使用的傾向，如反義連接詞(如：“pero/sin embargo”「但是」)或是附屬類型中的因果連接詞(如“porque”「因爲」)等，靜宜西文系的學習者所使用的比率則高於成大西文組學習者的使用率。並列的聯繫連接詞是較易使用的連接詞類型，成大的學生使用得比較多，正確率相對容易提高；而靜宜在較難使用的連接詞上使用得比較多，也容易出現錯誤，所以在與母語者修改的結果比較後，成大學生在原始文和修正文的差距反而略微小於靜宜學生的表現。此外，我們可以推論出連接詞學習的發展階層爲並列的聯繫連接詞“y/o”早於反義連接詞“pero/sin embargo”。

接著，我們進一步去觀察並列連接詞使用在第二語和第三語間的細部對應關係，我們以平行句的對應關係去瞭解在學習者原始文裡，西文對應到英文時連接詞於兩個語言出現的比例數，並與母語修改者的修正文裡西譯英連接詞使用比例做比較，結果如表三所示。

表三、西譯英並列連接詞使用之比較

學校	靜宜大學		成功大學	
文本	原始文	修正文	原始文	修正文
西譯英連接詞使用	1.75:1	2.01:1	1.40:1	1:53:1

從表三，我們可以觀察到西文對應至英文的並列連接詞在學習者原始文和母語者修正文的差異，其結果驗證成功大學外文系西語學習者(1.40:1.53)在連接詞使用的表現確實比靜宜大學西文系學習者(1.75:2.01)接近母語者修正文的西譯英比例。

有關西文學習時間長度較長的學習者和母語者的使用率差異高過西文學習時間較短的學習者的情形，在後續 4.4.3 章節中，我們將特別針對西文能力程度（而非西文學習時間長度）進行統計檢定驗證語言程度與連接詞使用正誤的相關性。

進一步，我們分析學習者所使用的連接詞特性時發現：學習者在連接詞的使用多集中於“y(e)/and”和“o(u)/or”並列連接詞的類型，當我們限定第一字之詞類爲連接詞，且同時兼顧互現訊息分數和結果筆數皆高者的範例如“y luego, y yo, y me”。以右側被搭配詞爲附屬性連接詞的“luego” (“y luego”)爲例，表四前兩列是兩組學習者在原始文和修正文

sin embargo （反義）	0.05%	0
pero （反義）	0.21%	0.12%
si （條件）	0.11%	0
aunque （讓步）	0.03%	0

中"y luego"兩詞搭配出現的統計結果。我們進一步查詢"y luego"一詞在「西班牙語語料庫」中的出現情形（表四第四列）[10]。

表四、"y luego"搭配詞於原始文和修正文出現的情形

y luego	搭配次數	MI 值	"y"出現個數	"luego"出現個數
原始文	3	3.76	188	6
修正文	5	83.56	202	6
西班牙語語料庫	4,239	6.6	1,249,549	19,283

從表四，我們可以觀察到：連接詞與其後被搭配成分"luego"共同出現之機率在原始文的搭配次數是 3，而在修正文是 5，"y luego"這個搭配詞在修正文的互現訊息值(83.56)高過原始文(3.76)，"luego"此一附屬性的連接詞在母語者的修正文中幾乎只要出現就跟"y"做搭配組合，但此一被搭配成份在學習者的原始文本中則僅有 50%的機率是與連接詞"y"共同出現，我們進一步查詢到"y luego"在「西班牙語語料庫」中的出現個數(4,239)與互見訊息值(6.6)。結果顯示其互見訊息值雖不如修正文高，但仍約爲原始文的兩倍，類似這樣的連接詞搭配範例，可以作爲教學過程中提醒學習者特別注意的重點之一。

4.4.2 錯誤類型

學習者在西文作文和英文翻譯的錯誤率（連接詞錯誤數/連接詞使用數）結果如表五所示。

表五、錯誤率分佈情況

語言	總錯誤率	誤用	多用	少用
西文	18.60%	4.32%	4.65%	10.30%
英文	10.63%	2.53%	1.72%	5.82%

由表五，我們觀察到學習者在西文連接詞使用的錯誤率高於英文。此外，我們也注意到根據誤用、多用和少用的不同錯誤類型分類：在學習者西文作文中，錯誤率所佔比例由高到低分別是少用、多用和誤用；而在其英文翻譯中，則依序爲少用、誤用和多用。因此，不管在學習者的西文作文或英文翻譯的文本中，少用的情形都是三類型中錯誤率最高者。如果再細分三類錯誤的連接詞屬性，我們特別注意到誤用和少用的情形多集中於補充類型屬性的連接詞；而多用的情形則偏屬於反義類型的連接詞。上述的分析觀察和 Granger y Tyson (1996)的研究發現剛好相反，因爲他們指出以英語爲外語之學習者的英語作文中，反義屬性的連接詞有使用不足的情形，而補充屬性的連接詞則有使用過多的現象。此外，Granger y Tyson (1996)和 Field & Yip (1992)指出在以非英語爲母語的學

[10] 感謝論文審查者建議加入「西班牙語語料庫」的檢索結果（表四第四列），輔助修正文與原始文之對比，更能凸顯該連接詞搭配的重要性。

習者在其第二語的英語寫作中，皆出現連接詞使用過多的情形，他們推論可能和受母語影響有關。上述研究結果的主要差異在於目的語不同，Granger y Tyson 和 Field & Yip 的研究對象是以學英文爲外語的學習者，而我們的研究對象則是以中文爲母語、英文爲第一外語、西文爲第二外語的台灣西語學習者。而 Blanco (2007)和我們一樣以台灣的西語學習者爲對象所收集語料的分析，其獲致的結果與我們有類似之處：學習者在反義連接詞有多用的現象。以上的比較結果顯示不同目的語間存在著連接詞使用的差異現象。

再者，針對母語這個影響因素的探討，如就連接詞全面性加總的結果而言，我們發現台灣的西語學習者傾向少用連接詞，連接詞整體性的使用傾向主要是由於聯繫連接詞在所有連接詞中佔有極高的比例。我們根據 Lu *et al*. (2009) 在平行語料中連接詞使用排序（西文>英文>中文）的研究結果(如表六第四列連接詞的總和與第一列之聯繫及補充之並列連接詞檢索結果所示)可以解釋台灣西語學習者連接詞少用的現象，即：在相對平行的語意中，中文少用連接詞的情形影響學習者第二外語西語的學習和使用結果。

表六、「西英漢平行語料庫」之檢索結果

	西文(37,407)	英文(38,895)	中文(52,546)
聯繫、補充	2,768	2,431	328
反義、因果	386	567	129
總和	3,154	2,998	557

如進一步對連接詞語意屬性進行較爲細項分類，我們發現不同類型連接詞間的差異：台灣的西語學習者在聯繫和補充的並列連接詞中有少用的情形，而反義和因果連接詞則有多用的情形。一方面，針對不同類型的連接詞比較，台灣西語學習者在並列和補充連接詞少用的情形可以被這兩類連接詞在平行對應語料中(如表六第二列所示)所呈現的分佈傾向排序（西文>英文>中文）所解釋：因爲對等語意中，中文傾向不用並列的聯繫或補充連接詞，所以台灣西語學習者在進行西文寫作時，受到母語影響的結果會導致少用的情形。另一方面，對於反義和因果的多用情形，我們則以中西文的對比分析來解釋：因爲中文「雖然...可是/但是」和「因爲...所以」的表達，在西文的使用上，屬反義和因果的連接詞成分“aunque, sin embargo, no obstante, pero”以及“como, porque, por eso, por lo tanto”只能擇一出現，我們根據 Anderson (1990, 1993)「學習原則」中的「一對一原則」解釋學習者在反義和因果連接詞多用的結果。此外，針對反義和因果兩類連接詞在平行語料中出現頻率的計算，我們獲致「英文>西文>中文」之使用頻率的高低排序(如表六第三列所示)，這個階層排序顯示西文和英文間的使用情形與前述的補充並列類型有異，我們因此推論台灣西語學習者的第一外語英文和第二外語西文間可能存有某種程度的互動關係，導致影響了學習者的使用結果，這部分則有待後續進一步的研究與探討。

4.4.3 語言程度和第二語與第三語的關係

西語能力、英文翻譯和西文作文的連接詞錯誤數之間的相關性檢定結果如表七所示。

表七、相關性檢定結果

	西語能力		英文錯誤數	
學校	靜宜大學	成功大學	靜宜大學	成功大學
西文錯誤數	0.570	0.220	0.265	0.488

透過表七，我們可以瞭解到西語能力檢定、英文連接詞使用錯誤數和西文連接詞使用錯誤數彼此之間的相關性。SPSS 的檢定結果顯示：三變數間的相關性皆未達顯著水準(P>0.05)者：西語程度和英語連接詞的錯誤使用數(P=0.002)皆分別與西語連接詞使用的錯誤數的多寡無顯著關係存在；亦即西語連接詞的使用錯誤與西語程度無顯著的關係性(靜宜 P=0.570；成大 P=0.220)，且和第二語（英語）之相關性亦未達顯著差異的水準(靜宜 P=0.265；成大 P=0.488)。這和 Wu (1996)研究學習者之英語連接詞使用與學習者第二語英語能力之間有關係的結果不同，我們發現西語連接詞的使用錯誤和第三語西語語言程度的高低沒有相關性存在。此外，我們的檢驗結果也顯示第二語和第三語連接詞錯誤使用彼此間亦無必然之關係。

5. 教學因應之道

余昇易(2006)和曾嬿竹(2004)在探討連接詞教學對學生寫作表現成效的影響的研究，其結果顯示連接詞的密集教學能有效改善語言程度較低學生之整體寫作表現。根據上述對學習者錯誤的分析和討論結果，我們提出以下的教學因應之道。而我們所參考的是依據母語者的「西班牙語語料庫」(Corpus del Español[11]，四千一百萬字)、「現代美語語料庫」(Corpus of Contemporary American，COCA[12]，兩千萬字)和「現代漢語平衡語料庫」(Academia Sinica Balanced Corpus of Modern Chinese[13]，八百萬字)的檢索結果，按照連接詞語意分類所做的排序。這三個極具代表性的母語語料庫的共同特色是語料量豐富且檢索功能強，特別值得一提的是前二者，因建置、註記與程式系統模式類似，對西英雙語的研究非常有幫助。我們根據學習者連接詞使用的分析研究結果，在教學因應策略上針對相關連接詞透過上述三個母語語料庫進行檢索、分類與排序，以作爲教學內容設計的參考，而這裡僅舉出補充和反義連接詞類型的常用連接詞作爲範例，結果如表八所示[14]。

[11] Corpus del Español: http://www.corpusdelespanol.org/x.asp

[12] http://www.americancorpus.org/

[13] http://dbo.sinica.edu.tw/SinicaCorpus/

[14] 西文的數據是在「西班牙語語料庫」(Corpus del Español)設定 19、20 世紀爲查詢範圍的檢索結果。

表八、連接詞於西、英和中文三語言之頻率高低排序

連接詞	西文	英文	中文
補充 連接詞	además (15,376) a la vez (2,361) asimismo (2,184) por otra parte (2,118) a su vez (1,902) por otro lado (962) por un lado (731) por una parte (517)	on the other hand (17,986) besides (16,747) moreover (16,166) on one hand (10,444) furthermore (10,351) on the one hand (4,159)	而且 (2,637) 並且 (995) 此外 (994) 一方面 (478) 還有 (338) 另一方面 (338) 除此之外 (113)
反義/ 讓步 連接詞	pero (131,191) aunque (28,397) sin embargo (14,566) mas (13,740) a pesar de (6,364) sino que (5,690) no obstante (3,588) empero (634)	but (1,811,191) though (164,532) however (142,356) yet (135,705) although (107,965) rather (100,144) even though (35,150) even if (33,802) nevertheless (13,204)	但是 (4,953) 雖然 (2,976) 可是 (2,532) 然而 (1,372) 而是 (932) 儘管 (451) 就算 (321)

表八呈現了台灣西語學習者的第一語中文、第二語英文和第三語西文三個語言的母語者在連接詞使用上高低頻的順位，如：在西文補充屬性的同義連接詞中依序以“además, a la vez”和“asimismo”的出現頻率排序最高，同義的英文連接詞依序爲“besides, moreover”和“furthermore”，中文則是「而且、並且、此外」。而我們也注意到中文表達中的「一方面」和「另一方面」的高低頻排序(前者高於後者)不同於西文的“por otra parte, por otro lado, por un lado, por una parte”和英文的“on the other hand”和“on the one hand”；且西文中的“por otra parte, por otro lado”高於“por un lado, por una parte”。又如西文的反義連接詞高低頻依序爲“pero, sin embargo, no obstante, empero”。教學者若善用上述表格中連接詞排序之跨語言的同異特質，根據聽、說、讀、寫不同需求的語言技能設計教案，預期學習者應可從中獲益。

接下來，即是由真實的文本中學習。這一個程序分兩方面進行，先是透過「語料庫工具」直接檢索出由西語母語者所修改的修正文，以便與學習者原始作文進行對比，分「多、少和誤用」類別以及根據「讓步、結果和補充」語意分類提醒學習者注意(如句子(1)a-b, (2)a-b 所示)。接著再間接搭配「Google 翻譯器」(Google Translator)[15]的英、中翻

[15] http://translate.google.com/#

譯(如句子(1)c-d, (2)c-d 所示)，讓學習者理解其母語（中文）和第一外語（英語）的語意。雖然自動翻譯的結果和人工翻譯仍有相當的改善空間，但不可諱言，「Google 翻譯器」仍是目前自動翻譯工具中成效較佳者。對於因應教學的大量範例，考量效率與便利的因素，我們採用人工修正自動翻譯結果的模式來節省時間與人力。

(1) 補充連接詞少用

a. habitualmente leer en la cama hasta me duermo_luego apaco la luz (學習者西文作文原始文)

b. habitualmente leo en la cama hasta que me duermo y luego apago la luz (母語者修改文)

c. usually I read in bed until I fall asleep and then turn off the light (英文翻譯)

d. 通常我會躺在床上看書，一直到我睡著，然後關燈。(中文翻譯)

(2) 反義連接詞多用

a. Aunque es un poco sosa, pero me gusta la. (學習者西文作文原始文)

b. Aunque es un poco aburrido,_me gusta. (母語者修改文)

c. Although it is a bit boring, I like it. (英文翻譯)

d. 雖然有點無聊，但是我喜歡。(中文翻譯)

最後，則是藉由西語母語者語料庫「西班牙語語語料庫」的關鍵字居間(Keywords in Context: KWIC)功能，透過同一類型的一連串連接詞範例，強化學習者的印象。我們進一步觀察在相關連接詞前後高頻率出現的成分特色，藉此教導學習者學習西語母語者在連接詞使用的慣用模式，以之前所討論過的搭配詞“y luego”爲例，搜尋結果如範例((3)a-c)的句子所示。

(3)

a. Cuba a los platillos volantes, de los que se habla durante mucho tiempo **y luego** pasan años sin que se vuelva a decir nada de ellos.[16]

b. y mi concentración no rinden al máximo. Si yo dirigiera una sola vez **y luego** nada, otra vez y luego nada, entonces no conseguiría buenos resultados.[17]

c. máquinas comerciales cada vez más. En primer lugar porque es muy práctico, **y luego** porque nosotros no somos una fábrica de ordenadores.[18]

---

16 Title: Entrevista (ABC). Author: Corral Pedro. Source: http://www.abc.es.

17 Title: Entrevista (ABC). Author: Rubio José Luis. Source: http://www.abc.es.

18 Title: Entrevista (ABC). Author: Rubio José Luis. Source: http://www.abc.es.

根據上述範例，我們發現主要包含“...(,)y luego....”和“Y luego...”兩種模式。如逐字解釋，“y”是連接詞，“luego”表時間「之後」之語意，當兩詞搭配時，則具有前後語意連接、口氣延順，以及進一步說明之意。舉例來說：3(a)和(b)句中的“ y luego”屬於連接前後語意和口氣；而 3(c)句中的“y luego”則有進一步說明原因之意。

6. 結論

本論文結合語料庫技術建構「西語學習者英西平行語料庫」(西文作文 6,185 字及英文翻譯 6,057 字），並以之爲本、輔以應用工具，進行學習者連接詞使用之分析研究。在語料庫建構方面，語料的後續處理包括修正與錯誤的註記。此外，爲達後續有效率系統化分析之目的，本研究也介紹了目前可利用的語料庫工具及與之配合的開發程式。

在以語料庫爲本的應用分析方面，和過去文獻之研究結果不同，我們獲致以下結論：(1)在西文作文和英文翻譯的原始文中，連接詞的使用比率皆低於在西語及英語母語者修正文中的連接詞使用率，學習者在連接詞的使用率還需加強。我們進一步透過母語者平行語料的比較，對連接詞的少用可能和跟學習者的母語（中文）有關提出驗證說明。(2)在錯誤類型上，我們發現少用的情形集中於「補充」類型的連接詞；多用的情形則多屬於「反義」類型的連接詞。(3)在可能的影響變數檢定中，連接詞的使用錯誤與西語程度間，未呈現顯著的相關性，且和第二語（英語）之相關性也未達顯著差異的水準。最後，我們根據分析結果，利用現有之工具，提出連接詞的教學範例以供參考。

本論文的限制主要和學習者語料的數量有關。受制於雙語平行語料處理本身的複雜度、現有人力與資源分配等外在因素的限制，在語料數量上無法快速擴增，是本研究的一大限制。此外，因爲目前所分析的筆數有限，日後也需要較大量的語料來進一步衍生、驗證與強化本論文與過去研究結果不同的結論，使之更具說服力。在未來工作方面，除了擴增語料庫的語料量外，我們還將從應用工具與分析議題兩方向努力。在應用工具方面，如「西班牙語詞語搭配工具」目前僅限於前後緊鄰成分的搭配關係(bi-gram)，我們會朝理想化的多詞搭配關係(N-gram)邁進。另外，我們將以相同的研究模式與路徑，對不同學習議題進行分析研究，以朝逐步涵蓋第三語之研究廣度與深度的目標努力。

感謝

首先，感謝成功大學頂尖計畫（R021）以及國科會人文學研究中心前置規劃案（「中介語的發展及變異性」整合型 (2007.4～2008.6)計畫之西班牙語子計畫）的經費補助。再者，感謝 UAM 所開發的 Corpus Tool 這個免費工具的資源共享。最後，對協助語料庫建構、作文修改（巴辛蒂、史瓦帝、艾索爾、羅布拉、德瑞克）及語料工具開發和分析的助理群們（王廷軒、仇志良、吳奕賢、張展毓、劉亭佑、張偉婷、葉倩廷、張以姍、林于涵、陳姿均、吳彥融）致上最誠摯的謝意。

參考文獻

余昇易（民 95）。連接詞教學及其對大學生寫作影響之探討。元智大學應用外語學系碩士論文。

周玉玲（民 91）。短篇故事在英語連接詞教學法效益之研究。台灣師範大學英語系在職進修碩士班論文。

孫義禎（民 82）。西班牙語實用語法新編。台北市：建宏出版社。

陳樹升（民 82）。西班牙文文法。台北市：中央圖書出版社。

曾嬿竹（民 93）。線上連接詞教學對大學生英文寫作之效能研究。清華大學外國語文研究所碩士論文。

劉啓分（民 82）。實用西班牙文文法。台北市：中央圖書出版社。

盧慧娟、呂羅雪（民 98）。語料庫為本之中介語研究。淡江外語論叢，13，30-47。

Anderson, R. (1990). Models, processes, principles, and strategies: Second Language acquisition in and out of the classroom. In B. VanPatten and J. Lee (Eds.), *Second Languages Acquisition: Foreign Language Learning* (pp.45-78). Clevedon, UK: Multilingual Matters.

Anderson, R. (1993). Tour operating principles and input distribution as explanations for underdeveloped and mature morphological systems. In Kenneth Hyltenstam and A. Viborg (Eds.), *Progresion and Regresion in Language* (pp.309-339). Cambridge: Cambridge University Press.

Blanco, P. D. (2007). Los conectores en noticias de sucesos, retórica constrastiva en textos de periodistas españoles y alumnos taiwaneses. *Círculo de Lingüística Aplicada a la Comunicación,* 30, 24-44.

Davies, M. (2002-) Corpus del Español (100 million words, 1200s-1900s). Available online at http://www.corpusdelespanol.org.

Davies, M. (2008-) The Corpus of Contemporary American English (COCA): 410+ million words, 1990-present. Available online at http://www.americancorpus.org.

De Angelis, G. (2005). Interlanguage transfer of function words. *Language Learning,* 99(3), 379-414.

__________. (2007). *Third or Additional Language Acquisition*. Clevedon: Multilingual Matters.

Field, Y. & Yip, L. M. O. (1992). A comparison of internal cohesive conjunction in the English essay writing of Cantonese speakers and native speakers of English. *RELC Journal,* 23, 15-28.

Gómez Torrego, L. (1997). *Gramática Didáctica del Español*. Madrid: SM.

Granger, S. & S. Tyson. (1996). Connectors usage in the English essay writing of native and non-native EFL speakers of English. *World Englishes,* 15(1), 17-27.

Leung, Y-K. I. (2005). L2 vs. L3 initial state: A comparative study of the acquisition of French DPs by Vietnamese monolinguals and Cantonese-English bilinguals. *Bilingualism: Language and Cognition,* 8(1), 39-61.

__________. (2006). Full transfer vs. partial transfer in L2 and L3 acquisition. In R. Slabakova, S. Montrul, & P. Prévost (Eds.), *Inquiries in Linguistic Development: In Honor of Lydia White* (157-188). Amsterdam: John Benjamins.

Lu, Hui-Chuan, Cheng-Yu Chang & Lo Hsueh Lu. (2009). Creation and application of CPATEI, a parallel Spanish-English learner corpus with Mutual Relevance Analysis between multiple interlanguages. Abstract accepted by AACL in Alberta, Canada.

Lu, Hui-Chuan & Lu Lo Hsueh. (2005). Estudio sobre la interacción entre las estructuras textuales y oracionales. *Glosas Didácticas,* 15, 145-156.

__________. (2006). Estudio del léxico sobre un corpus de textos de lectura. *TamKang Studies of Foreign Languages and Literatures,* 28, 35-50.

__________. (2009). Creación del Corpus Paralelo de Aprendices Taiwaneses de Español e Inglés (CPATEI). *Foreign Language Studies,* 9, 135-156.

Mark, K. (2001). A parallel learner corpus: Using computers in a humanistic approach to language teaching and research. *Japanese Journal of Language and Society*, 4(1), 5-16.

Martínez, R. (1997). *Conectando Texto. Guía para el Uso Efectivo de Elementos Conectores en Castellano*. Barcelona: Octaedro.

Martín, M. & J. Portolés. (1999). Los marcadores del discurso. In I. Bosque & V. Demonte (Eds.), *Gramática Descriptiva de la Lengua Española* (pp.4051-4213). Madrid: Espasa Calpe.

Montolío, E. (1998). La teoría de la relevancia y el estudio de los marcadores discursivos. In Martín Zorrquino, M. & E Montolio (Eds.), *Los Marcadores del Discurso. Teoría y Análisis* (pp.93-119). Madrid: Arco Libros.

__________. (2000). *Manual Práctico de Escritura Académica*. Vol III. Barcelona: Ariel.

__________. (2001). *Conectores de la Lengua Escrita*. Barcelona: Ariel.

Poblete, M. T. (1999). La cohesión de los marcadores discursivos en distintos tipos de discurso. *Estudios Filológicos*, 34, 165-180.

Romero, M. G. (2005). Análisis de los conectores en ensayos argumentativos escritos por estudiantes universitarios. Ponencia presentada en XXIV ENDIL de Sán Cristóbal, Venezuela.

Ringbom, H. (2007). *Cross-linguistics Similarity in Foreign Language Learning*. Clevedon: Multilingual Matters.

Sánchez, I. (1993). ¿Por qué son tan incoherentes los ensayos que escriben los estudiantes? *Tierra Nueva,* 1, 87-93.

Shimizu, M., Du, F., & Dantsuji, M. (2004). Construction of the Chinese Learners' Parallel Corpus of Japanese and its preliminary analysis. *K@ta,* 6(2), 91-105.

Wu, Jwo-Shien. (1996). The Relationship between conjunction use and language proficiency. *Studies in English Language and Literature,* 1(8), 50-57.

語料庫與工具

Academia Sinica Balanced Corpus of Modern Chinese, http://dbo.sinica.edu.tw/SinicaCorpus/

Google Translator, http://translate.google.com/#

Corpus de Aprendices Taiwaneses de Español (CATE), http://corpora.flld.ncku.edu.tw/

Corpus del Español, Mark Davies, http://www.corpusdelespanol.org/x.asp

Corpus Escrito del Español L2 (CEDEL2), UAM, http://www.uam.es/proyectosinv/woslac/cedel2.htm

Corpus of Contemporary American (COCA), Mark Davies, http://www.americancorpus.org/

ParaConc, Michael Barlow, http://paraconc.com/

Spanish Learner Language Oral Corpus (SPLLOC), http://www.splloc.soton.ac.uk/

Tree Tagger, University of Stuttgart, http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/

UAM Corpus Tool, Michael O'Donnell, http://www.wagsoft.com/CorpusTool/

WordSmith Tool v.5, Mike Scott, http://www.lexically.net/wordsmith/

The individuals listed below are reviewers of this journal during the year of 2009. The IJCLCLP Editorial Board extends its gratitude to these volunteers for their important contributions to this publication, to our association, and to the profession.

Guo-Wei Bian
Chia-Ping Chen
Chao-Jan Chen
Hao-Jan Howard Chen
Zhao-Ming Gao
Iwen Huang
Chih-Hsien Huang
Jeih-weih Hung
Un-Gian Iunn
Gwo-Hwa Ju
Chia-Feng Juang
June-Jei Kuo
Yuan-Fu Liao
Chi-Jen Lin
Yu-Hong Lin
Bor-Shen Lin
Chuan-Jie Lin
Hao-Chiang Koong Lin
Shu-Yen Lin
Hsien-Chin Liou
Chris Shei
Wei-Ho Tsai
Richard Tzong-Han Tsai
Shih-Ping Wang
Jenq-Haur Wang
Wing-Kwong Wong
Jiun-Shiung Wu
Jie Chi Yang
Jui-Feng Yeh
Liang-Chih Yu

2009 Index
International Journal of Computational Linguistics & Chinese Language Processing
Vol. 14

This index covers all technical items---papers, correspondence, reviews, etc.---that appeared in this periodical during 2009.

The Author Index contains the primary entry for each item, listed under the first author's name. The primary entry includes the coauthors' names, the title of paper or other item, and its location, specified by the publication volume, number, and inclusive pages. The Subject Index contains entries describing the item under all appropriate subject headings, plus the first author's name, the publication volume, number, and inclusive pages.

AUTHOR INDEX

SUBJECT INDEX

The Association for Computational Linguistics and Chinese Language Processing

(new members are welcomed)

Aims：

1. To conduct research in computational linguistics.
2. To promote the utilization and development of computational linguistics.
3. To encourage research in and development of the field of Chinese computational linguistics both domestically and internationally.
4. To maintain contact with international groups who have similar goals and to cultivate academic exchange.

Activities：

1. Holding the Republic of China Computational Linguistics Conference (ROCLING) annually.
2. Facilitating and promoting academic research, seminars, training, discussions, comparative evaluations and other activities related to computational linguistics.
3. Collecting information and materials on recent developments in the field of computational linguistics, domestically and internationally.
4. Publishing pertinent journals, proceedings and newsletters.
5. Setting of the Chinese-language technical terminology and symbols related to computational linguistics.
6. Maintaining contact with international computational linguistics academic organizations.
7. Dealing with various other matters related to the development of computational linguistics.

To Register：

Please send application to:

The Association for Computational Linguistics and Chinese Language Processing
Institute of Information Science, Academia Sinica
128, Sec. 2, Academy Rd., Nankang, Taipei 11529, Taiwan, R.O.C.

payment： Credit cards(please fill in the order form), cheque, or money orders.

Annual Fees：

regular/overseas member： NT$ 1,000 (US$50.-)
group membership： NT$20,000 (US$1,000.-)
life member：ten times the annual fee for regular/ group/ overseas members

Contact：

Address： The Association for Computational Linguistics and Chinese Language Processing
Institute of Information Science, Academia Sinica
128, Sec. 2, Academy Rd., Nankang, Taipei 11529, Taiwan, R.O.C.

Tel.：886-2-2788-3799 ext. 1502 Fax：886-2-2788-1638

E-mail: aclclp@hp.iis.sinica.edu.tw Web Site: http://www.aclclp.org.tw

Please address all correspondence to Miss Qi Huang, or Miss Abby Ho

The Association for Computational Linguistics and Chinese Language Processing

Membership Application Form

Member ID#：________________

Name：__________________________________ Date of Birth：_______________

Country of Residence：___________________ Province/State：___________________

Passport No.：___________________________ Sex: _________________________

Education(highest degree obtained)：____________________________________

Work Experience：____

____

Present Occupation：____

Address：____

____

Email Add：____

Tel. No：____________________________ Fax No：___________________________

Membership Category：☐ Regular Member ☐ Life Member

Date：_____/_____/_____（Y-M-D）

Applicant's Signature：

Remarks： Please indicated clearly in which membership category you wish to register, according to the following scale of annual membership dues：

Regular Member ： US$ 50.- （NT$ 1,000）

Life Member ： US$500.-（NT$10,000）

Please feel free to make copies of this application for others to use.

Committee Assessment：

中華民國計算語言學學會

宗旨：

（一） 從事計算語言學之研究
（二） 推行計算語言學之應用與發展
（三） 促進國內外中文計算語言學之研究與發展
（四） 聯繫國際有關組織並推動學術交流

活動項目：

（一）定期舉辦中華民國計算語言學學術會議（Rocling）
（二）舉行有關計算語言學之學術研究講習、訓練、討論、觀摩等活動項目
（三）收集國內外有關計算語言學知識之圖書及最新發展之資料
（四）發行有關之學術刊物，論文集及通訊
（五）研定有關計算語言學專用名稱術語及符號
（六）與國際計算語言學學術機構聯繫交流
（七）其他有關計算語言發展事項

報名方式：

1. 入會申請書：請至本會網頁下載入會申請表，填妥後郵寄或E-mail至本會
2. 繳交會費：劃撥：帳號：19166251，戶名：中華民國計算語言學學會
   信用卡：請至本會網頁下載信用卡付款單

年費：

終身會員： 10,000.- （US$ 500.-）
個人會員： 1,000.- （US$ 50.-）
學生會員： 500.- （限國內學生）
團體會員： 20,000.- （US$ 1,000.-）

連絡處：

地址：台北市115南港區研究院路二段128號 中研院資訊所(轉)
電話：(02) 2788-3799 ext.1502 傳真：(02) 2788-1638
E-mail：aclclp@hp.iis.sinica.edu.tw 網址: http://www.aclclp.org.tw
連絡人：黃琪 小姐、何婉如 小姐

中 華 民 國 計 算 語 言 學 學 會
個 人 會 員 入 會 申 請 書

會員類別	☐終身 ☐個人 ☐學生	會員編號		（由本會填寫）	
姓　　名		性別		出生日期	年　月　日
				身分證號碼	
現　　職		學　　歷			
通訊地址	□□□				
戶籍地址	□□□				
電　　話		E-Mail			
申請人:　　　　（簽章） 中　華　民　國　　年　　月　　日					

審查結果:

1. 年費：

   終身會員： 10,000.-
   個人會員： 1,000.-
   學生會員： 500.-（限國內學生）
   團體會員： 20,000.-

2. 連絡處：

   地址：台北市南港區研究院路二段128號 中研院資訊所(轉)
   電話：(02) 2788-3799 ext.1502 傳真：(02) 2788-1638
   E-mail：aclclp@hp.iis.sinica.edu.tw 網址: http://www.aclclp.org.tw
   連絡人：黃琪 小姐、何婉如 小姐

3. 本表可自行影印

The Association for Computational Linguistics and Chinese Language Processing (ACLCLP)

PAYMENT FORM

Name :____(Please print)   Date: ___________________

Please debit my credit card as follows: US$______________________

❑ VISA CARD ❑ MASTER CARD ❑ JCB CARD   Issue Bank:_________________________

Card No.: __________- __________ - __________ - __________ Exp. Date:_______________

3-digit code: ______________ (on the back card, inside the signature area, the last three digits)

CARD HOLDER SIGNATURE : ____________________________________

Tel.: ___________________________________ E-mail: ___________________________________

Add: ____

PAYMENT FOR

US$ _____ ❑ Computational Linguistics & Chinese Languages Processing (CLCLP)

Quantity Wanted: ______

US$ _____ ❑ Publications:___________________________________

US$ _____ ❑ Text Corpora: __________________________________

US$ _____ ❑ Speech Corpora:_________________________________

US$ _____ ❑ Others: _______________________________________

US$ __________ ❑Life Member Fee ❑ New Member ❑Renew

US$ __________ = Total

Fax : 886-2-2788-1638 or Mail this form to :

ACLCLP

℅ Institute of Information Science, Academia Sinica

R502, 128, Sec.2, Academia Rd., Nankang, Taipei 115, Taiwan

E-mail: aclclp@hp.iis.sinica.edu.tw

Website: http://www.aclclp.org.tw

中 華 民 國 計 算 語 言 學 學 會
信用卡付款單

姓名: ________________________________(請以正楷書寫)　日期:：________________

卡別：❑ VISA CARD ❑ MASTER CARD ❑ JCB CARD　發卡銀行：________________

卡號:__________-_________-_________- ________　有效日期：________________

卡片後三碼：____________________（卡片背面簽名欄上數字後三碼）

持卡人簽名：　________________________________(簽名方式請與信用卡背面相同)

通訊地址：____

聯絡電話：______________________　E-mail：________________________________

備註：為順利取得信用卡授權，請提供與發卡銀行相同之聯絡資料。

付款內容及金額：

NT$__________❑ 中文計算語言學期刊(IJCLCLP)

NT$__________❑ 中研院詞庫小組技術報告

NT$__________❑ 中文（新聞）語料庫

NT$__________❑ 平衡語料庫

NT$__________❑ 中文詞庫八萬目

NT$__________❑ 中文句結構樹資料庫

NT$__________❑ 平衡語料庫詞集及詞頻統計

NT$__________❑ 中英雙語詞網

NT$__________❑ 中英雙語知識庫

NT$__________❑ 語音資料庫______________________________

NT$__________❑ 會員年費　❑續會　❑新會員　❑終身會員

NT$__________❑ 其他: ________________________________

NT$______________=　合計

填妥後請傳真至 02-27881638 或郵寄至:
115台北市南港區研究院路2段128號中研院資訊所(轉)中華民國計算語言學學會 收
E-mail: aclclp@hp.iis.sinica.edu.tw
Website: http://www.aclclp.org.tw

Publications of the Association for Computational Linguistics and Chinese Language Processing

		Surface	AIR (US&EURP)	AIR (ASIA)	VOLUME	AMOUNT
1.	no.92-01, no. 92-04(合訂本) ICG 中的論旨角色與 A Conceptual Structure for Parsing Mandarin -- Its Frame and General Applications--	US$ 9	US$ 19	US$15	_____	_____
2.	no.92-02 V-N 複合名詞討論篇 & 92-03 V-R 複合動詞討論篇	12	21	17	_____	_____
3.	no.93-01 新聞語料庫字頻統計表	8	13	11	_____	_____
4.	no.93-02 新聞語料庫詞頻統計表	18	30	24	_____	_____
5.	no.93-03 新聞常用動詞詞頻與分類	10	15	13	_____	_____
6.	no.93-05 中文詞類分析	10	15	13	_____	_____
7.	no.93-06 現代漢語中的法相詞	5	10	8	_____	_____
8.	no.94-01 中文書面語頻率詞典(新聞語料詞頻統計)	18	30	24	_____	_____
9.	no.94-02 古漢語字頻表	11	16	14	_____	_____
10.	no.95-01 注音檢索現代漢語字頻表	8	13	10	_____	_____
11.	no.95-02/98-04 中央研究院平衡語料庫的內容與說明	3	8	6	_____	_____
12.	no.95-03 訊息為本的格位語法與其剖析方法	3	8	6	_____	_____
13.	no.96-01 「搜」文解字－中文詞界研究與資訊用分詞標準	8	13	11	_____	_____
14.	no.97-01 古漢語詞頻表(甲)	19	31	25	_____	_____
15.	no.97-02 論語詞頻表	9	14	12	_____	_____
16.	no.98-01 詞頻詞典	18	30	26	_____	_____
17.	no.98-02 Accumulated Word Frequency in CKIP Corpus	15	25	21	_____	_____
18.	no.98-03 自然語言處理及計算語言學相關術語中英對譯表	4	9	7	_____	_____
19.	no.02-01 現代漢語口語對話語料庫標註系統說明	8	13	11	_____	_____
20.	Computational Linguistics & Chinese Languages Processing (One year) (Back issues of *IJCLCLP*: US$ 20 per copy)	---	100	100	_____	_____
21.	Readings in Chinese Language Processing	25	25	21	_____	_____
				TOTAL	_____	_____

10% member discount: ___________Total Due:__________

- **OVERSEAS USE ONLY**
- PAYMENT： ☐ Credit Card (Preferred)
  ☐ Money Order or Check payable to "The Association for Computation Linguistics and Chinese Language Processing " or "中華民國計算語言學學會"
- E-mail：aclclp@hp.iis.sinica.edu.tw

Name (please print): ____________________________ Signature: ____________________________

Fax: ____ E-mail: ______________________________

Address：____

中華民國計算語言學學會
相關出版品價格表及訂購單

編號	書目	會　員	非會員	冊數	金額
1.	no.92-01, no. 92-04 (合訂本)　ICG 中的論旨角色 與 A conceptual Structure for Parsing Mandarin--its Frame and General Applications--	NT$ 80	NT$ 100	_____	_____
2.	no.92-02, no. 92-03 (合訂本) V-N 複合名詞討論篇 與V-R 複合動詞討論篇	120	150	_____	_____
3.	no.93-01　新聞語料庫字頻統計表	120	130	_____	_____
4.	no.93-02　新聞語料庫詞頻統計表	360	400	_____	_____
5.	no.93-03　新聞常用動詞詞頻與分類	180	200	_____	_____
6.	no.93-05　中文詞類分析	185	205	_____	_____
7.	no.93-06　現代漢語中的法相詞	40	50	_____	_____
8.	no.94-01　中文書面語頻率詞典（新聞語料詞頻統計）	380	450	_____	_____
9.	no.94-02　古漢語字頻表	180	200	_____	_____
10.	no.95-01　注音檢索現代漢語字頻表	75	85	_____	_____
11.	no.95-02/98-04　中央研究院平衡語料庫的內容與說明	75	85	_____	_____
12.	no.95-03　訊息為本的格位語法與其剖析方法	75	80	_____	_____
13.	no.96-01　「搜」文解字－中文詞界研究與資訊用分詞標準	110	120	_____	_____
14.	no.97-01　古漢語詞頻表（甲）	400	450	_____	_____
15.	no.97-02　論語詞頻表	90	100	_____	_____
16	no.98-01　詞頻詞典	395	440	_____	_____
17.	no.98-02　Accumulated Word Frequency in CKIP Corpus	340	380	_____	_____
18.	no.98-03 自然語言處理及計算語言學相關術語中英對譯表	90	100	_____	_____
19.	no.02-01 現代漢語口語對話語料庫標註系統說明	75	85	_____	_____
20	論文集 COLING 2002 紙本	100	200	_____	_____
21.	論文集 COLING 2002 光碟片	300	400	_____	_____
22.	論文集 COLING 2002 Workshop 光碟片	300	400	_____	_____
23.	論文集 ISCSLP 2002 光碟片	300	400	_____	_____
24.	交談系統暨語境分析研討會講義（中華民國計算語言學學會1997第四季學術活動）	130	150	_____	_____
25.	中文計算語言學期刊（一年四期）　年份：______（過期期刊每本售價500元）	---	2,500	_____	_____
26.	Readings of Chinese Language Processing	675	675	_____	_____
27.	剖析策略與機器翻譯 1990	150	165	_____	_____
			合　計	_____	_____

※　此價格表僅限國內 (台灣地區) 使用

劃撥帳戶：中華民國計算語言學學會　　劃撥帳號：19166251

聯絡電話：(02) 2788-3799 轉1502

聯絡人：黃琪 小姐、何婉如 小姐　　E-mail:aclclp@hp.iis.sinica.edu.tw

訂購者：______________________　收據抬頭：______________________

地　　址：____

電　　話：______________________　E-mail: ______________________

Information for Authors

International Journal of Computational Linguistics and Chinese Language Processing (IJCLCLP) invites submission of original research papers in the area of computational linguistics and speech/text processing of natural language. All papers must be written in English or Chinese. Manuscripts submitted must be previously unpublished and cannot be under consideration elsewhere. Submissions should report significant new research results in computational linguistics, speech and language processing or new system implementation involving significant theoretical and/or technological innovation. The submitted papers are divided into the categories of regular papers, short paper, and survey papers. Regular papers are expected to explore a research topic in full details. Short papers can focus on a smaller research issue. And survey papers should cover emerging research trends and have a tutorial or review nature of sufficiently large interest to the Journal audience. There is no strict length limitation on the regular and survey papers. But it is suggested that the manuscript should not exceed 40 double-spaced A4 pages. In contrast, short papers are restricted to no more than 20 double-spaced A4 pages. All contributions will be anonymously reviewed by at least two reviewers.

Copyright：It is the author's responsibility to obtain written permission from both author and publisher to reproduce material which has appeared in another publication. Copies of this permission must also be enclosed with the manuscript. It is the policy of the CLCLP society to own the copyright to all its publications in order to facilitate the appropriate reuse and sharing of their academic content. A signed copy of the IJCLCLP copyright form, which transfers copyright from the authors (or their employers, if they hold the copyright) to the CLCLP society, will be required before the manuscript can be accepted for publication. The papers published by IJCLCLP will be also accessed online via the IJCLCLP official website and the contracted electronic database services.

Style for Manuscripts: The paper should conform to the following instructions.
1. Typescript: Manuscript should be typed double-spaced on standard A4 (or letter-size) white paper using size of 11 points or larger.
2. Title and Author: The first page of the manuscript should consist of the title, the authors' names and institutional affiliations, the abstract, and the corresponding author's address, telephone and fax numbers, and e-mail address. The title of the paper should use normal capitalization. Capitalize only the first words and such other words as the orthography of the language requires beginning with a capital letter. The author's name should appear below the title.
3. Abstracts and keywords: An informative abstract of not more than 250 words, together with 4 to 6 keywords is required. The abstract should not only indicate the scope of the paper but should also summarize the author's conclusions.
4. Headings: Headings for sections should be numbered in Arabic numerals (i.e. 1.,2....) and start form the left-hand margin. Headings for subsections should also be numbered in Arabic numerals (i.e. 1.1. 1.2...).
5. Footnotes: The footnote reference number should be kept to a minimum and indicated in the text with superscript numbers. Footnotes may appear at the end of manuscript
6. Equations and Mathematical Formulas: All equations and mathematical formulas should be typewritten or written clearly in ink. Equations should be numbered serially on the right-hand side by Arabic numerals in parentheses.
7. References: All the citations and references should follow the APA format. The basic form for a reference looks like

    Authora, A. A., Authorb, B. B., & Authorc, C. C. (Year). Title of article. *Title of Periodical, volume number*(issue number), pages.

Here shows an example.

    Scruton, R. (1996). The eclipse of listening. *The New Criterion, 15*(30), 5-13.

The basic form for a citation looks like (Authora, Authorb, and Authorc, Year). Here shows an example. (Scruton, 1996).
Please visit the following websites for details.
(1) APA Formatting and Style Guide (http://owl.english.purdue.edu/owl/resource/560/01/)
(2) APA Stytle (http://www.apastyle.org/)

No page charges are levied on authors or their institutions.

Final Manuscripts Submission: If a manuscript is accepted for publication, the author will be asked to supply final manuscript in MS Word or PDF files to clp@hp.iis.sinica.edu.tw

Online Submission: http://www.aclclp.org.tw/journal/submit.php

Please visit the IJCLCLP Web page at http://www.aclclp.org.tw/journal/index.php

International Journal of
Computational Linguistics & Chinese Language Processing
vol.14 no.4 December 2009

Contents

Papers

ISSN: 1027-376X
The Association for Computational Linguistics and Chinese Language Processing