

國立政治大學語言學研究所碩士論文

National Chengchi University

Graduate Institute of Linguistics

Master's Thesis

指導教授：陳正賢 博士

萬依萍 博士

Advisors: Dr. Alvin Cheng-Hsien Chen

Dr. I-Ping Wan

以量化語料庫方法研究中文“導致”的三個近義詞在不同主題下之語義韻

A study of semantic prosody of three near-synonyms of *cause* in Mandarin Chinese under
different topics:

A quantitative corpus-based perspective

研究生：巫孟宸 撰

Student: Meng-Chen Wu

中華民國一百零八年一月

January 2019

A study of semantic prosody of three near-synonyms of *cause* in Mandarin Chinese under
different topics:

A quantitative corpus-based perspective



January 2019



Copyright © 2019

Meng-Chen Wu

All Rights Reserved

Acknowledgements

致謝

終於到了撰寫此論文的最後階段，此刻我的內心卻意外地平靜，不過我明確地知道這平靜並不是無緣無故而來，反而是奠定在這一路上來出現在我學術旅途的老師、同學、朋友與家人們所給予我的幫助與溫暖。

首先，我要深深感謝我的兩位指導教授，陳正賢老師與萬依萍老師。若沒有兩位老師的教導與耐心的陪伴，這篇論文是不可能完成的。感謝陳老師願意跨校當我的指導教授，並且給予我當研究助理的寶貴機會。從中，我不但能盡情發揮我對編程的熱情，也能時常與您討論與分享計算語言學的理論面以及其未來實際應用的可能性。同時，非常感謝您在量化語料庫語言學的領域中不斷地給予我最正確及最新的知識，也在我撰寫論文的過程中，以超乎常理的耐心與細心，導引我如何確切地修改文體架構與其中的邏輯闡述。謝謝萬老師，老師除了指導撰寫論文外，也不斷地給予我人工智慧與語言學結合的應用資訊，使我對自然語言處理以及計算語言學有更多的憧憬。謝謝老師提供許多業界的珍貴一手消息，讓我更加了解相關領域在國內的狀況。最後，也謝謝老師給予我精神上的支持，因此，我在撰寫論文的過程中，能持續地奔跑不放棄。

我要感謝兩位本論文的口試委員，戴智偉與許展嘉老師。感謝兩位老師在繁忙的學期間撥空參與我的論文口試，並給予許多有建設性、明確的建議，點出了論文中的盲點，讓我能更加提升論文的整體級別。另外，我要感謝曾在政大語言所及師大英語系語言組指導過我的教授，黃瓊之老師、何萬順老師、蕭宇超老師、戴智偉老師、甯俐馨老師。你們的教導扎深了我的語言學專業知識。另外，我要特別感謝助教惠玲學姊，您總是不厭其煩地提醒了我諸多行政上面重要的事項，讓我能順利地走完研究所的道路。

我要感謝政大 105 級的同學，吳璐、佩璇、淑丰、迎婕，以及師大的同學，仲恩、Vicky、筱芸、又萱。感謝課堂中有你們，讓我的研究生生涯充滿了色彩。

我要感謝最愛我的家人，爸爸、媽媽、妹妹。感謝你們對我的關愛與照顧，以及精神上的支持。謝謝你們永遠都站在我這邊，我愛你們。感謝教會的弟兄姊妹們，不停地為我禱告，陪伴我屬靈上的成長。感謝 Steve、

Allison、Yawen、Celine、Tina、Naomi、Gloria、Mark、Guru、Philip、Mitzuris、Cristina、Ginny、Jenny、David、Grace、Patrick、Deb、Max、Cate、Jamison、Maisie、Fei、Carmen、Angel、Yena。

最後，我要將一切榮耀歸於最愛我的上帝，我感謝撰寫論文過程中所發生的一切，不管是喜還是悲，我都知道它們是滋潤我生命的養分，感謝主。



Table of Contents

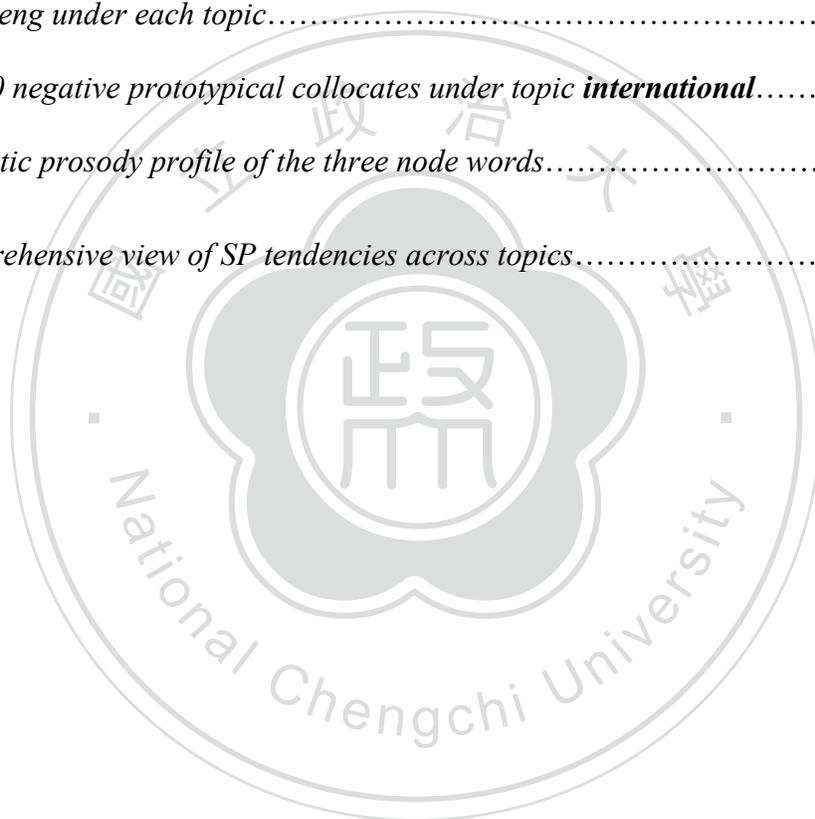
Abstract.....	12
1 Introduction.....	13
2 Literature Review.....	18
2.1 Semantic Prosody.....	18
2.2 Context-dependent Semantic Prosody	20
2.3 Semantic Prosody of CAUSE.....	21
2.3.1 Significant Collocate Analysis.....	21
2.3.2 Concordance Line Analysis.....	22
2.4 Semantic Network Analysis.....	24
3 Methodology.....	27
3.1 Data Collection and Preprocessing.....	27
3.1.1 Data Source.....	27
3.1.2 Data Preprocessing.....	28
3.2 Automatic Concordance Line Analysis	31
3.2.1 ANTUSD	31
3.2.2 Judgement of Evaluative Meanings of Concordances and the SP of a Node Word	34
3.3 Semantic Network Analysis.....	37
3.3.1 Attracted Collocates Extraction.....	37
3.3.2 Semantic Network Construction.....	41

	7
4 Results.....	41
4.1 Semantic Prosody of Chansheng	45
4.1.1 Concordance Semantic Prosody Analysis	45
4.1.2 Network Analysis.....	48
4.2 Semantic Prosody of Niangcheng.....	63
4.2.1 Concordance Semantic Prosody Analysis	63
4.2.2 Network Analysis.....	65
4.3 Semantic Prosody of <i>Cucheng</i>	71
4.3.1 Concordance Semantic Prosody Analysis	71
4.3.2 Network Analysis.....	73
4.4 Internal Summary.....	74
5. Discussion.....	77
5.1 Topics Triggering Positive SP	79
5.2 Topics Triggering Negative SP.....	86
5.3 Topics Triggering Positive and Negative SP	94
6. Conclusion	98
References.....	101
Appendix.....	106

List of Tables

Table 1	<i>Distribution of article number under each of the nine topics in ADN corpus</i>	29
Table 2	<i>Number of concordances of each node word across different topics</i>	31
Table 3	<i>Entry of milian ‘obsessed’ in the ANTUSD</i>	33
Table 4	<i>Entry of wugu ‘innocent’ in the ANTUSD</i>	33
Table 5	<i>Frequency distribution of the max value (>0) in the evaluation fields of each word in the ANTUSD</i>	34
Table 6	<i>Cohen’s kappa statistics on interrater agreement in the three node words</i>	37
Table 7	<i>An example of four types of frequencies necessary for calculating the Delta P value between the node word chansheng and cuojui ‘illusion’</i>	38
Table 8	<i>Number of collocates of chansheng across each topic before/after the filtering</i>	39
Table 9	<i>Number of collocates of niangcheng across each topic before/after filtering</i>	40
Table 10	<i>Number of collocates of cucheng across each topic before/after filtering</i>	40
Table 11	<i>Standardized Residuals in a Chi-Square Contingency Table for the SP distribution of the node word chansheng under each topic</i>	47
Table 12	<i>Negative prototypical collocates under topic society</i>	57
Table 13	<i>Top 20 positive prototypical collocates under topic entertainment</i>	58
Table 14	<i>Top 20 negative prototypical collocates under topic international</i>	59
Table 15	<i>Top 20 positive prototypical collocates under topic sports</i>	60
Table 16	<i>Positive prototypical collocates under topic finance</i>	61
Table 17	<i>Top 20 negative prototypical collocates under topic finance</i>	62

Table 18 <i>Standardized Residuals in a Chi-Square Contingency Table for the SP distribution of the node word niangcheng under each topic</i>	64
Table 19 <i>Top 20 negative prototypical collocates under topic international</i>	69
Table 20 <i>Positive prototypical collocates under topic sports</i>	70
Table 21 <i>Positive prototypical collocates under topic lifestyle</i>	70
Table 22 <i>Standardized Residuals in a Chi-Square Contingency Table for SP distribution the node word cucheng under each topic</i>	72
Table 23 <i>Top 10 negative prototypical collocates under topic international</i>	74
Table 24 <i>Semantic prosody profile of the three node words</i>	75
Table 25 <i>Comprehensive view of SP tendencies across topics</i>	76



List of Figures

<i>Figure 1.</i> The flowchart of method of procedure	27
<i>Figure 2.</i> Distribution of the evaluative meanings of the collocates of the node word <i>chansheng</i> in the ADN.....	46
<i>Figure 3.</i> Subset semantic network under topic society with negative collocates highlighted	49
<i>Figure 4.</i> Subset semantic network under topic entertainment with positive collocates highlighted	50
<i>Figure 5.</i> Subset semantic network under topic international with negative collocates highlighted	51
<i>Figure 6.</i> Subset semantic network under topic sports with positive collocates highlighted	52
<i>Figure 7.</i> Subset semantic network under topic finance with positive collocates highlighted	53
<i>Figure 8.</i> Subset semantic network under topic finance with negative collocates highlighted ...	54
<i>Figure 9.</i> Distribution of the evaluative meanings of the concordances of the node word <i>niangcheng</i> under the seven topics	63
<i>Figure 10.</i> Subset semantic network under topic international with negative collocates highlighted	66
<i>Figure 11.</i> Subset semantic network under topic sports with positive collocates highlighted	67
<i>Figure 12.</i> Subset semantic network under topic lifestyle with positive collocates highlighted .	68
<i>Figure 13.</i> Distribution of the evaluative meanings of the collocates of the node word <i>cucheng</i> in the ADN	71
<i>Figure 14.</i> Subset semantic network under topic international with negative collocates highlighted	73

國立政治大學研究所碩士論文提要

研究所別：語言學研究所

論文名稱：以量化語料庫方法研究中文“導致”的三個近義詞在不同主題下之語義韻

指導教授：陳正賢 博士、萬依萍 博士

研究生：巫孟宸

論文提要內容：(共一冊，20344 字，分六章)

本論文之主要目的為探討文本主題如何影響一個詞彙的語義韻 (semantic prosody)。主題在此定義為在新聞類文體中不同類別的文章，其涵蓋文本的範圍小於語域 (register) 之範圍。我們查驗了一個混合語義韻之詞 (產生)，與兩個強語義韻之詞 (釀成、促成) 在蘋果日報中不同主題下的語義韻分布。我們以規則化的詞語所引列方法來決定這三個詞的語義韻分布，並運用語義網路分析來探尋它們的典型語意場 (semantic field)。研究結果指出，主題對產生的語義韻有中等強烈的影響，但對釀成與促成反而影響程度不大，因此建議了詞彙的語義韻之主題依賴。我們的分析結果指出，新聞文章下的某一主題之內容可能是強化正/負語義韻趨勢的來源，同時揭示了主題下某一詞彙的常規用法。

關進詞：語義韻、語義偏好、主題、規則化評估、語義網路分析

Abstract

The objective of this study is to investigate how the semantic prosody (SP) of a lexical item may be mediated by the topic of the texts. Topic is defined as different categories of articles in news genre, covering a smaller scope of texts than register. In particular, we examine the SP distributions of three near synonyms: a mixed-SP node word, i.e., *chansheng*, and two strong-SP node words, i.e., *niangcheng* and *cucheng*, under different topics in the self-collected Apple Daily News corpus. We determine their SP distributions via a rule-based concordance line analysis on the Apple Daily News corpus and utilize semantic network analysis to further discover their prototypical semantic fields. The results indicate that topic has moderately strong effect on *chansheng*, a mixed-SP node word, but weak effect on *niangcheng* and *cucheng*, strong-SP node words, suggesting the topic-dependency of the lexical SP. Our analysis suggests that the subject matters of the news articles under a given topic may be the source that intensifies the positive/negative SP tendency of a node word, and also reveals the conventionalized usage of the node word under the topic.

Keywords: semantic prosody, semantic preference, topic, rule-based evaluation, semantic network analysis

1 Introduction

Semantic prosody (SP) has been a popular topic in corpus linguistics research over the past three decades. Several researchers have studied the SP of a lexical item (i.e., a word or a phrase) based on the evaluative meaning of its concordances or its most typical collocates in corpora and have shed light on how a node item has a connotational meaning (Hunston, 1995; Hunston, 2007; Louw, 1993; Louw, 2000; Partington, 1998; Partington, 2004; Sinclair, 1991; Sinclair, 2004; Stubbs, 1995). For instance, Louw (1993) examined the collocational patterns of *utterly*, *bent on*, and *symptomatic of*, and found that they tended to co-occur with words with negative evaluative meaning, indicating negative SP. Also, words such as *persistent* (Hunston, 2007), *Commit* and *dealings* (Partington, 1998, pp. 66-74), *budge* (Sinclair, 2004, pp. 142-147), and *cause* (Stubbs, 1995) were found to be mainly associated with negative contexts, suggesting an unfavorable connotation.

The fruitful SP studies on English words have also sparked scholars' interests in Mandarin Chinese (Chinese) words via cross-linguistic analysis. One representative case is Xiao and McEnery (2006), who conducted a comparative study of SP and semantic preference (i.e., semantic features that are associated with the collocates of a node item) of three sets of near synonyms in both English and their notional equivalents in Chinese and found that these synonyms showed different SP patterns and semantic preferences in each language. Specifically, synonyms of *cause* in Chinese such as *zhishi*, *niangcheng*, and *zaocheng* had a very strong negative SP, with very few concordances with positive/neutral evaluation; *cucheng* had a very strong positive SP; however, other synonyms such as *chansheng* and *dailai* had a rather mixed SP, with similar proportions of evaluative negative, positive, and neutral concordances. Similarly, Wei and Li (2014) tested the assumption whether the SP of a word in one language

would also be the same as the translation equivalent of the word in another language. They studied the SP and semantic preference of four English words and their translation equivalents in Chinese and revealed that each translation pairs had, in certain degree, different SP patterns and semantic preference. In addition to comparative studies, Li and Jiao (2017) investigated the SP of Chinese logical resultative formulae (e.g., *yinwei* ‘because’ and *youyu* ‘due to’) and discovered that *yinwei* had a neutral SP and *youyu* mixed SP.

Research on SP has also been extended to the analysis of the relationship between SP and register (e.g., categories such as hard news, soft news, and sports reporting are considered as registers in news genre) (Hunston, 2007; O'Halloran, 2007). That is, a lexical item may showcase a positive SP in one register but a negative SP in another. For example, the word *erupted* had a positive SP in the sports report register, as it was associated with contents of cheering, but a negative SP in the hard news register that associates with natural forces (O'Halloran, 2007). In addition, Partington (2017) suggested topic as a possible factor that influences the SP of words. He found that the SP of *orchestrate* had positive evaluation in the topics of music and sports, but negative evaluation in the topic of politics. However, the evidence he provided was limited and the classification of topic type was based on researcher's intuition.

Therefore, it is possible that the existence of a mixed-SP word, as revealed in the study of Xiao and McEnery (2006), may have certain affinity to text categories such as register or topic. In this study, topic is defined as different categories of articles in news genre, and covers a smaller scope of texts than register. Because a given lexical item may have a mixed SP, and little research has addressed the link between topic and the mixture of SP in a large scale, the current study thus aims to examine more closely the relationship between TOPIC and SEMANTIC PROSODY. In particular, we investigate three near-synonymous verbs of *cause* in Chinese, i.e., 產生

chansheng, 釀成 *niangcheng*, 促成 *cucheng*, regarding their SP to the topics of their co-occurring texts in a news genre corpus.

The objective of this study is to investigate how the SP of a lexical item may be mediated by the topic of the texts. We examined the SP distributions of *chansheng*, a mixed-SP node word, and both *niangcheng* and *cucheng*, strong-SP node words, under different topics in the Apple Daily News corpus. Meanwhile, the relationship between TOPIC and SEMANTIC PROSODY are examined under three hypotheses. The strong hypothesis predicts that topic has an effect on both a mixed-SP and strong-SP words. The moderate hypothesis predicts that topic only affects mixed-SP words. The null hypothesis predicts that topic has limited influence on the SP of words. If the strong hypothesis is true, we expect to see the SP distributions of *chansheng*, *niangcheng*, and *cucheng* vary significantly across different topics. If the moderate hypothesis is true, we expect to see only *chansheng* having a topic-dependent SP distribution. If the null hypothesis is true, we expect to see the SP distribution of three target words are unresponsive to topic change. Finally, the existence of either strong or moderate hypothesis would shed new light on how the semantic representation of a lexical item may be contextually dependent, even at the level of topic.

To explore the SP of a lexical item, two methods were conventionalized in the previous studies. Before introducing these methods, we define a few related terminologies. A node word is a central, target lexical item within a certain window size (e.g., 3:3 window size indicates a span of three words to the left and right of a node word). Collocates are words that co-occur with a node word within a stretch of texts. A concordance is a span of words that contain both a node word and its collocates within a pre-defined window size. Beginning with significant collocate analysis (Stubbs, 1995; Stubbs, 2001a), the collocates of a node word within a certain window

size which pass a researcher-defined threshold are classified into one of the three evaluations (i.e., positive, neutral, and negative). The SP of a node word is determined by the semantic distribution of these collocates. Another more common approach to determine the SP of a node word is concordance line analysis (Hunston, 2004; Hunston, 2007; Louw, 1993; Partington, 1998; Partington, 2017; Wei & Li, 2014; Xiao & McEnery, 2006). In this approach, the concordance lines of a node word in a corpus are often manually examined and categorized into positive, neutral, or negative evaluation by researchers. However, concordance line analysis is often time-consuming and labor-intensive, and often needs to limit the samples to a tractable size. Also, criteria for determining SP may be ambiguous across different annotators, leading to an issue of consistency.

In the present study, we adopt a hybrid method: rule-based concordance line analysis. This method was based on concordance line analysis with some modifications. Since it is impossible to manually annotate the entire concordances of three node words (e.g., more than 27,000 instances) in the corpus within the limited time, we developed a set of rules to efficiently classify each concordance of a node word into one of three evaluative meaning groups under each topic. The idea was inspired by the sentiment analysis, a computational linguistics approach, which has been effectively applied in mining opinions within customer reviews or finance news (Khedr, Salama, & Yaseen, 2017; Pang & Lee, 2008).

If the results from the above method show that there is an intimate interaction between TOPIC and SEMANTIC PROSODY, we can further identify the source that amplifies the SP distributions of the three node words, specifically under each topic in the corpus. As the research of Partington (2004) and Wei and Li (2014) suggests, SP of a node word has a strong relationship with semantic features of the collocates of that node word. Based on their

proposition, we hope to articulate the connection between the topic-dependent SP and semantic features. Hence, we utilized semantic network analysis to find the top 20 prototypical collocates of the node words under topics which significantly enhance either positive or negative SP tendency of a node word. Subsequently, we categorized the semantic features of those prototypical collocates and examined a few concordances of the node words under those topics, so as to verify whether such semantic features contribute to the rising of positive/negative SP.



2 Literature Review

2.1 Semantic Prosody

First made public in 1993, Louw described semantic prosody (SP) as “a consistent aura of meaning with which a form is imbued by its collocates” (1993, p. 157) and “a form of meaning which is established through the proximity of a consistent series of collocates” (2000, p. 57). A node item may acquire a connotational attribute of positive or negative evaluative meaning from the collocates it habitually co-occur with. In another perspective, Partington added that SP is “the spreading of connotational coloring beyond single word boundaries” (1998, p. 68). Thus, the favorable/unfavorable meaning of a node word is the product of the node word coordinating with its collocates within a certain span. The main function of SP is to express language user’s evaluation, attitude, or opinion toward a given topic in a pragmatic setting (Louw, 2000; Sinclair, 2004, p. 34; Stubbs, 2001b, p. 98).

Another important notion in research on SP is semantic preference, that is, the common semantic features of the collocates associated with a node item (Stubbs, 2001b). For example, Wei and Li (2014) compared the semantic preferences of the English-Chinese translation pair of *orchestrate* and *cehua* and found that the collocates of *orchestrate* typically come from the semantic subsets of (1) destructive/violent operations and events and (2) political and social activities/events, and the collocates of *cehua* from (1) destructive/violent operations events, (2) malicious plots and (3) constructive events/activities. Partington (2004, p. 151) further explained the interplay between semantic prosody and semantic preference: semantic prosody “dictates the general environment which constraints the preferential choices of the node item,” while semantic preference “contributes powerfully to building semantic prosody.” Therefore, there is an

intimate relationship between SP and semantic preference, such that the SP of a node word is conditioned by the semantic fields of its co-occurring words.

In terms of the grading scheme for the evaluative meaning of SP, most scholars adopt a trinary distinction: positive, neutral and negative, (Li & Jiao, 2017; Partington, 2004; Stubbs, 1995; Wei & Li, 2014; Xiao & McEnery, 2006). On the other hand, Sinclair (2004) regarded SP as attitudinal, and thus the categorical grading scheme is not applicable. For instance, the sequence of [attempt + negation + *budge*] as in a sentence “the kid tried to roll the stone, but it did not budge” expresses the sense of frustration at the failure of endeavor on something difficult. The SP of this sequence is rather complex and cannot be simply classified into one of the three evaluative categories.

There are also two streams of methods to discover the SP of a given node item: significant collocate analysis and concordance line analysis. The first method focuses on the consistent evaluative polarity of collocates of a node item. For instance, Stubbs (1995) extracted the significant collocates of a given node item within a certain span (e.g., 3 words before and after a node item) and determined the SP of the node word according to the majority of the evaluative meaning the collocates reveal. However, such a method may ignore the phenomenon of evaluative embedding discussed in Morley and Partington (2009) and Partington (2017). The concept of evaluative embedding notes that evaluative polarity is contingent upon the higher syntactic level in a string of words. Let us consider an example provided by Partington (2017, p. 196): in “*Global poverty is falling rapidly,*” while the phrase *global poverty* conveys a negative evaluation, the overarching meaning of the sentence is instead positive.

On the other hand, the concordance line analysis focuses on the consistent evaluative polarity of concordances where a node item exists. Researchers adopting this method extract

either all or a random-sampled subset of concordances of a given node item and manually examine the evaluative meaning each concordance conveys (Louw, 1993; Louw, 2000; Partington, 1998; Partington, 2004; Sinclair, 2004; Wei & Li, 2014; Xiao & McEnery, 2006). Nonetheless, the concordance analysis has limitation on its scalability since it is time-consuming to examine every concordance of a node item. Meanwhile, the operational criteria for determining the evaluative meaning of concordances may not be consistent and may differ from one scholar to another.

2.2 Context-dependent Semantic Prosody

Researchers have also drawn attention to the connection between SP and text category. O'Halloran (2007) proposed the notion of *register prosody* and stated “some prosodies have probabilistic relationships to register.” At the lower level to genre, a register is defined as a “configuration of meanings that are typically associated with a particular situational configuration of field, tenor, and mode” (Halliday & Hasan, 1989, p. 89). Thus, within the genre of news journalism, categories such as hard news, soft news, sports report, and recipes are considered as register (O'Halloran, 2007). In his study, the concordances of words such as *erupted* and *simmering* were examined under different registers. The results showed that *erupted* had negative SP under hard news register and positive SP under sport report register; *simmering* was associated with negative SP under hard news register but neutral SP under recipe register. This suggested that both words have a register-dependent SP. In parallel, Hunston (2007) pointed out the word *cause* lost its negative SP but gained the neutral SP in scientific register. She suggested that the different SP phenomena of *cause* were the results of register selection.

Furthermore, at the lower level of register, Partington (2017) pointed out the role of topic in the SP of the word *orchestrate*. Under topics of music and sports, since the word was used in

its literal meaning ‘combine harmoniously,’ it transmitted a positive SP; under topic politics, it was used metaphorically ‘plotting,’ thus leading to a negative SP. He claimed that the two senses of the node word *orchestrate* is the cause of topic-dependent SP. However, there are issues regarding his argumentation: only few number of concordance was examined, and the judgement of topic type of the concordances was purely subjective. What sets our study significantly different are that (1) three target node words do not have metaphorical senses, (2) there are large number of instances being investigated, and (3) each article in the news genre corpus was classified by the editors into different topics at the time of publication.

2.3 Semantic Prosody of CAUSE

2.3.1 Significant Collocate Analysis

Stubbs (1995) studied the SP of the node word *cause* in both its nominal (e.g., *the cause of*) and verbal (e.g., *cause, caused*) uses by examining the 120-million-word Cobuild Corpus and the 1-million-word LOB Corpus. In the former corpus, he identified the top 50 collocates within the 4:4 window (four words to the left and right of the node word) that were most associated with the node word *cause* in terms of *t*-score and MI scores, and the top 50 frequent collocates within the window of 3:3. Among the top 50 collocates ranked by *t*-scores, all the content words were predominantly negative, e.g., *anxiety, concern, crisis, damage, distress, embarrassment, aids, blood, cancer, and death*. Among the adjective collocates, however, more of them were neutral, e.g., *common, considerable, great, major, root*. On the other hand, the top 50 collocates ranked by MI scores revealed more low-frequency collocates with strong association with the node word, e.g., *consternation, grievous, uproar; célèbre, irreparable*. In regard of the top 50 frequent collocates, many of them have negative meaning, e.g., *problem(s), damage, death(s), disease, cancer, pain, trouble, and serious* (Stubbs, 1995, p. 15).

In general, the node word *cause* has clear negative SP as it co-occurs with negative words, which are attested in collocates based on statistical association and raw frequencies. However, Stubbs (1995) pointed out the disadvantage of using significant collocates analysis. Take the top 12 frequent collocates for example, although 8 of the 12 collocates are negative, the other words, e.g., *great*, *major*, and *common*, are apparently not negative. These words are in fact not modifier of the node word *cause*, but of the other nouns that happened to be included in the 4:4 window. For instance, the word *natural* in *cause for natural great concern*. Also, they are not the direct object of the verb *cause*, but the adjectival modifier of the object argument of *cause*, such as *grievous* in *cause grievous bodily harm* (Stubbs, 1995). Moreover, the researcher needed to manually remove those highly frequent functional words within the window, which were not considered as the contributing feature to the SP of *cause*.

2.3.2 Concordance Line Analysis

Xiao and McEnery (2006) conducted a comparative study of the SP of the *cause* group: *cause* and its near-synonyms (i.e. *arouse*, *lead to*, *result in/from*, *give rise to*, *bring about*) in English and eleven close translations of *cause* in Mandarin Chinese (i.e. *zhishi*, *niangchen*, *zaocheng*, *yinfa*, *daozi*, *dailai*, *yinqi*, *chansheng*, *xingchen*, *cushi*, *cucheng*). For English words, they used the Freiburg-LOB Corpus of British English (FLOB), the Freiburg-Brown Corpus of American English, and supplemented the corpora by the Brown University Corpus of American English and the British National Corpus when the former two corpora could not provide enough occurrences of the target words. For Mandarin Chinese, they used the Lancaster Corpus of Mandarin Chinese (LCMC) and supplemented the corpus by the *People's Daily* Corpus (2000).

Adopting the concordance line analysis, they investigated the SP of the target words according to the distributions of the evaluative meanings of the all the concordances of the *cause*

group in the corpora. Moreover, they used 4:4 window size to capture the significant collocates (with MI score greater than 3 and joint frequency at least 3 times) of the target words to investigate the collocational behavior of the *cause* group.

According to the proportions of evaluative meaning of the concordances reported in Xiao and McEnery (2006, p. 117), the synonym *arouse* had similar strong negative SP as *cause* (65% of the concordances of *arouse* were negative and 78% of the concordances of *cause* were negative) while the others synonyms such as *lead to*, *result in/from* and *give rise to* had weaker negative SP (*lead to*, *result in/from*, and *give rise to* have 49%, 47% and 46% negative concordances respectively), and *bring about* had mixed negative and positive SP (38% negative and 46% positive).

On the other hand, five of the eleven *cause* group in Chinese generally showcased the strong negative SP, such as *zhishi*, *niangcheng*, *zaocheng*, *yinfa*, and *daozihi*. However, the words such as *dailai*, *yinqi*, *chansheng* and *cushi* had inconsistent SP: *dailai* had 49% negative, 27% positive and 24% neutral concordances; *yinqi* had 43% negative, 15% positive, and 42% neutral concordances; *chansheng* had 31% negative, 24% positive, and 45% neutral concordances; *cushi* had 5% negative, 59% positive, and 36% neutral concordances). The node word *xingcheng* had a strong neutral SP with 64% neutral concordances, and *cucheng* had a very strong positive SP with 98% positive concordances.

According to the results, Xiao and McEnery (2006) concluded that SP is also phenomenal in Chinese language, and the near-synonyms of *cause* cannot be replaced with each other in both languages because of their different SPs and collocational patterns. Meanwhile, the SP profiles for the three target node words in this study are based on their findings: *chansheng*, the mixed-SP node word, and *niangcheng* and *cucheng*, the strong-SP node words.

2.4 Semantic Network Analysis

To conduct a semantic network analysis (SNA), we first need to identify significant collocates (i.e., collocation) of node words. The notion of collocation refers to the relation between two lexical items where the two co-occur together ‘with greater than random probability in its (textual) context’ (Hoey, 1991, pp. 6-7). To capture the collocation, many association measures (AMs) have been provided. The AMs can further be classified into bidirectional and directional. The bidirectional AMs include the MI, PMI, X^2 , t, z, LLR, Fisher-exact test; the directional AMs include *Delta P* (Evert, 2008; Gries, 2013; Manning & Schütze, 1999, pp. 169-183).

If it is the significant relation between a grammatical construction and a lexical item instead, such pattern is called collocation (Stefanowitsch & Gries, 2003). Here, the lexical item(s) appearing in the grammatical slot(s) of a construction is called collexeme(s). In parallel to collocation, the same AMs introduced above can also be applied in finding collocation, and such approach is collocation analysis (Stefanowitsch & Gries, 2003).

Ellis and his collaborators have utilized the collocation analysis and the SNA to study the learning of verb argument constructions (VACs) in language acquisition, e.g., [Verb Locative] construction (*A kangaroo jumped over the fence*) and [Verb + Object + Object] construction (*She baked me a cookie*) (Ellis & Ogden, 2017; Ellis, Römer, & O'Donnell, 2016). They were interested in the statistical patterns in language use that may account for the acquisition of verbs in different types of the VAC, including verb frequency, verb contingency, and semantic prototypicality. They derived the values of these three factors from the British National Corpus (BNC).

Verb contingency is the degree of faithfulness a verb has toward a construction. In the collostructional analysis, they used *Delta P* as the AM to test how strong a verb (i.e. collexeme) is attracted to a construction or how likely a verb can be predicted given the construction. The formula of *Delta P* is detailed in section 3.5 (Gries, 2013).

Semantic prototypicality of a verb is defined by its degree of betweenness centrality (BC) within the semantic network of a VAC where that verb appears (Ellis et al., 2016, pp. 82-86). Verbs that are attracted to a particular VAC can together form a semantic network belonging to that construction. Originally, BC was developed to quantify an individual's contribution to a solution in human communication network (Freeman, 1977). Likewise, a verb's relation to other verbs in a semantic network of a VAC can also indicate the degree of BC. This identifies the number of shortest paths between other verbs that pass through that verb over the number of shortest paths between any other pairs of verbs. The higher degree of BC a verb has, the more prototypical the verb will be. As in prototype theory, where a prototype of a category often summarizes most of the features shared by the members in the category (Rosch, 1975), a prototypical verb should bear the most resemblance to most of the other verbs in a semantic network (Ellis et al., 2016, pp. 92-94).

Ellis and Ogden's studies (2017; 2016) found that the three factors are crucial in both L1 and L2 acquisition. In L1 acquisition, the verb slot in children's VACs were mainly occupied by verbs that were more frequent and more faithful to their VACs in adult language, and were semantically prototypical. These verbs were also believed to drive the acquisition of VACs.

In L2 acquisition, Ellis et al. (2016) found that the first-acquired verb in each type of VAC (1) had more frequent occurrences in the verb slot in that construction, (2) was more faithful to that construction, and (3) was a more prototypical representation of the construction's

functional meaning. Moreover, the acquisition of the verb in a construction related significantly to its frequency and contingency, but not semantic prototypicality.

In our study, we employed the same approach of SNA illustrated in Ellis et al. (2016, pp. 82-86) to retrieve the prototypical collocates of a node word. Subsequently, we investigated the semantic preference of those prototypicality collocates so that we may grasp the idea of typical contents with which a given node word is often associated in a topic.



3 Methodology

Two analyses were conducted in the current study: automatic concordance line analysis and semantic network analysis. Figure 1 presents a work flowchart of our method of procedure. The detail of each step is explained in the following sub-sections.

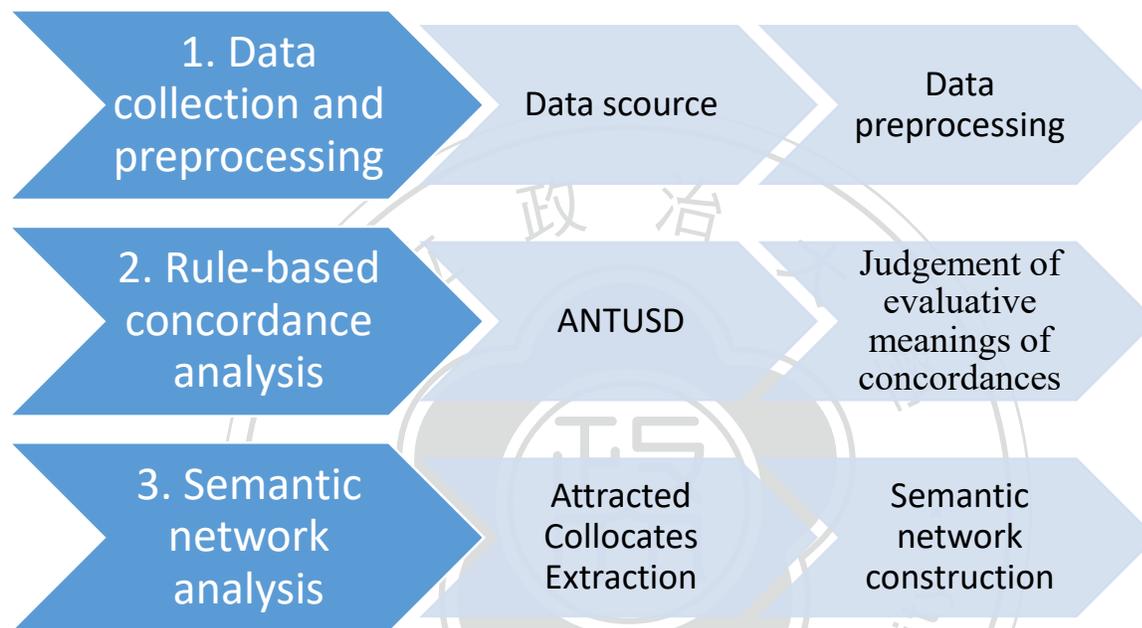


Figure 1. The flowchart of method of procedure

3.1 Data Collection and Preprocessing

3.1.1 Data Source

All the news articles¹ published online during 2003-05-02 to 2018-7-19 in the Apple Daily News²(ADN) were automatically collected using a self-developed web crawler written in Python language. We categorized the articles in the ADN corpus into different topics³ based on the

¹ Total number of articles during 2003-05-02 to 2018-7-19 are 1,354,185

² <https://tw.appledaily.com/>

³ The topic categories include: 生活 *shenghuo* ‘life’, 地方綜合 *difangzonghe* ‘sum of local news’, 投訴 *tousu* ‘sue’, 法庭 *fating* ‘court’, 社會 *shehui* ‘society’, 政治 *zhengzhi* ‘politics’, 要聞 *yaowen* ‘important news’, 搶鏡頭

annotations from the news provider. Among these topics, we filtered out general topics that included cross-topic contents (e.g., *yaowen* ‘important news’, *difangzonghe* ‘sum of local news’, *luntan* ‘forum’ and *luntanyuzhuanlan* ‘forum’) or ones which had few articles (e.g., *nuanliu* ‘warm current’ and *pinglunzhenxian* ‘Apple comment ranks’). Also, we merged topics with similar contents together, i.e., *fating* ‘court’ and *tousu* ‘file a lawsuit’ as topic **forum**. In the end, seven topics were chosen for our analysis, i.e., **society, politics, entertainment, international, sports, finance, and lifestyle**.

3.1.2 Data Preprocessing

First, we removed all the metadata in the articles, including hyperlinks, the author names of the articles, and image links and captions. Second, the JiebaR package⁴ was utilized to conduct word segmentation and part-of-speech (POS) tagging. The POS information was useful for retrieving content words (i.e., noun, verb, adjective, and adverb), which would be crucial to our later collocate extraction procedure. Also, based on the punctuations, we were able to break texts into chunk-size units. A chunk is delimited by punctuations⁵ and symbols, and may correspond to a sentence, clause, or phrase. The statistics of the article number and token number in each topic are summarized in Table 1.

qiangjingtou ‘scene stealing’, *暖流 nuanliu* ‘warm current’, *號外 haowai* ‘extra’, *網路新聞 wangluxinwen* ‘internet news’, *論壇 luntan* ‘forum’, *頭條 toutiao* ‘headline’, *蘋果爆破社 pingguobaoposhe* ‘Apple blasting club’, *蘋論陣線 pinglunzhenxian* ‘comment ranks’, *娛樂 yule* ‘entertainment’, *蘋果國際 pingguoguoji* ‘international’, *體育 tiyu* ‘sports’, *財經 caijing* ‘finance’, *地產 dichan* ‘home’, *副刊 fukan*, *論壇與專欄 luntanyuzhuanlan* ‘forum’.

⁴ <https://github.com/qinwf/jiebaR> The typical tags in JeibaR include n (noun), v (verb), a (adjective), t (time), r (pronoun), m (quantifier), d (adverb), p (preposition), c (conjunction), u (particle), e (exclamation), o (onomatopoeia), x (symbol), and w (punctuation).

⁵ Punctuations are included in the square brackets: [—, 、, : , ; , °, !, ? , !?]; symbols include ◎ ●

Table 1

Distribution of article number under each of the nine topics in ADN corpus

Topic name	Society	Politics	Enter.	Inter.	Sports	Finance	Lifestyle	Total
Number of articles	62527	36923	225314	97487	157972	139069	151729	871021
%	6.47	3.82	23.32	10.09	16.35	14.39	15.71	100
Token number	179798 20	117317 61	525698 40	237673 87	361804 43	5611637 6	6477659 7	2631222 24
Average token number	287.55	317.74	233.32	243.8	229.03	403.51	426.92	305.98

Third, we extracted the concordances of the three node words for SP analyses. We defined the span of the concordance line as 1:1 chunk-based window size. The concordance lines did not span over paragraph boundaries. An example was provided in (2):

(2)

- a. 今年賽制轉變產生差異在於成績領先車手先出發，可能因賽道上抓地力較差而吃虧，

jinnian saizhi zhuaibian chansheng chayi zaiwu chengji lingxian cheshou xian chubo,

genai yin saidao shang zhuadeli jiaocha er chikui,

This-year game-system change caused difference at score leading driver first start, may because track on grip worse and suffer-loose,

‘The difference in the system of this year 's competition is that the drivers with the leading scores start first and may suffer disadvantage from poor grip on the track,’

- b. 砸中在下方的工人，釀成意外。

za zhong zai xiafang de gongren, niangcheng yiwai.

Hit among at below ATTR worker, cause accident

‘hitting the worker below, causing the accident.’

- c. 美國和俄羅斯居中斡旋、促成的敘利亞停火協議，昨凌晨零時生效，

meiguo he eluoshi jizhong guanxuan, chuocheng de xulie tinghuo xieyi, zuo lingchen lianshi shengxiao,

US and Russia amidst mediate、cause/promote DE Syria ceasefire agreement, yesterday early-morning 12:00 am take effective

‘The ceasefire agreement of Syria meditated and caused/promoted by US and Russia was effective last night 12:00 am’

In (2a) and (2b), since the current chunk of the node word *chansheng/niangcheng* was at the beginning/end of a paragraph, only the current chunk (i.e., the chunk where a node word exists) and the following/preceding chunk were included as part of the concordance line. When the chunk of the node word was surrounded by punctuations, as *cucheng* in (2c), the concordance lines included the current chunk as well as one chunk preceding/following it. Table 2 shows the number of concordances of each node word under the seven topics.

Table 2

Number of concordances of each node word across different topics

Topic name	Society	Politics	Enter.	Inter.	Sports	Finance	Lifestyle
<i>chansheng</i>	1102	1112	2074	2808	1730	5593	8999
<i>niangcheng</i>	606	50	120	556	125	73	141
<i>cucheng</i>	29	403	631	586	200	805	149

3.2 Rule-based Concordance Line Analysis

3.2.1 ANTUSD

We classified the collocates in each concordance line into one of the three evaluative meaning groups according to the manual annotation of the sentiment labels (i.e., positive, neutral and negative) of words in the ANTUSD⁶ (Wang & Ku, 2016). The ANTUSD is the combination of words from The National Taiwan University Sentiment Dictionary (NTUSD), the NTCIR MOAT Task Dataset (NTCIR), the Chinese Opinion Treebank (COT), and the Advanced Chinese Bi-Character Word Morphological Analyzer Corpus (ACBiMA). It provided five sentiment-related fields for each annotated word: Pos, Neu, Neg, Non, and Not. These five fields were the number of times a given word was tagged as positive, neutral, negative, non-opinion word, and non-word respectively.

The NTUSD was the former version of the ANTUSD, containing 2812 positive and 8276 negative words, and was constructed for the purpose of sentiment analysis (Ku, Liang, & Chen, 2006). The sentiment of each word was determined by the majority among the three annotators.

⁶ <http://academiasinicanlplab.github.io/>

The two sentiment corpora are the NTCIR MOAT Task Dataset (NTCIR), the Chinese Opinion Treebank (COT). Three annotators labeled each sentence as positive, neutral, and negative in the two corpora based on the majority decision. Meanwhile, each annotator individually labeled words that he/she considered as sentiment words in each sentence as one of the three evaluations. Lastly, ACBiMA is the Chinese word morphological type. The ACBiMA contains about 11-thousand Chinese words along with their morphological types, and was developed to improve sentiment analysis by provided morphological information (Huang, Chen, & Kong, 2015). The words in the dataset were either randomly extracted from the Chinese Treebank 5.1⁷ and the NTCIR CIRB040 news corpus (Huang, Ku, & Chen, 2010), and were labeled as positive, neutral, negative, non-opinioned, and not-a-word by the agreement of the two annotators.

Due to its multiple sources, there were words with duplicated entries or with multiple values under the five fields in the ANTUSD. We followed a procedure to assign each matched collocate with a consistent sentiment label. First, if a collocate was found to have only one entry in the ANTUSD (94.5% of words have only one entry), we assigned the collocate with a sentiment label of the max value within the evaluation fields (i.e., Pos, Neu, and Neg). However, in cases where two evaluation fields shared the same max value, we assigned the collocate with the positive label if Pos and Neu share the same max value; the negative label if Neg and Neu share the same max value; and the neutral label if Pos and Neg share the same max value. Let us consider one example in Table 3. Since the word *milian* ‘obsessed’ was tagged 2 times as positive and negative, it was labeled as neutral.

⁷ <https://catalog.ldc.upenn.edu/LDC2005T01>

Table 3

Entry of milian 'obsessed' in the ANTUSD

Word	Pos	Neu	Neg
迷戀 <i>milian</i>	2	1	2

Second, if a collocate was found to have more than one entry in the ANTUSD, we chose the one of the max value within the evaluation fields and assigned the collocate with the sentiment label based on that entry. Consider an example in Table 4. There were two entries in the ANTUSD for the word *wugu* 'innocent'. The word was labeled as negative based on the first entry.

Table 4

Entry of wugu 'innocent' in the ANTUSD

Word	Pos	Neu	Neg
無辜 <i>wugu</i>	1	0	3
無辜 <i>wugu</i>	2	0	2

Third, if a matched collocate was found to have 0 value across the evaluation fields, it was not considered as sentiment word, and thus no sentiment label was given. The frequency distribution of Pos, Neu, and Neg with the max value among the evaluation fields of each word in the ANTUSD are summarized in Table 5.

Table 5

Frequency distribution of the max value (>0) in the evaluation fields of each word in the ANTUSD

Evaluation fields	<i>pos</i>	<i>neu</i>	<i>neg</i>	<i>pos and neu</i>	<i>neu and neg</i>	<i>pos and neg</i>	<i>pos, neu and neg</i>
Number of max value	11489	1791	13243	244	145	43	0

Note. There are 2708 instances that 0 are across the evaluation fields.

3.2.2 Judgement of Evaluative Meanings of Concordances and the SP of a Node Word

With the sentiment labels of the collocates established from the ANTUSD, the evaluative meaning of a given concordance was automatically determined. Inspired by the method of Li and Jiao (2017), We developed a set of heuristics to optimize the accuracy of the judgement of the evaluation of a concordance. First, if a concordance consisted of collocates which were uniformly tagged as positive/neutral/negative, that concordance was classified as positive/neutral/negative. Second, if none of collocates in a concordance were found in the ANTUSD, it was classified as neutral. Third, since the three target node words are all verbs, objects are main contributors of evaluative meaning of concordances. Thus, if one of the collocates following a node word in the current chunk was tagged as positive/negative, that concordance was classified as positive/negative evaluation; however, if there were positive and negative collocates following the node word, the evaluation of that concordance is determined by the majority number of positive/negative sentiment tags. Moreover, we took into account the effect of evaluative embedding discussed in Partington (2017), who pointed out that if a node

word was preceded by a negator (shown in Appendix 1), either positive or negative evaluative meaning of a concordance determined from the above rules would be reversed (cf. (3a)⁸).

Similarly, if a node word was preceded by a word expressing the idea of prevention (shown in Appendix 2), the originally negative evaluation of a concordance would be reversed to positive (cf. (3b)). Finally, if the chunk after the node word started with concessive conjunctions, e.g., *dan/danshi* ‘but’, and there were negative collocate(s) at the same time, it would be classified as negative evaluation (cf. (3c)).

(3)

- a. **快速** 上車 離去, **幸好** 沒(釀成)更**火爆**的**衝突**。

kuaisu shangche chiqu, xinhao mei (niangcheng) geng huobao de chongtu.

Quickly go-on-car leave, fortunate not cause more hot ATTR conflict

‘Get on the car quickly, and fortunately no more serious conflicts were caused.’

- b. 有通鼻、**預防** 鼻子 因 溫度 **變化**(產生) **過敏** **反應** 的 **功效**,

you tongbi, **yufang** bizi yin wendu bianhua (chansheng) guomin fanying de **gongxiao**,

have opening-nose, prevent nose because temperature change cause allergic reaction ATTR effect,

⁸ In (3), collocates tagged as negative were underlined and those tagged as positive are in bold, and the node words were in parentheses.

‘It has the effects of opening nose and preventing the allergic reaction of the nose due to temperature changes.’

- c. 他們也 **希望** 藉由 歌曲 (促成) 兩韓 **統一**，但 **擔心** 留在北韓的家人會 **遭** 到 當局 **報復**。

tamen ye xiwang jiyao gequ (chuo Cheng) lianghan tongyi, dan danxin liuzai
beihan de guren hui zaodao dangju baofu.

3-PL also hope through song cause two-Koreas unification, but worry stay North
Korea ATTR will suffer authority retaliate

‘They also hope that the two Koreas will be united through songs, but they fear
that the family members who stay in North Korea will be retaliated by the
authorities.’

In our first rule-based concordance line analysis, the SPs of the node words were determined based on the distribution of the three evaluative meanings of the concordances across different topics. In particular, by a cut-off rate of 50 percent (Li & Jiao, 2017), the SP of the node word was determined if one of the evaluative meanings reached over 50 percent. If the proportions of the three evaluative meanings of a node word were all below 50 percent, the node word was claimed to have a mixed SP.

To evaluate the credibility of our rule-based concordance line analysis, we first manually annotated one hundred random concordances of each of three node words. The number of these concordances under a topic is proportional to the percentage of number of concordances in that topic within the corpus. Then, we used Cohen’s kappa coefficient (k) to test the interrater

agreement between the rule-based and the human annotated results. In each case of the three node words, Cohen's kappa statistics all showed that there are moderate agreements between the two types of annotated results as shown in Table 6.

Table 6

Cohen's kappa statistics on interrater agreement in the three node words

	<i>z</i>	<i>p</i>	kappa
<i>chansheng</i>	6.6	<.001	0.51
<i>niangcheng</i>	4.85	<.001	0.52
<i>cucheng</i>	4.14	<.001	0.46

3.3 Semantic Network Analysis

3.3.1 Attracted Collocates Extraction

Our second analysis, semantic network, was built upon the collocates attracted to a given node word. For all the content words (i.e., noun, verb, adjective, and adverb) co-occurring with the node words within the 1:1 chunk-based window, we used *Delta P* to assess the directional association between a node word and its collocates under each topic. It measured how faithful a collocate was toward a node word (Gablasova, Brezina, & McEnery, 2017; Gries, 2013). If *Delta P* value of the collocate was greater than 0, it was attracted by the node word; conversely, if *Delta P* was lower than 0, the collocate was repulsed by the node word. We defined as significant collocates those collocates with *Delta P* > 0, and at the same time, with a joint frequency > 1 to remove hapax legomena (i.e., collocates occurring only once in the corpus).

Let us consider an example of the node word *chansheng* and its collocate *cuojui* ‘illusion’ under topic **society** in our corpus. The computation of *Delta P* value relied on the distributional information of the node word and its collocate, often cross-tabulated as a contingency table as shown in Table 7. *a* was the joint frequency of *chansheng* ‘cause’ and *cuojui* ‘illusion’; *b* was the frequency of the node word co-occurring without other collocates; *c* was the frequency of the collocate co-occurring with the other node items. Finally, *d* was the corpus size $N - (a+b+c)$, where *N* in the our analysis was defined as the total number of words tagged as verbs since our target node item was on a clause-level (Stefanowitsch & Gries, 2003, p. 218). Based on the formula provided in (4), the *Delta P* value between the node word *chansheng* ‘cause’ and *cuojui* ‘illusion’ was 0.05.

(4) *Delta P* Formula

$$\Delta P = P(O|C) - P(O|-C) = \frac{a}{a+b} - \frac{c}{c+d}$$

Table 7

An example of four types of frequencies necessary for calculating the *Delta P* value between the node word *chansheng* and *cuojui* ‘illusion’

	Outcome: <i>cuojue</i>	No Outcome: not <i>cuojue</i>
Cue: <i>chansheng</i>	a=40	b=758
No Cue: other node words	c=191	d=2,794,678

Note. $N = 2,795,667$

The type/token frequency of co-occurring words and attracted collocates of the node words *chansheng*, *niangcheng*, and *cucheng* are shown in Table 8, Table 9, and Table 10

respectively. Attracted collocates referred to those which passed the cut-offs of Delta P and joint frequencies.

Table 8

Number of collocates of chansheng across each topic before/after the filtering

Topic	Society	Politics	Enter.	Inter.	Sports	Finance	Lifestyle
Type number (co-occurring)	4389	3689	6281	8309	4659	9624	13487
Type number (attracted)	375	176	459	290	277	241	626
	(8.5%)	(4.8%)	(7.3%)	(3.5%)	(5.9%)	(2.5%)	(4.6%)
Token number (co-occurring)	11656	11173	20533	28507	15590	58887	88678
Token number (attracted)	2518	1123	3995	2474	1497	2986	15277
	(21.6%)	(10.1%)	(19.5%)	(8.7%)	(9.6%)	(5.1%)	(17.2%)

Note. The percentage of the number of type/token frequency of the collocates remained after the filtering are recorded in the parentheses.

Table 9

Number of collocates of niangcheng across each topic before/after filtering

Topic	Society	Politics	Enter.	Inter.	Sports	Finance	Lifestyle
Type number (co-occurring)	1954	314	746	2071	508	487	723
Type number (attracted)	138 (7.1%)	24 (7.6%)	71 (9.5%)	167 (8.1%)	66 (13%)	41 (8.4%)	62 (8.6%)
Token number (co-occurring)	4863	414	1068	4533	946	695	1104
Token number (attracted)	1323 (27.2%)	78 (18.8%)	250 (23.4%)	1211 (26.7%)	296 (31.2%)	155 (17.8%)	218 (19.7%)

Table 10

Number of collocates of cucheng across each topic before/after filtering

Topic	Society	Politics	Enter.	Inter.	Sports	Finance	Lifestyle
Type number (co-occurring)	257	1740	2425	2511	1035	3078	1013
Type number (attracted)	15 (5.8%)	103 (5.9%)	150 (6.2%)	162 (6.5%)	77 (7.4%)	75 (2.4%)	67 (6.6%)
Token number (co-occurring)	286	3749	5092	5219	1770	7659	1443
Token number (attracted)	13 (12.2%)	513 (13.7%)	690 (13.6%)	767 (14.7%)	229 (12.9%)	445 (5.8%)	203 (14.1%)

3.3.2 Semantic Network Construction

Semantic network analysis allows us to find a collocate with the prototypical semantics among the selected collocates of a node word. *Delta P* gave us a lexicon list, consisting of the collocates statistically attracted to the three node words. These attracted collocates were often semantically interrelated, and thus formed a semantic network, where some collocates were more prototypical. A prototypical collocate tends to encapsulate most of the semantic features shared by most collocates included in the network. Therefore, of particular importance to the construction of semantic network was the similarity in-between the collocates.

In the analysis of VACs, Ellis et al. (2016, pp. 82-86) used WordNet, an English lexical database built on an ontological framework where semantically similar words were grouped together, to measure the distance between each pair of verbs in the semantic space. Based on these distances, they built up a semantic network of verbs in a VAC construction. However, due to the limited Chinese tokens in Chinese Wordnet (Huang & Hsieh, 2010), we instead utilized GloVe (Pennington, Socher, & Manning, 2014), an effective computational learning algorithm, to transform words into vectors and obtained the semantic distance between each pair of words via cosine similarity.

GloVe is an unsupervised algorithm that learns the semantic relations between words from a corpus and generates representations for each word in the vector space. The training was initialized by converting a given raw corpus into a word-word co-occurrence matrix, in which rows corresponded to words and columns corresponded to context words (i.e., collocates in a window size). The semantic relation between two words was captured by the ratio of co-occurrence probabilities of the two words with each context word, which was mathematically encoded by the vector differences. The objective function of the model was to minimize the

proposed weighted least square function of the difference between the dot product of the word vectors of each two words and their co-occurrence probability in a corpus.

During the training process, there were three main parameters for users to set: context window size, symmetric/asymmetric window size, and dimension. Context window size referred to the range to consider collocate, and symmetric/asymmetric window size referred to whether the range to the left/right of a given word should be symmetric. Lastly, dimension was the dimensionality of word vector for each word. In our case, we used the ADN corpus as the training corpus to obtain the word representations in the vector space. We set the context window size as 5 with symmetric context window and set the dimension to 200. Then, based on the word representations in the vector space, we calculated the distance between each pair of words using cosine similarity, based on (5), where A and B referred to the vectors of two words respectively.

(5) Collocate Similarity (Cosine Similarity)

$$\text{cosine similarity} = \cos \theta = \frac{\vec{A} \cdot \vec{B}}{\|\vec{A}\| \|\vec{B}\|}$$

Third, we constructed different semantic networks for each topic based on the cosine similarity between each pair of collocates under different topics. A semantic network was composed of vertex (word type) and edge (connection between each vertex). The length of edge between each pair of words was based on their cosine similarity value. The higher a cosine similarity value shared between the two vertexes, the closer the two vertexes will be. We set the cosine similarity value 0.6 as a threshold to the connection between each node: Only pair of collocates with a cosine similarity value greater than or equal to 0.6 are connected. The details of building a semantic network can be found in Ellis et al. (2016, pp. 82-86).

After attaining the cosine distance between each pair of collocates in a network, we measured the semantic prototypicality of each collocate in the network by calculating the degree of betweenness centrality (Ellis et al., 2016, pp. 82-86). The betweenness centrality of a vertex is computed as follows:

(6)

$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Where σ_{st} is the total number of the shortest paths from vertex s to vertex t and $\sigma_{st}(v)$ is the number of those paths that involve v .

We used the visNetwork package (Almende, Benoit, & Titouan, 2018) to plot semantic networks based on the parameters computed above, i.e., word embedding, cosine similarity, and degree of betweenness centrality. In the visual representation of the network, each vertex referred to a collocate type and the thickness of the edges between each pair of vertexes were their mutual distances measured by cosine similarity. The size of a vertex was proportional to its semantic prototypicality which was calculated by the degree of betweenness centrality of that collocate in the network. The more semantically prototypical a collocate, the larger vertex size of the collocate. The evaluative meaning of a collocate was indicated by one of the three distinctive colors: blue for positive, grey for neutral, and red for negative.

The results from the SNA allow us to objectively identify the prototypical collocates of a node word in terms of their semantic commonalities. Given the fact that the semantic features shared by collocates may indicate the typical contexts a node word is often associated with, we

can therefore extrapolate the possible sources of the topic-dependent positive/negative SP for each node word.



4 Results

Our research question is to investigate whether topic has an effect on the SPs of both mixed-SP node word, *chansheng*, and strong-SP node words, *niangcheng* and *cucheng*, in Chinese. There were two main analyses for each node word. The first approach is the rule-based concordance line analysis. This had given us the SP distributions of each node word under each topic. Based on the results, a chi-square test was applied to test the association between TOPIC and SEMANTIC PROSODY. In the chi-square test, the standardized residuals (SR) in each combination of TOPIC and SEMANTIC PROSODY were computed, which allowed us to identify which topic had a greater impact on the SP. Any topic with an absolute SR value greater than 1.96 in one of the three meaning groups suggests that such topic-evaluative meaning pair greatly contributes to the rejection of the null hypothesis of no interaction existing between TOPIC and SEMANTIC PROSODY. The follow-up was the network analysis. We utilized the network analysis to find the prototypical collocates of each node word under topics that significantly enhanced either positive or negative SP tendency of that node word. Meanwhile, the figures of the semantic networks of these topics were provided. Furthermore, semantic features of the prototypical collocates in these topics would be generalized so that we can grasp the topical effect on the contents a node word emerge.

4.1 Semantic Prosody of Chansheng

4.1.1 Concordance Semantic Prosody Analysis

Figure 2 shows the overall distribution of the evaluative meanings of the concordances of the node word *chansheng* in the entire ADN corpus. The concordances with positive evaluation account for 29.1 %, neutral evaluation 10.6 % and negative evaluation 60.3 % of all the data. Our data suggest that *chansheng* has a primary negative SP with a less common positive SP.

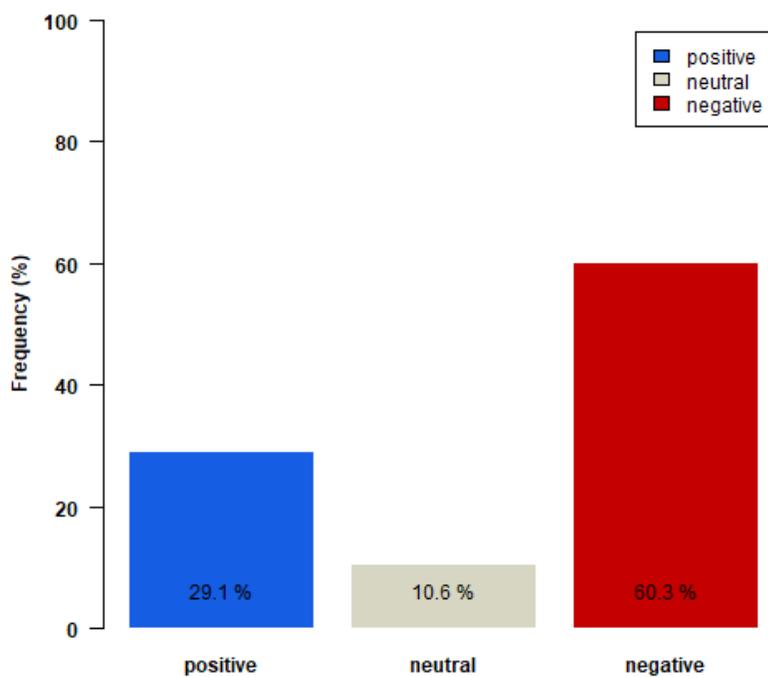


Figure 2. Distribution of the evaluative meanings of the collocate types of the node word *chansheng* in the ADN

Table 11 shows the SP distribution (i.e., positive, neutral, and negative) of the concordances of the node word *chansheng* across the seven topics. A chi-square test shows that there is a very significant association between TOPIC and SEMANTIC PROSODY ($\chi^2 = 403.48$, $df=12$, $p < .001$, Cramer's $V=0.26$).

Table 11

Standardized Residuals in a Chi-Square Contingency Table for the SP distribution of the node word chansheng under each topic

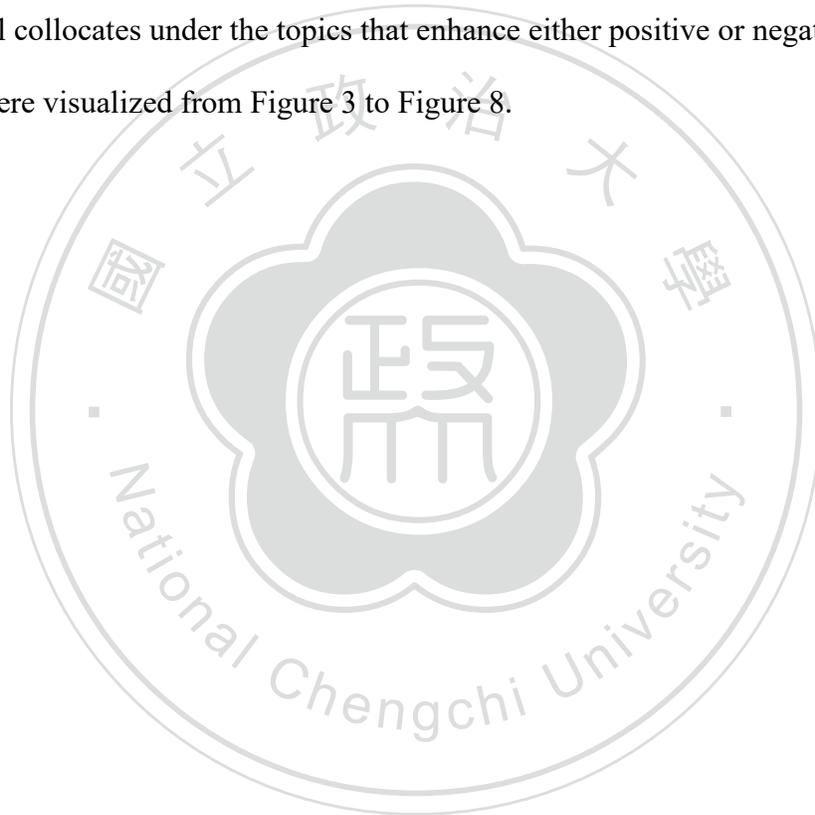
$\chi^2=548.93, df=12, p<.001,$		Society	Politics	Enter.	Inter.	Sports	Finance	Lifestyle
Cramer's $V=0.26$								
Positive	Observed	188	300	658	669	585	1709	2702
	Expected	320.189	323.1	603.411	816.773	503.57	1625.977	2617.981
	R	-7.387	-1.285	2.222	-5.171	3.629	2.059	1.642
Neutral	Observed	93	158	194	356	299	358	1031
	Expected	117.009	118.073	220.509	298.48	184.024	594.194	956.71
	R	-2.22	3.674	-1.785	3.329	8.476	-9.69	2.402
Negative	Observed	819	652	1221	1781	846	3519	5261
	Expected	662.802	668.827	1249.08	1690.747	1042.407	3365.828	5419.309
	R	6.067	-0.651	-0.795	2.195	-6.083	2.64	-2.15

The post-hoc residual analyses indicate that the proportions of positive evaluative meaning are much more than expected under the topics of **entertainment**, **sports**, and **finance**, and the proportions of negative evaluative meaning are much more than expected under the topics of **society**, **international** and **finance**. While *changsheng* demonstrates a primary negative

SP and a secondary positive SP in the entire corpus, the positive SP tendency of *chansheng* seems to be invoked under the topics of **entertainment**, **sports**, and **finance** and the negative SP tendency was intensified under the topics of **society**, **international** and **finance**. On the other hand, the residuals also suggest that *chansheng* has a weaker negative SP tendency under the topic **sports**.

4.1.2 Network Analysis

The prototypical collocates under the topics that enhance either positive or negative SP tendency of *chansheng* were visualized from Figure 3 to Figure 8.



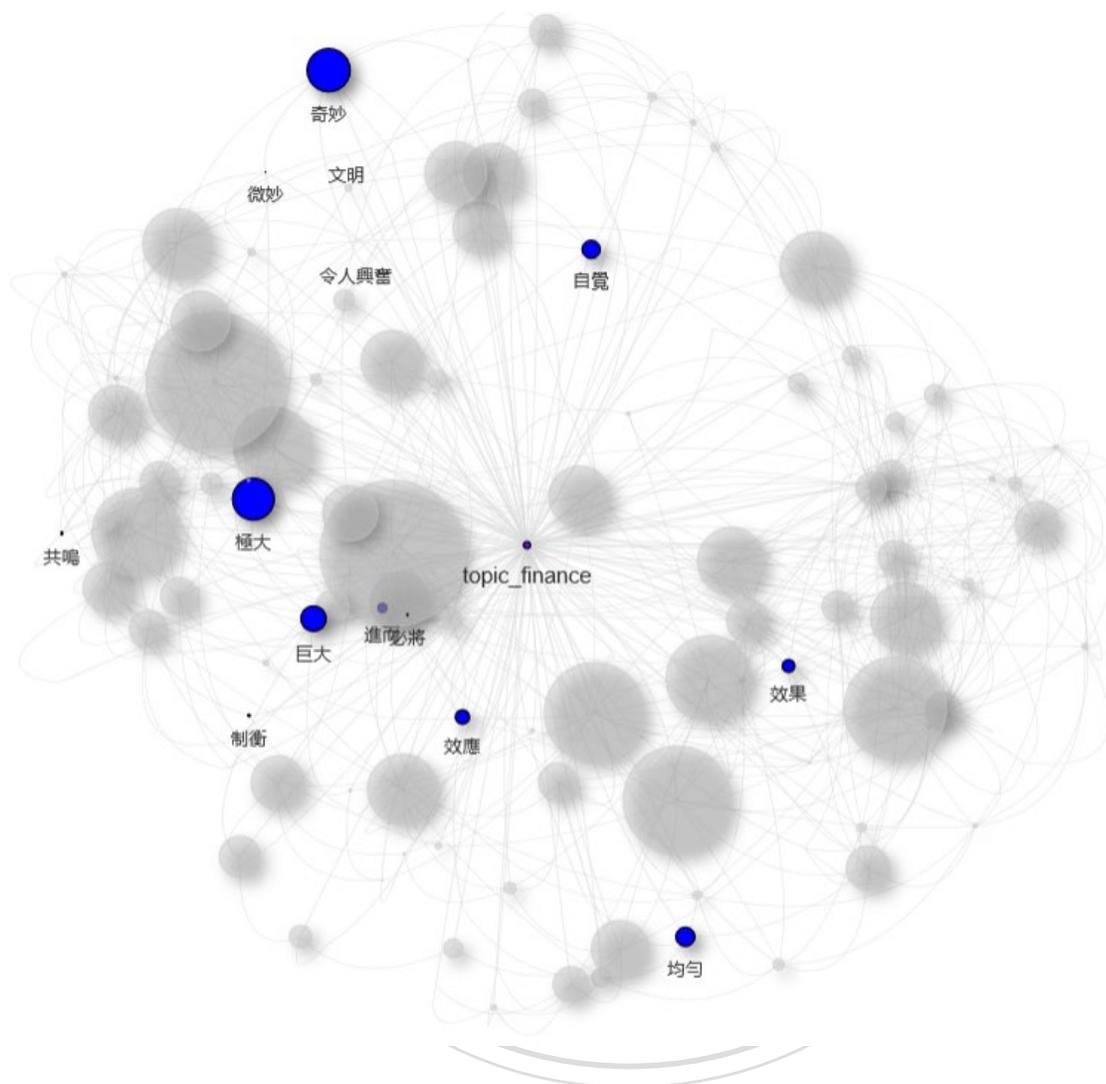


Figure 7. Subset semantic network under topic **finance** with positive collocates highlighted

(e.g., *gaishan* ‘improve’). The semantic features of the positive collocates for topic **sports** are (i) acts and behavior (e.g., *tihui* ‘experience’ and *jiexiao* ‘disclose’), (ii) abstract concept (e.g., *shizhi* ‘essence’ and *xiaoyon* ‘effectiveness’), (iii) optimal changes (e.g., *cujin* ‘promote’), (iv) prevention (e.g., *yufan* ‘prevent’) and (v) Attributes of physical entity/abstract concept (e.g., *jida* ‘very big’). The positive collocates within topic **finance** have semantic features of (i) acts and behavior (e.g., *zijue* ‘aware’), (ii) abstract concept (e.g., *xiaoguo* ‘effect’), (iii) attributes of physical entity/abstract concept (e.g., *qimiao* ‘amazing’), and (iv) balance (e.g., *zhiheng* ‘checks and balances’). These semantic features may be the contributor in increasing the positive SP tendency under the three topics.

In terms of topics that enhanced the negative SP tendency, we observed that the negative collocates under topic **society** have semantic features of (i) unpleasant emotion, thought and mental states (e.g., *jaolu* ‘anxiety’ and *kongju* ‘fear’), (ii) acts and behavior (*fanyin* ‘reaction’), (iii) health issues (e.g., *bingbian* ‘pathological changes’ and *fayan* ‘inflammation’), (iv) destructive/harmful/annoying events (e.g., *suanshan* ‘damage/injure’), and (v) change of state (e.g., *jiandi* ‘lower’). The negative collocates under topic **international** were found to have semantic features of (i) unpleasant emotion, thought and mental states (e.g., *jaolu* ‘anxiety’ and *yuanhen* ‘hatred’), (ii) acts and behavior (e.g., *yaohuang* ‘swaying’), (iii) health issues (e.g., *guomin* ‘allergic’), (iv) change of state (e.g., *biancha* ‘getting worse’), and (iv) natural phenomenon (e.g., *yanwu* ‘smoke’). Lastly, the semantic features of negative collocates for topic **finance** are (i) unpleasant emotion, thought and mental states (e.g., *wuhui* ‘misunderstanding’ and *maoduen* ‘paradox’), (ii) acts and behavior (e.g., *yapo* ‘oppress’), (iii) bad consequence (e.g., *fuzuoyong* ‘side-effect’), (iv) undesirable description about physical entity/abstract concept (e.g., *buliang* ‘not-well’), and (v) loss (*suenshi* ‘loss’ and *luocha* ‘undesirable difference’). Also, it is

possible that these semantic features are the source that intensify the negative SP tendency under the three topics.

Taken together, we found that the shared semantic features in topics where the positive SP tendency was increased are (i) acts and behavior and (ii) abstract concept. For the topics that intensified the negative SP tendency, the common semantic features are: (i) unpleasant emotion, thought and other mental states, (ii) acts and behavior, and (iii) health issues.



Table 12

*Negative prototypical collocates under topic **society***

Semantic features	Prototypical collocates
	Topic Society × negative SP
Unpleasant emotion, thought and mental states	焦慮 jaolu ‘anxiety’, 負面 fujian ‘negative’, 恐懼 kongju ‘fear’, 壓力 yali ‘pressure’, 誤解 wujie ‘misunderstanding’, 敏感 minggan ‘sensitive’, 疑慮 yilu ‘doubt’
Acts and behavior	反應 fanyin ‘reaction’, 干擾 ganrao ‘disturb’
Health issues	惡化 ehua ‘deteriorate’, 病變 bingbian ‘pathological changes’, 發炎 fayan ‘inflammation’, 中毒 zhongdu ‘poisoned’, 不適 bushi ‘discomfort’
Destructive/harmful/annoying events	損傷 suanshan ‘damage/injure’, 引爆 yinbao ‘detonate’, 爆炸 baozha ‘explode’, 噪音 zaoyin ‘noise’
Change of state	降低 jiandi ‘lower’
Others	強烈 qianglie ‘intense’

Table 13

Top 20 positive prototypical collocates under topic entertainment

Semantic features	Prototypical collocates
	Topic Entertainment × positive SP
Acts and behavior	促使 <i>cushi</i> ‘motivate’, 修復 <i>xoufu</i> ‘fix/reconcile’, 舒緩 <i>shuhuan</i> ‘smooth’, 信任 <i>xinren</i> ‘trust’
Abstract concept	認知 <i>renzhi</i> ‘cognition’, 效果 <i>xiaoguo</i> ‘effect’, 信心 <i>xinxin</i> ‘faith’
Description about physical entity or abstract concept	容易 <i>ronyi</i> ‘easy’, 明顯 <i>mingxian</i> ‘obvious’, 極大 <i>jida</i> ‘very-big’, 重大 <i>zhongda</i> ‘significant’, 濃厚 <i>nonghou</i> ‘rich’
Optimal change	改善 <i>gaishan</i> ‘improve’, 增長 <i>zengzhang</i> ‘increase’
Others	進而 <i>jiner</i> ‘so-as-to’, 造成 <i>zaocheng</i> ‘bring-about’, 有助 <i>youzhu</i> ‘helpful’, 將會 <i>jianhui</i> ‘will-do’, 光澤 <i>guanze</i> ‘luster’

Table 14

Top 20 negative prototypical collocates under topic international

Semantic features	Prototypical collocates
	Topic International × negative SP
Unpleasant emotion, thought and mental states	焦慮 jaolu ‘anxiety’, 負面影響 fumianyingxiang ‘negative impact’, 誤解 wujie ‘misunderstand’, 疑慮 yilu ‘doubt’, 怨恨 yuanhen ‘hatred’, 衝擊 chongji ‘shock’, 自卑 zibei ‘self-abased’
Acts and behavior	摩擦 moca ‘friction’, 搖晃 yaohuang ‘swaying’, 衝動 chongdong ‘impulsive’
Health issues	副作用 fuzuoyong ‘side-effect’, 過敏 guomin ‘allergic’, 後遺症 hoyizheng ‘sequela’, 病變 bingbian ‘pathological changes’, 暈眩 yunxuan ‘dizzy’
Change of state	變差 biancha ‘getting worse’, 變化 bianhua ‘change’, 減低 jiandi ‘reduce’
Natural phenomenon	煙霧 yanwu ‘smoke’, 燃燒 ranshao ‘burning’

Table 15

Top 20 positive prototypical collocates under topic sports

Semantic features	Prototypical collocates
	Topic Sports × positive SP
Acts and behavior	體會 tihui ‘experience’, 揭曉 jiexiao ‘disclose’, 吸收 xishou ‘absorb’, 決議 jueyi ‘decide’, 提名 timing ‘nominate’
Abstract concept	實質 shizhi ‘essense’, 效用 xiaoyon ‘effectiveness’, 效果 xiaoguo ‘effect’, 動機 dongji ‘motivation’, 商機 shanji ‘business opportunity’, 成就感 chengjiougan ‘sense-of- achivement’
Optimal change	促進 cujin ‘promote’, 新增 xinzeng ‘add’
Prevention	預防 yufan ‘prevent’
Attributes of physical entity/abstract concept	極大 jida ‘very big’, 常見 chanjian ‘often-seen’, 獨特 dute ‘unique’, 自然 ziran ‘natural’
Others	進而 jiner ‘so-as-to’, 現行 xiangxing ‘be-in-effect’

Table 16

*Positive prototypical collocates under topic **finance***

Semantic features	Prototypical collocates
	Topic Finance × positive SP
Acts and behavior	自覺 zijue ‘aware’
Abstract concept	效果 xiaoguo ‘effect’, 效用 xiaoyon ‘effectiveness’
Attributes of physical entity/abstract concept	極大 jida ‘very big’, 奇妙 qimiao ‘amazing’, 巨大 juda ‘gigantic’, 微妙 weimiao ‘mystical’
Balance	制衡 zhiheng ‘checks and balances’, 均勻 junyun ‘balance’
Others	進而 jiner ‘so-as-to’, 必將 bijian ‘inevitably’, 共鳴 gongming ‘resonance’

Table 17

Top 20 negative prototypical collocates under topic finance

Semantic features	Prototypical collocates
	Topic Finance × negative SP
Unpleasant emotion, thought and mental states	疑慮 yilu ‘doubt’, 誤會 wuhui ‘misunderstanding’, 憂鬱 yolu ‘melancholic’, 矛盾 maoduen ‘paradox’, 疑惑 yihuo ‘doubt’, 罪惡感 zuiegan ‘guilty’, 負面 fumian ‘negative’, 負面影響 fumianyingxiang ‘negative impact’
Acts and behavior	對立 duili ‘oppose’, 摩擦 moca ‘friction’, 危害 weihai ‘harm’, 壓迫 yapo ‘oppress’, 扭曲 niuqu ‘distort’
Bad consequence	副作用 fuzuoyong ‘side-effect’, 後遺症 hoyizheng ‘sequela’
Undesirable description about physical entity/abstract concept	不良 buliang ‘not-well’
Loss	損失 suenshi ‘loss’, 落差 luocha ‘undesirable difference’
Others	細菌 xijun ‘bacteria’, 臭味 chowei ‘smelly odor’

4.2 Semantic Prosody of Niangcheng

4.2.1 Concordance Semantic Prosody Analysis

Figure 9 shows the overall distribution of the evaluative meanings of the concordances of the node word *niangcheng* under the ADN corpus. The concordances with negative evaluation account for about 74.5 %, positive evaluation 22.1 and neutral evaluation sentiments 3.4 % of all the data. The node word *niangcheng* has a strong negative SP.

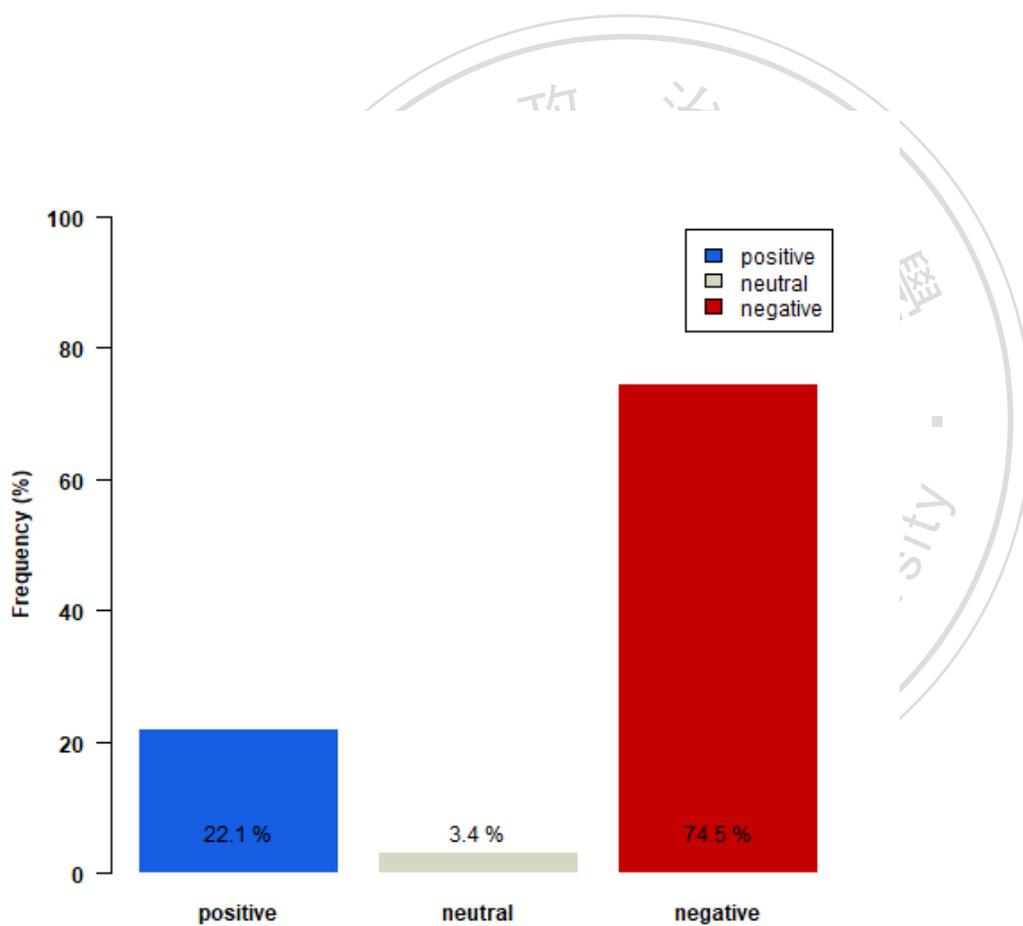


Figure 9. Distribution of the evaluative meanings of the concordances of the node word *niangcheng* under the seven topics

Table 18 shows the SP distribution of the concordances of the node word *niangcheng* across the seven topics. A chi-square test shows that there is a significant association between TOPIC and SEMANTIC PROSODY ($\chi^2=128.17$, $df=12$, $p=p<.001$, Cramer's $V=0.15$).

Table 18

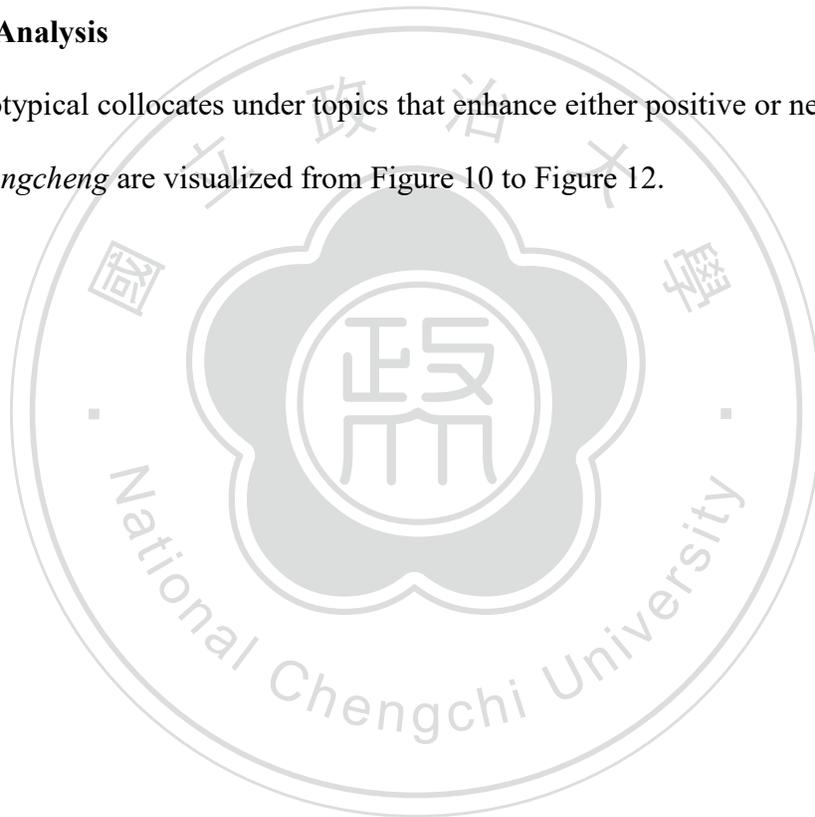
Standardized Residuals in a Chi-Square Contingency Table for the SP distribution of the node word niangcheng under each topic

		Society	Politics	Enter.	Inter.	Sports	Finance	Lifestyle
$\chi^2=128.17$, $df=12$, $p<.001$, Cramer's $V=0.15$								
<i>Positive</i>	Observed	136	9	32	82	43	17	50
	Expected	133.82	11.041	26.499	122.779	27.603	16.12	31.136
	<i>R</i>	0.188	-0.614	1.069	-3.68	2.931	0.219	3.381
<i>Neutral</i>	Observed	8	2	6	14	2	3	22
	Expected	20.671	1.706	4.093	18.966	4.264	2.49	4.81
	<i>R</i>	-2.787	0.225	0.942	-1.14	-1.096	0.323	7.838
<i>Negative</i>	Observed	462	39	82	460	80	53	69
	Expected	451.508	37.253	89.408	414.255	93.133	54.39	105.054
	<i>R</i>	0.494	0.286	-0.783	2.248	-1.361	-0.188	-3.518

The post-hoc residual analyses indicate that the proportions of positive evaluative meaning are much more than expected under the topics of **sports** and **lifestyle**, and the proportion of negative evaluative meaning is much more than expected under topic **international**. With the strong negative SP, the positive SP tendency of *niangcheng* is elicited under the topics of **sports** and **lifestyle** and negative SP tendency is further intensified under topic **international**.

4.2.2 Network Analysis

Again, the prototypical collocates under topics that enhance either positive or negative SP tendency of *niangcheng* are visualized from Figure 10 to Figure 12.



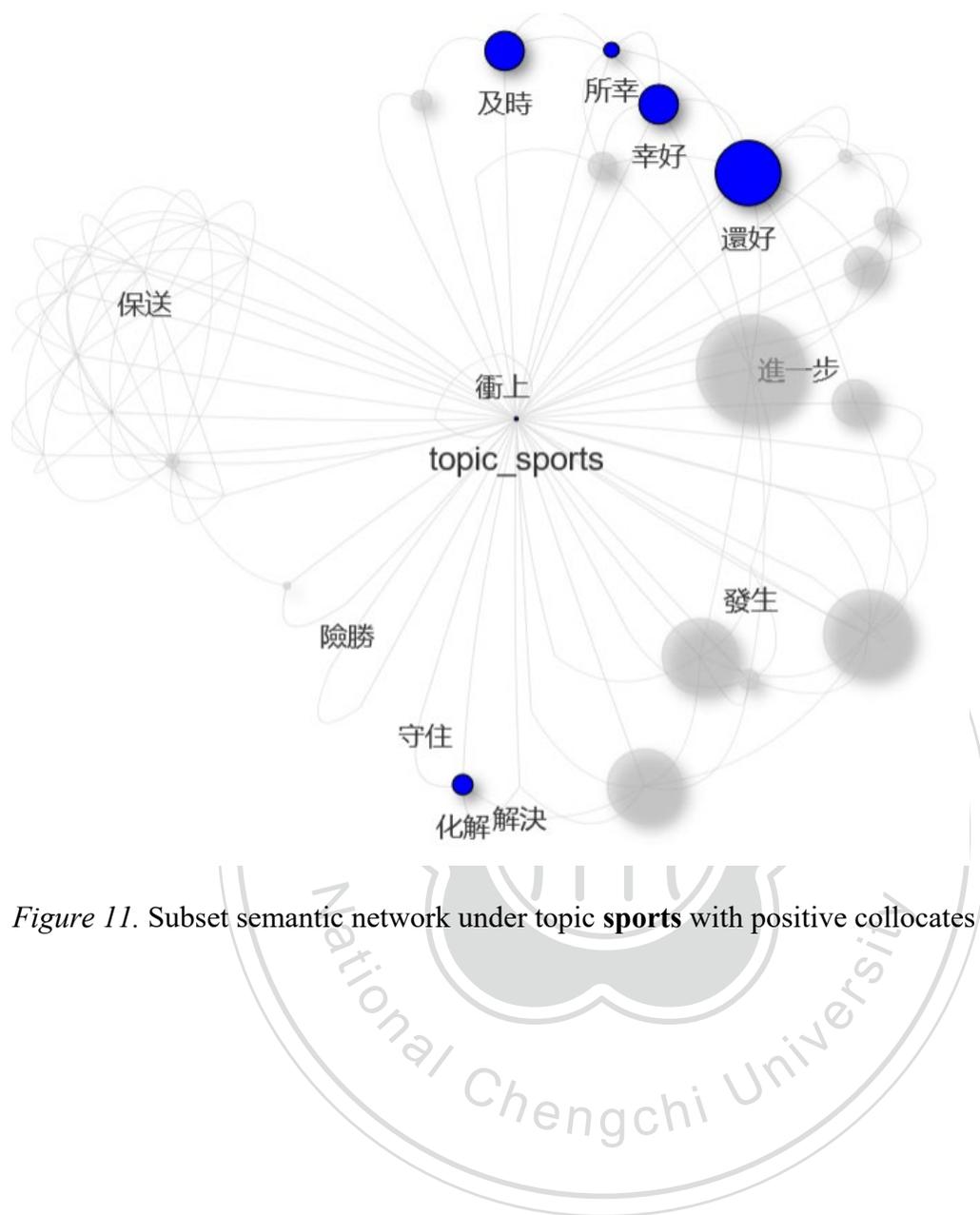


Figure 11. Subset semantic network under topic **sports** with positive collocates highlighted

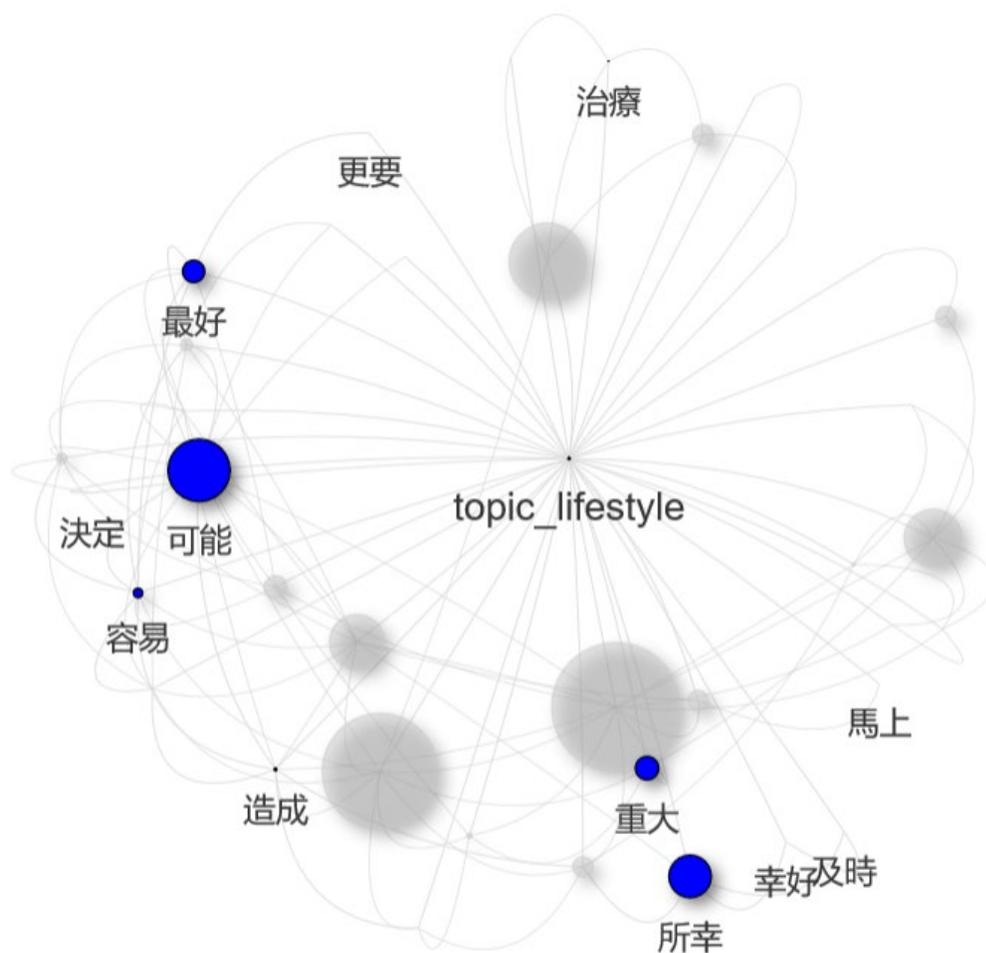


Figure 12. Subset semantic network under topic **lifestyle** with positive collocates highlighted

According to semantic features, the prototypical collocates of *niangcheng* under these topics are classified into different semantic subsets as shown in from Table 19 to Table 21. In topics where the positive SP tendency was elicited, the semantic features of the positive collocates for topic **sports** are (i) desirable description (e.g., *xinghao* ‘fortunate’) and (ii) settling issues. For topic **lifestyle**, the positive collocates have semantic features of (i) desirable description (e.g., *ronyi* ‘easy’ and *zhongda* ‘significant’) and (ii) medical manner (e.g., *zhiliao* ‘treatment’). These semantic features may be the source that triggered the positive SP tendency. The only shared semantic feature in the two topics is desirable description.

In terms of a topic where the negative tendency was intensified, we found that in topic **international**, the semantic features of the negative collocates are (i) acts and behavior (e.g., *tuiji* ‘push’), (ii) destructive/harmful events, accidents and disaster (e.g., *zainan* ‘disaster’ and *baozha* ‘explode’) and (iii) natural disaster (e.g., *haixiao* ‘tsunami’). The enhancement of negative SP tendency may due to these semantic features. Noticeably, most of the negative collocates are associated with the second semantic set.

Table 19

Top 20 negative prototypical collocates under topic international

Semantic features	Prototypical collocates
	Topic International × negative SP
Acts and behavior	推擠 <i>tuiji</i> ‘push’
Destructive/harmful events, accidents and disaster	意外 <i>yiwai</i> ‘accident’, 災難 <i>zainan</i> ‘disaster’, 傷亡 <i>shanwang</i> ‘casualties’, 事故 <i>shigu</i> ‘incident’, 車禍 <i>chewo</i> ‘car accident’, 火災 <i>huozai</i> ‘fire disaster’, 爆炸 <i>baozha</i> ‘explode’, 失控 <i>shikong</i> ‘out of control’, 縱火 <i>zonghuo</i> ‘arson’, 災害 <i>zaihai</i> ‘disaster’, 喪生 <i>sangsheng</i> ‘die’, 動亂 <i>dongran</i> ‘turmoil’, 起火 <i>qihuo</i> ‘outbreak of fire’, 死傷 <i>sishang</i> ‘death and injury’, 喪命 <i>sangming</i> ‘die’, 衝突 <i>chongtu</i> ‘conflict’, 悲劇 <i>beijui</i> ‘tragedy’
Natural disaster	海嘯 <i>haixiao</i> ‘tsunami’
Others	不慎 <i>bushen</i> ‘accidentally’

Table 20

Positive prototypical collocates under topic sports

Semantic features	Prototypical collocates
	Topic Sports × positive SP
Desirable description	幸好 xinghao ‘fortunate’, 還好 haihao ‘fortunate’, 及時 jishi ‘on-time’
Settling issues	化解 huajie ‘resolve’

Table 21

Positive prototypical collocates under topic lifestyle

Semantic features	Prototypical collocates
	Topic Lifestyle × positive SP
Desirable description	容易 ronyi ‘easy’, 重大 zhongda ‘significant’, 最好 zuihao ‘the best’, 所幸 suoxing ‘fortunate’, 可能 keneng ‘possible’
Medical manner	治療 zhiliao ‘treatment’
Others	造成 zaocheng ‘bring-about’

4.3 Semantic Prosody of *Cucheng*

4.3.1 Concordance Semantic Prosody Analysis

Figure 13 shows the overall distribution of the concordances of the node word *cucheng* under the ADN corpus. The concordances with positive evaluation accounts for 74.2 %, neutral evaluation 5.8 %, and negative evaluation 19.9 % of the data. The node word *cucheng* shows a strong positive SP.

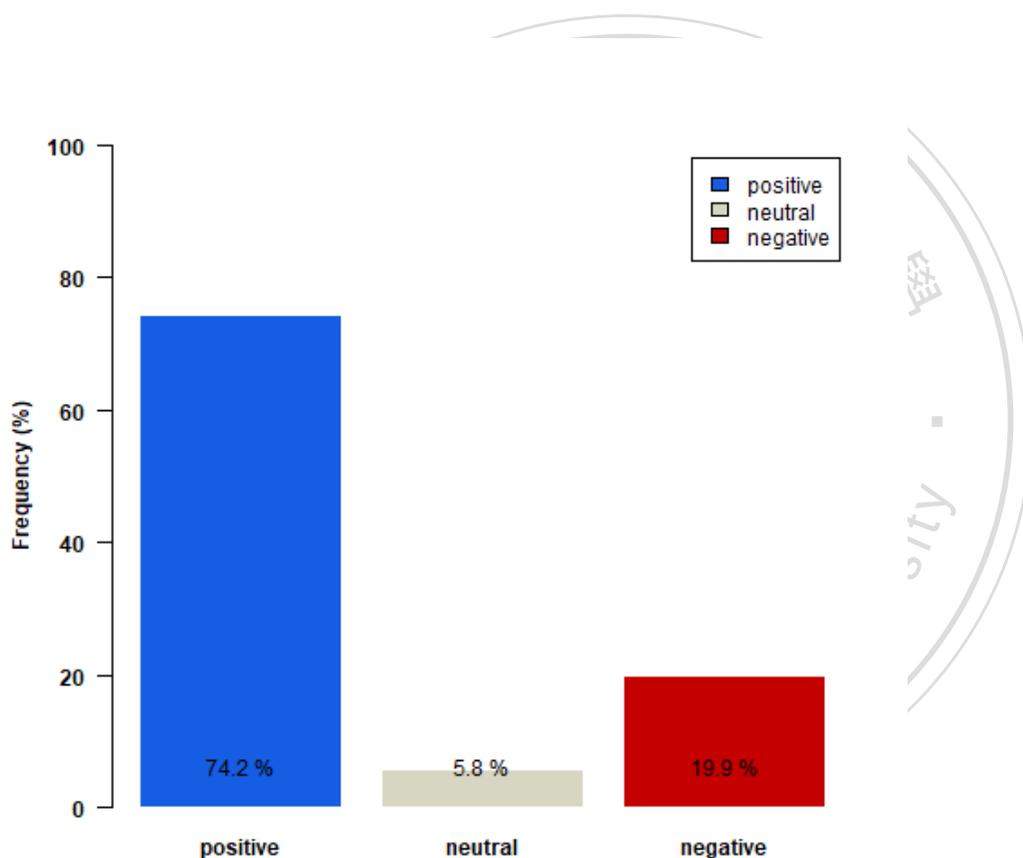


Figure 13. Distribution of the evaluative meanings of the collocates of the node word *cucheng* in the ADN

Table 22 shows the SP distribution of the concordances of the node word *cucheng* across the seven different topics. A chi-square test shows that there is a significant association between TOPIC and SEMANTIC PROSODY ($\chi^2=54.93$, $df=12$, $p < .001$, Cramer's $V=0.1$).

Table 22

Standardized Residuals in a Chi-Square Contingency Table for SP distribution the node word cucheng under each topic

		Society	Politics	Enter.	Inter.	Sports	Finance	Lifestyle
$\chi^2=54.93$, $df=12$, $p < .001$, Cramer's $V=0.1$								
<i>Positive</i>	Observed	16	303	459	403	151	633	116
	Expected	21.53	299.195	468.466	435.057	148.484	597.647	110.62
	<i>R</i>	-1.192	0.22	-0.437	-1.537	0.206	1.446	0.511
<i>Neutral</i>	Observed	3	20	58	21	13	41	7
	Expected	1.686	23.435	36.694	34.077	11.63	46.812	8.665
	<i>R</i>	1.012	-0.71	3.517	-2.24	0.402	-0.85	-0.566
<i>Negative</i>	Observed	10	80	114	162	36	131	26
	Expected	5.783	80.37	125.84	116.866	39.886	160.54	29.715
	<i>R</i>	1.753	-0.041	-1.055	4.175	-0.615	-2.331	-0.681

The topic-dependent negative prototypical collocates under topic **international** can be generalized as (i) unpleasant emotion, thought and other mental states (e.g., *maoduen* ‘paradox’), (ii) termination of events (e.g., *ezhi* ‘stop’), (iii) warfare (e.g., *lenzhan* ‘cold war’ and *jiangju* ‘deadlock’), and (iv) change of state (e.g., *zhuanbian* ‘change’) (see Table 23). These semantic features may account for the elicitation of the negative SP tendency.

Table 23

Top 10 negative prototypical collocates under topic international

Semantic features	Prototypical collocates
	Topic International × negative SP
Unpleasant emotion, thought and other mental states	矛盾 <i>maoduen</i> ‘paradox’
Termination of events	終結 <i>zhongjie</i> ‘terminate’, 遏止 <i>ezhi</i> ‘stop’, 落幕 <i>luomu</i> ‘ending’
Warfare	冷戰 <i>lenzhan</i> ‘cold war’, 僵局 <i>jiangju</i> ‘deadlock’, 敵對 <i>didui</i> ‘hostility’, 派系 <i>paixi</i> ‘faction’
Change of state	轉變 <i>zhuanbian</i> ‘change’, 削減 <i>xiaojian</i> ‘reduce’

4.4 Internal Summary

Throughout the entire corpus, the SP distribution of the concordance lines for each node was summarized in Table 24. The node word *chansheng* has a primary negative SP with a secondary

positive SP. The node words *niangcheng* and *cucheng* have a strong negative SP and positive SP respectively.

Table 24

Semantic prosody profile of the three node words

Node word	Positive %	Neutral %	Negative %	Semantic prosody
<i>chansheng</i>	29.1	10.6	60.3	Negative
<i>niangcheng</i>	22.1	3.4	74.5	Negative
<i>cucheng</i>	74.2	5.8	19.9	Positive

The topic-dependent SP distribution of the concordance lines of the three node words was summarized in Table 25. In the diagram, “positive” indicates a positive SP, and “negative” refers to a negative SP. A red label with an upward-pointing arrow indicates the SP distribution has a residual value greater than 1.96 in a topic; conversely, a green label with a down-facing arrow indicates the SP distribution has a residual value lower than -1.96 in a topic. Two important patterns should be noted. First, a consistent increment of the proportion of the negative SP was observed under topic **international**. Secondly, the arousal of positive SP of *chansheng* and *niangcheng* was discovered under topic **sports**.

Table 25

Comprehensive view of SP tendencies across topics

	Society	Politics	Enter.	Inter.	Sports	Finance	Lifestyle
產生 chansheng	↓ positive		↑ positive	↓ positive	↑ positive	↑ positive	
	↑ negative			↑ negative	↓ negative	↑ negative	↓ negative
釀成 niangcheng				↓ positive	↑ positive		↑ positive
				↑ negative			↓ negative
促成 cucheng							
				↑ negative			

In terms of the semantic features of the prototypical collocates (“prototypical features” for convenience) of a node word under topics that intensify the positive/negative SP tendency of that node word, some phenomena were detected. The common positive prototypical features for *chansheng* were (i) acts and behavior and (ii) abstract concept; for *niangcheng* was desirable description. On the other hand, the common negative ones for *chansheng* were (i) unpleasant emotion, thought and other mental states, (ii) acts and behavior, and (iii) health issues; for *niangcheng* was destructive/harmful events, accidents and disaster.

5. Discussion

The current study has shown that (1) *chansheng* has a primary negative SP and a secondary positive SP, (2) *niangcheng* a strong negative SP, and (3) *cucheng* a strong positive SP in the ADN corpus. In parallel to Xiao and McEnery (2006), the last two node words echo the findings in their study; nonetheless, *chansheng* contradicts to their finding where the node word was found to have a mixed SP (31% negative and 24 % positive). A possible reason for the surge of negative SP is the influence of negative disposition under news genre. In journalism, it is argued that “the more negative the event in its consequences, the more probable that it will become a news item” (Cohen & Young, 1981, p. 56). Thus, the negative SP tendency of *chansheng*, which has a mixed SP in domain-general texts (hereafter ‘general’), may be amplified under the ADN corpus. Compatible to the SP trend of news, *niangcheng*, which has a negative SP in general, continues to show a strong negative SP. On the other hand, despite exposing to the overall sentimental orientation of news genre, *cucheng* nonetheless carries a strong positive SP. Its strong positive SP in general may entail its insensitivity to the topical influence of the news genre.

The SP tendency of the mixed-SP node word, *chansheng*, varies to a great degree across different topics; conversely, the SP tendencies of the strong-SP node words, *niangcheng* and *cucheng*, are relatively stable despite exposing to diverse topic types. As presented in the individual result section of each node word, the Cramer’s V value informs that only the strength of the association between TOPIC and SEMANTIC PROSODY is moderately strong for *chansheng*, but weak for the other two node words. Thus, we suggest that a mixed-SP node word in general is susceptible to the topic change in a way that it serves as a vehicle to resonate the overall sentiment and content nature of a given topic. On the other hand, a strong-SP node word in

general is rather insensitive to the topic change and has a more monotonous SP usage. Since only the mixed-SP node word is found to be influenced by topic, the moderate hypothesis is confirmed in our study. The weak influence of topic on the strong-SP node words rejects the strong hypothesis. Lastly, the null hypothesis of no correlation existing between topic and the SPs of any node words is not supported by the evidence of topic-dependent SP of *chansheng*.

Along with the well-validated notion of *register prosody* by O'Halloran (2007), we further propose the idea of *topic prosody*—context-dependent SP operates at the level of topic. Based on the findings in the present study, the phenomenon of a topic-dependent SP is confined to *chansheng*, the mixed-SP node word. We might also argue that topic has higher probability in selecting the SP phenomenon of mixed-SP node words more often than strong-SP node words.

In semantic network analysis, topic-dependent prototypical features generalized from prototypical collocates of the three target node words may contribute to the increase of positive/negative SP tendency for these node words. To find out how the prototypical features are connected to the main themes of the news article under a given topic, we decided to utilize the latent Dirichlet allocation (LDA) algorithm (Blei, Ng, & Jordan, 2003) to explore the typical semantic content/matter of the news coverage under each topic. LDA is an effective data-mining technique which offers analysts a quick look of the thematic distribution across documents under a collection of texts. Moreover, LDA has been employed for tasks such as topic discovery in scientific papers (Griffiths & Steyvers, 2004) and recommendation for data annotation (Krestel, Fankhauser, & Nejd, 2009). We applied LDA to the six topics where the SP of one of the node words is intensified significantly. We selected an arbitrary number and classified the news collection of each topic into ten themes. The naming for each theme, along with topic 15 relevant

words under each theme, under each topic (i.e., **society, entertainment, international, sports, finance, lifestyle**) are shown from Appendix C to Appendix H.

Listed below are themes generalized by LDA within the aforementioned topics; these themes contain the news subject matter that journalists typically report about. Among them, we provide a few concordances demonstrating how the prototypical semantic features of a node word relate to the content of news articles.

5.1 Topics Triggering Positive SP

Under topic **entertainment**, it is observed that the positive tendency of *chansheng* rises. The typical ten themes are *jingpin* ‘boutique’, *biaoyan* ‘performance’, *falujiufen* ‘lawsuit’, *baoyang* ‘health maintenance’, *shenghuoyule* ‘life entertainment’, *yiliao* ‘medical care’, *ganqingbagua* ‘gossiping’, *yingshi* ‘television’, *jiating* ‘family’, and *dianying* ‘movie’ (see Appendix D). Examples expressing positive evaluation for the topic are provided in (7).

(7)

- a. 踢踏舞等新元素，和兒童劇產生了全新的「化學變化」。

dietawu deng xin yuansu, he ertongju chansheng le quanxin de

「huaxuebianhua」。

tap-dance so-on new element, CONJ children's drama cause PFV brand-new

GEN 「chemical-change」

‘New elements such as tap dance and children's drama have created a new

"chemical change.”

- b. 透過二倍活化每一層的年輕蛋白，有效打擊隨著年齡增長產生的靜態紋，讓鬆弛不再，

touguo erbei huohua mei yiceng de nianqing danbai, youxiao daji suizhe nianling zengchang chansheng de jingtai wen, rang songchi buzai,

through two-time activation each layer GEN young protein, effective strike with age grow cause GEN static wrinkle, let looseness no-more,

‘Through doubling young protein that activates each layer, it effectively combats the static lines that grow with age, stopping the looseness.’

- c. 綺麗風光收錄於衣裙之間，並融入60年代未來風格輪廓，產生出一種明快俐落的度假意象。

qili fengguang shoulu wu yiqun zhijian, bing rongru 60 niandai weilai fengge lunkuo, chansheng chu yi zhong mingkuai lila de dujia yixiang.

beautiful scenery record in dress within, and blend 60 period future style outline, create out one CLF bright clear-cut GEN vacation image.

‘The beautiful scenery is included in the dress and blended into the futuristic outline of the 1960s, creating a bright and clear-cut holiday image.’

It is possible that (7a) belongs to the theme of *biaoyan* ‘performance’. The idea of introducing new elements into a performance brings out chemical change in the context shadows the prototypical features of abstract concept (e.g., *huaxuebianhua* ‘chemical change’) and description about physical entity or abstract concept (e.g., *quanxin* ‘brand new’), as shown in Table 13. (7b) relates to the theme of *baoyang* ‘health maintenance’ and delineates the optimal

health change on skin through using a cosmetic. Thus, it involves the prototypical features of acts and behavior (e.g., *huohua* ‘activate’) and optimal change (e.g., *songchi buzai* ‘no more looseness’) (cf. Table 13). (7c) is likely to come from the theme of *dianying* ‘movie’. Its content associates with filming techniques and projected effect, and thus reflects the prototypical feature of abstract concept (e.g., *yixiang* ‘image’). What we mean by “reflect” here is the indication that a prototypical feature is observable in a given context.

Under topic **sports**, the positive SP of both *chansheng* and *niangcheng* is enhanced. The typical ten themes are *lanqiu* ‘basketball’, *titanjinkuang* ‘sports news updates’, *qiandu* ‘gambling’, *taiwanzhibang* ‘Taiwanese baseball league’, *bangqiu* ‘baseball’, *jingsai* ‘competition’, *saishi* ‘games’, *jianshen* ‘workout’, *wangqiu* ‘tennis’, and *shangyezanzhu* ‘business partnership’ (see Appendix F). Examples for each node word are provided in (8).

(8)

- a. 但搭配滑球與時速100公里出頭的慢速曲球，產生不錯速差。
 dan dapei guqiu yu shisu 100 gongli chutou de mansu quqiu, chansheng bucu
 sucha.
 but match-with slide-ball and speed-per-hour NUM kilometer a-little-more GEN
 slow curve-ball, cause good-speed-difference.
 ‘but, with a slide ball and a curved ball that is 100 kilometers per hour, it produces
 a good speed difference.’
- b. 莫瑞絲摩打法明顯進步，產生個人特色，她常適當上網，
 moruisima dafa mengxian jinbu, chansheng geren tese, chi chang didang
 shangwang,

PN play obvious improve, cause personal characteristic, 3SG.F often appropriate approach-net,

‘Morrismo has made obvious improvement in her playing and had personal characteristics. She often approaches net appropriately.’

- c. 威廉絲姊妹在廣告中打的並不是網球而是桌球，兩人產生不少特殊的笑果。

weiliansi zimei zai guanggao zhong da de bingbushi wangqiu ershi zhuoqiu, liangren chansheng bushao teshu de xiaoguo.

PN sisters at advertisement among play AUX not tennis but table-tennis, two-people produce many special GEN funny-effect.

‘The Williams sisters played table tennis in the advertisements instead of tennis. The two produced a lot of special funny effect.’

- d. 人每天平均會產生1~10萬個腫瘤細胞，正常情況下免疫系統會自動判斷細胞，

ren meitian pingjun hui chansheng 1~10 wan ge zhongliu xibao, zhengchang qingkuang xia mianyijitong hui zidong panduan xibao,

people every-day average will produce 1~10 10-thousand CLF tumor cells, normal condition under immune-system will automatic judge cells

‘On average, people will produce 1 to 100,000 tumor cells per day. Under normal circumstances, the immune system will automatically judge cells.’

- e. 接著與球迷爆發衝突，最後被安全人員護送到轉播室才沒釀成更多的意外。

jiezhe yu qiumi baofa chongtu, zuihou bei anquan renyuan husong dao zhuaibo shi cai mei niangcheng gengduo de yiwai.

then with ball-fans burst conflict, eventually PASS secure personnel escort to broadcast room only not cause more GEN accidents.

‘Then there was a conflict with the fans, and s/he was finally escorted by the security personnel to the broadcast room before it caused more accidents.’

- f. 隆利、布雷克再接力安打釀成大局。

longli, buleike zai chali anda niangcheng daju.

PN, PN again relay hit cause big-game.

‘Longli and Blake also hit, leading to a big baseball game.’

For *chansheng*, (8a) relates to either theme of *taiwanzhibang* ‘Taiwanese baseball league’ or *bangqiu* ‘baseball’. (8b) is likely to come from the theme of *wangqiu* ‘tennis’. The two instances share the idea of making an optimal change in plays leading to a good consequence or characteristic, and the idea incorporates the prototypical features of acts and behavior (e.g., *dapei* ‘match with), abstract concept (e.g., *tese* ‘characteristics’), and optimal change (e.g., *jinbu* ‘improve’), as shown in Table 15. (8c) is likely to come from the theme of *shangyezanzhu* ‘business partnership’. Its content is basically describing the humor in an advertisement, which creates a special funny effect to the audiences. The prototypical feature of attributes of physical entity/abstract concept (e.g., *teshu* ‘special’) is therefore reflected (cf. Table 15). Lastly, (8d)

associates with the theme of *jianshen* ‘workout’. It contains the idea of immune systems preventing infections, which reflects the prototypical feature of prevention (e.g., *mianyijitong* ‘immune system’) (cf. Table 15).

In terms of *niangcheng*, (8e) may belong to the theme of *saishi* ‘games’. The idea of official security guards escorting game players to avoid conflict in the text suggests the prototypical feature of settling issues (e.g., *husong* ‘escort’), as provided in Table 20. (8f) associates with either the theme of *bangqiu* ‘baseball’ or *jingsai* ‘competition’. Here, consecutive hits in an inning leads to positive results, which matches the prototypical feature of desirable description (e.g., *daju* ‘big game’).

Under topic **lifestyle**, the positive SP tendency of *niangcheng* is prompted. The typical ten themes are *shying* ‘photography’, *yiliaozhaohu* ‘medical care’, *fuzhuang* ‘cloths’, *xiuxian* ‘recreation’, *luyou* ‘travel’, *shicai* ‘ingredients’, *zhanbo* ‘diviation’, *qiche* ‘car’, *shangpintehui* ‘discount products’, *meishi* ‘food’. Examples expressing favorable meaning are provided in (9).

(9)

- a. 工作要把話說清楚，才不會釀成遺憾。
gongzuo yao ba hua shuo qingchu, cai buhui niangcheng suida.
work must hold words speak clear, so will-not cause regret
‘making a clear statement must be done at work, so it will not cause regret.’
- b. 應去看醫生治療，以免拖到最後釀成大病。
yingqu kan yisheng zhiliao, yimian tuodao zuihou niangcheng dabing.
should see doctor treatment, so-as-not-to dragged-to last cause serious-illness.

‘You should go to the doctor for treatment, so as not to have a serious illness in the end.’

- c. 葡萄釀成紅酒後，紅酒多酚也有抗氧化功效，
beitao niangcheng gongjiu hou, gongjiu duofen yeyou kangyanghua gongxiao,
grape brew wine after, wine polyphenols also have antioxidant effects,
‘After the grapes are made into red wine, the red wine polyphenols also have antioxidant effects.’
- d. 或以陳年糙米醋、薰衣草冰糖釀成舒爽的「天醋香草飲品」，以及用牛蒡和地瓜絲炸過，
huo yi chennian caomi cu, xunyicao bingtang niangcheng shushuang de 「tian cu xiangcao yinpin」, yiji yong niubang he degua sizhaguo,
or with aged brown-rice vinegar, lavender crystal-sugar brewed refreshing GEN “sky vinegar vanilla beverage,” and use burdock and sweet potato,
‘Or use the aged brown rice vinegar and lavender crystal sugar to make a refreshing "sky vinegar vanilla drink" and fried with burdock and sweet potato.’

The content of (9a) is about skill at work, which presumably leads to a desirable outcome. The most possible theme for (9a) is *shangpintehui* ‘discount products’, since it is intuitively compatible to a work-related context. Moreover, the prototypical feature of desirable description (e.g., *shuo qingchu* ‘speak clearly’) illustrated in Table 21 is reflected here. (9b) might come from the theme of *yiliaozhaohu* ‘medical care’. Its content suggests seeking for

medical care due to discomfort, involving the prototypical feature of medical manner (e.g., *zhiliao* ‘treatment’) (cf. Table 21).

In terms (9c) and (9d), we find that the other meaning of the node word—*brew*—is present. (9c) seems to come from the theme of *meishi* ‘food’, and expresses the idea of the benefit of wine polyphenols; (9d) likely comes from the theme of *shicai* ‘ingredients’ and describes how ingredients are used to make a beverage. Also, the two instances all reflect the prototypical feature of desirable description (e.g., *kangyanghua* ‘antioxidant’ and *shushuang* ‘refreshing’) shown in Table 21. It is possible that the interference from the polysemic nature of the node word might account for the phenomenon that *niangcheng* appears frequently in positive contexts.

5.2 Topics Triggering Negative SP

Under topic **society**, we find that the negative SP tendency of *chansheng* is boosted. The typical ten themes within the topic are *jiaotongshijian* ‘traffic events’, *dupinfanzui* ‘drug crime’, *xingshi* ‘crime’, *xiaoyuan* ‘school’, *zhapian* ‘fraud’, *baolishijian* ‘violent events’, *xiaofang* ‘firefighting’, *falujiufen* ‘lawsuit’, *qingseshijian* ‘sexual harassment’, and *jiatingjiufen* ‘family problem’ (see Appendix C). We provide four instances of *chansheng* in (10).

(10)

- a. 沒想到對方母親返家後，雙方產生誤解。

meixiangdao duifang muqin fangu hou, shuangfang chansheng wujie.

Unexpectedly other-side mother return-home after, both-sides cause misunderstanding.

‘Unexpectedly, after the other’s mother returned home, two people had misunderstanding.’

- b. 影響智力，甚至以後會產生經常流眼淚、抽筋的後遺症，
yingxiang zhili, shendie sihou hui chansheng jingchang liuwenlei, choujin de
housuizheng

influence intelligence, even afterward will cause often tearing, cramp GEN
sequelae,

‘Affecting intelligence, and even it will cause the sequelae of frequent tears and
cramps in the future,’

- c. 兩瓶已乾燥成顆粒狀的化學藥品，倒進水溝時，引起化學變化而產生氣爆，將他炸傷。

liang ping yi ganzao cheng kelizhuang de huaxueyaopin, daojin shuigou shi, yinqi
huaxuebianhua er chansheng qibao, jiang ta zhashang.

two bottles already dry become granular GEN chemical, pour-into ditch when,
cause chemical-change and cause gas-explosion, let 3SG.M injury.

‘When two bottles of chemicals, which have been dried into pills, was poured into
the ditch, it caused chemical changes with a gas explosion which injured him.’

- d. 身心創傷遠比異性間的侵害更大，甚至在青春期時易產生對性別混淆，建議家長隨時注意小孩言行舉止，

shenxinchuangshang yuanbi yixing jian de qinhai gengda, shenzhi zai qingchunji
 shi yi chansheng dui xingbie hunxiao, jianyi jiachang suishi zhuyi xiaohai
 yanhangjuzhi,

physical-and-mental-trauma far-more heterosexual within GEN invasion bigger,
 even in adolescence when easy cause toward gender confusion, suggest parents
 anytime watch children words-and-deeds,

‘Physical and mental trauma is far more intense than the invasion among
 heterosexual people. Even in adolescence, it is easy to cause gender confusion.
 Parents are advised to pay attention to children’s words and deeds at any time.’

The four instances unambiguously express an unfavorable evaluative meaning. (10a) is likely to come from the theme of *jiatingjiufen* ‘family problem’, and relates to the negative thought toward others, reflecting the prototypical features of unpleasant emotion, thought and mental state (e.g., *wujie* ‘misunderstanding’) (see Table 12). (10b) relates to the theme of *dupinfanzui* ‘drug crime’ and is about the consequence of taking drugs. The content involves the prototypical features of acts and behavior (e.g., *liuwenlei* ‘tearing’), and health issues (e.g., *housuizheng* ‘sequelae’) (cf. Table 12). (10c) is connected to the theme of *xiaofang* ‘firefighting’, which is often associated with its harmful events in real life. This may be closely connected to the prototypical feature of destructive/harmful/annoying events (e.g., *qibao* ‘gas-explosion’) (cf. Table 12). Lastly, (10d) may belong to the theme of *xiaoyuan* ‘school’. It concerns the change during puberty, and thus reflects the prototypical feature of change of state (e.g., *xingbie hunxiao* ‘gender confusion’) (cf. Table 12).

Under topic **international**, the three node words all together reveal an increase of negative SP, when compared to each of their negative SP counterparts in the ADN corpus. The typical ten themes are *jingxuan* ‘election’, *junshi* ‘military’, *touzimaoyi* ‘investment and trade’, *xing’an* ‘criminal case’, *shengjiyutaikong* ‘bio-technology and astronomy’, *mingren* ‘celebrity’, *wanglu* ‘internet’, *shigu* ‘accident’, *yule* ‘recreation’, and *richang* ‘daily’ (see Appendix E). Examples for each node word are presented in (11).

(11)

- a. 上海逮捕數十名遊行滋事分子，似乎已產生嚇阻作用，
 shanghai daibu shushiming youhang cishi banzi, shihu si chansheng hezhu
 zuoyong,
 LOC arrest dozen parade event-initiating people, seems PFV cause deterrent
 effect
 ‘Shanghai has arrested dozens of parade initiators and seems to cause the effect of
 deterrent’
- b. 且爆炸當晚至昨上午都吹東北風，研判該爆炸案產生的空氣污染
 物質不致影響位處上風處的金門地區空氣品質。
 qie baozha dangwan zhi zuo shangwu dou chui dongbeifeng, yanpan gai baozha
 an chansheng de kongqi wuran wuzhi buzhi yingxiang weichu shangfeng chu de
 jinmen deou kongqi pinzhi.

yet explosion that-night till yesterday morning all blow east-north-wind, judge
DET explosion case cause GEN air pollution substance will-not affect locate
upper-wind place GEN LOC area air quality.

‘Yet the northeast wind was blowing from the night of the explosion to the
morning of yesterday morning. It was determined that the air pollutants produced
by the explosion did not affect the air quality of the Kinmen area at the upper
wind.’

- c. 北韓經濟更難有起色，時間一久人民可能對金正恩政府產生不滿。

beihan jingji geng nanweiqise, shijian yijiu renmin genai dui jinzhengen zhengfu
chansheng buman.

LOC economy more difficult-to-get-better, time for-a-while people may toward
PN government cause dissatisfaction.

‘The North Korean economy is harder to be improved, and after a while the
people may be dissatisfied with the Kim Jong-un government.’

- d. 1輛火車撞上停靠在車站備用列車，釀成14死26傷的慘劇。

1 liang huochē zhuangshang tingkao zai chezhan beiyong liche, niangcheng 14 si
26 shang de canju.

1 CLF train bump-into stop at station back-up train, cause NUM death NUM
injury GEN tragedy.

‘One train crashed into the station's backup train, causing a tragedy of 14 deaths
and 26 people wounded.

- e. 並將病毒傳給其他城鎮民眾，釀成全面災難。

ban jiang bingdai chuangei jita chengtian minzhong, niangcheng quanmian zainan.

and will virus pass other town people, cause comprehensive disaster.

‘And the virus was passed on to other urban people, causing a comprehensive disaster.’

- f. 其中一具引擎吸入飛鳥，因此被迫緊急降落，衝出跑道，才會釀成意外。

qizhong yiju yinqing xiru feidiao, yinci beipai jinji jiangla, chongchu paodao, caihui niangcheng yiwai.

among one CLF engine suck-in flying-bird, thus was-forced emergent landing, rush-out runway, will cause accident.

‘One of the engines sucked in a flying bird, so they were forced to land in immediately and rushed out of the runway, causing an accident.’

- g. 可能促成「反對勢力團結，甚至促成軍方將領投靠反對勢力」。

genai chucheng 「fandui shili tuanjie, shenzhi cucheng junfang jiangling toukao fandue shili」.

May cause 「opposition force unify, even cause military leader join opposition force」.

‘It may lead to "the unification of opposition forces and even cause the military generals to join the opposition.”’

- h. 再再觸發中國敏感神經，因而促成此次試射以展現中國武裝實力。

zaizai chubo zhongguo mingan shenjing, yiner cucheng cici shishe si zhanxian zhongguo wuzhuang shili.

again-again trigger LOC sensitive nerve, thus cause this-time firing-test so-as-to show LOC armed power.

‘The sensitive nerves of China were triggered again and again, causing the country to do firing test in order to show its armed power.’

- i. 曼谷市民中，38.2%希望反對派停止集會、34.9%認為可能促成政治變化、24.3%認為會對經濟造成負面影響。

mangu shimin zhong, 38.2 % xiwang fanduipa tingzhi jihui, 34.9 % renwei keneng cucheng zhengzhi bianhua, 24.3 % renwei hui dui jingji zaocheng fumianyingxiang.

LOC citizens among, 38.2 % hope opposition stop rally, 34.9 % think may cause politics change, 24.3 % think will toward economy cause negative-impact.

‘Among the citizens of Bangkok, 38.2% of them hope that the opposition stops the rally, 34.9% of them think it may lead to political change, and 24.3% of them think it will have a negative impact on the economy.’

Beginning with *chansheng*, (11a) is connected to the theme of *xingan* ‘criminal case’ and conveys an idea that an event causes an unpleasant mental state and emotion, connoting the

prototypical features of unpleasant emotion, thought and mental states (e.g., *hezhu zuoyong* ‘detrant effect’) illustrated in Table 14. (11b) belongs to the theme of *shigu* ‘accident’, where the air pollution caused by an explosion is connected to prototypical features of health issues (e.g., *kongqi wuran* ‘air pollution’) and natural phenomenon (e.g., *dongbeifeng* ‘northeast wind’) (cf. Table 14). (11c) arises from the theme of *touzimaoyi* ‘investment and trade’. Here, espousing unpleasant opinions against a leader due economic decline is relevant to the negative change of state (e.g., *chansheng buman* ‘chause dissatisfaction’) (cf. Table 14).

In terms of *niangcheng*, (11d), (11e), and (11f) are all connected to the theme of *shigu* ‘accident’. The first two instances illustrate destructive accidents and natural disasters respectively, which are associated with the prototypical features of harmful events, accidents and disaster and natural disaster (e.g., *canju* ‘tragedy’, *zainan* ‘disaster’, and *yiwai* ‘accident’) (cf. Table 19). The third one describes in more detail about the process of accident which are relevant to the prototypical feature of acts and behavior (e.g., *jiangla* ‘landing’) (cf. Table 19).

In terms of *cucheng*, both (11g) and (11h) come from the theme of *junshi* ‘military’. (11g) is about the change between different parties in a warfare and thus associates with the prototypical feature of warfare (e.g., *chucheng junfang* ‘military leader’) and change of state (e.g., *chucheng...tuanjie* ‘cause...unify’) (see Table 23). The content of (11h) provokes the unpleasant feeling of a country, where they might incite war. Thus, the prototypical features of unpleasant emotion, thought and other mental states (e.g., *mingan shenjing* ‘sensitive nerve’) are involved (cf. Table 23). Lastly, (11i) is likely to come from the theme of *touzimaoyi* ‘investment and trade’, and is about the poll regarding termination and change of political situation and economy. This is connected to the prototypical feature of termination of events (e.g., *tingzhi jihui*

‘stops the rally’) and unpleasant emotion, thought and other mental states (e.g., *fumianyixiang* ‘negative impact’) (cf. Table 23).

5.3 Topics Triggering Positive and Negative SP

Under the topic **finance**, we find that both positive and negative SP tendencies of *chansheng* are elicited. That is, the usage of the node word is somehow preferred to describe things toward the two ends of the evaluative polarity. Again, the ten typical themes of the topic are *waibihuili* ‘foreign currency rate’, *yinhang* ‘bank’, *shehuixinxianren* ‘fresh graduate’, *baoxian* ‘insurance’, *xiaoshou* ‘sales’, *canyinbaihuo* ‘dinning and department store’, *qiyebinggou* ‘mergers and acquisition’, *fangdichan* ‘real estate’, *jinronggushi* ‘finance and stocks’, and *dianzichanpin* ‘electronic device’ (see Appendix G). Few instances conveying positive and negative meanings are provided in (12).

(12)

- a. 深度互動，產生連結，讓客戶會推薦客戶。
 shendu hudong, chansheng lanji, rang kehu hui tuijian kehu.
 deep interaction, cause connection, let customer will recommend customer.
 ‘Deep interaction, creating connection that allow customers to recommend customers’
- b. 配合 Win8 平臺推出，才能對消費者產生較大的吸引力。
 peige Win8 bengtai tuichu, cainai dui xiaobizhe chansheng jiaoda de xiyinli.
 coordinate Win8 platform launch, so to customer cause bigger GEN
 attractiveness.
 ‘With the launch of the Win8 platform, it will be more attractive to consumers.’

- c. 是最深暗或最明亮的畫面，都能確保顏色與光彩不產生偏差；

shi zui shenan huo zui mingliang de huamian, dou neng quebao yanse yu
guangcai bu chansheng piancha;

is most darkest or most brightest GEN picture, all can ensure color and brilliance
not cause deviation;

‘in the darkest or brightest picture, it can ensure that color and brilliance do not
deviate;’

- d. 隨著資金轉移至原物料股，產生排擠作用，導致電子股陷入弱勢盤
整，

suichu zijin zhuaichi die yuanwuliao gu, chansheng paiji zuoyong, dao zhi
dianzigu xianru ruoshi panzheng,

With fund transfer to original-material-stocks, cause crowding-out effect, lead
electronic-stocks fall-into weak consolidation,

‘As the funds were transferred to the original material stocks, it caused crowding
out effect, leading the electronic stocks to fall into a weak consolidation.’

- e. 財經單位的激烈作法，也可能產生負面效果。她說若投信業者拆成
多筆募集，

caijing chanwei de jilie zuofa, ye keneng chansheng fumianxiaoguo. ta shui ruo
touxin yezhe cacheng duobi muji, huo guonei touzi ren

finance-and-economics unit GEN intense practice, also may cause negative-impact. 3SG.F said if securities-investment-trust businessman split many collections,

‘The intense practice of units of finance and economics may also have negative effects. She said that if the credit agent splits into multiple collections,’

f. 但因近期國際銅價期貨大跌，市場產生未來可能降價的疑慮。

dan yin jinji guoji tongjia jihuo dadie, shichang chansheng weilai keneng jiangjia de yilu.

but because recent international copper-price futures plunge, market create future may price-reduction GEN doubt.

‘However, due to the recent plunge in international copper futures, it causes the market to doubt about possible future price cuts.’

Starting with the instances that transmit the positive SP, (12a) belongs to the theme of *xiaoshou* ‘sales’, and conveys the idea that building relationship with customers via interaction may help promote business. The prototypical feature of acts and behavior (e.g., *hudong* ‘interact’) is thus mirrored here (see Table 16). Both (12b) and (12c) are associated with the theme of *dianzichanpin* ‘electronic device’. (12b) is about the strategy of promoting a newly-released product which may attract customers’ interest. This is connected to the prototypical features of abstract concept (e.g., *xiyinli* ‘attractiveness’) and attributes of physical entity/abstract concept (e.g., *jiaoda* ‘bigger’) (cf. Table 16). (12c) relates to the balancing technique of exposure in photography, reflecting the prototypical feature of balance (e.g., *bu chansheng piancha* ‘don’t creat diviation’) (cf. Table 16).

Moving to the negative SP, (12d) comes from the theme of *jinronggushi* ‘finance and stocks’. The idea of transferring funds to one stock, leading to the decline of another stock, involves the prototypical features of acts and behavior (e.g., *zhuaichi* ‘transfer’), bad consequence (e.g., *ruoshi panzheng* ‘weak consolidation’), undesirable description about physical entity/abstract concept (e.g., *paiji zuoyong* ‘crowding out effect’), and loss (e.g., *xianru ruoshi panzheng* ‘fall into weak consolidation’) (see Table 17). (12e) associates with either the theme of *yinhang* ‘bank’, or *jinronggushi* ‘finance and stocks’. Since it concerns the practices of securities investment trust and consulting representative, the prototypical feature of acts and behavior (e.g., *zuofa* ‘practice’) is involved (cf. Table 17). (12f) is likely to come from the theme of *waibihuilu* ‘foreign currency rate’ and is about the plunge of price of futures, leading to the doubt toward future market. This is connected to the prototypical features of unpleasant emotion, thought and mental states (e.g., *yilu* ‘doubt’) and acts and behavior (e.g., *dadie* ‘plunge’) (cf. Table 17).

To sum up, the prototypical features of a node word, which resemble typical ideas underlying the contexts where a node word emerges, are one of many components that build up the overall content of a given topic. The spreading of those prototypical features in multiple themes under a given topic may be the source that intensifies the positive/negative SP tendency of a node word, and also reveals the conventionalized usage of that node word under that topic.

6. Conclusion

The purpose of the present study is to examine whether a lexical item may have a topic-dependent SP. Within the three proposed hypotheses, the strong hypothesis predicts the SP tendencies of both mixed-SP node word and strong-SP node words will be subject to topic type; the moderate hypothesis predicts only the SP tendency of the mixed-SP node word will be influenced by topic; the null hypothesis predicts no interaction between the SP tendencies of any node words and topic. In our news genre corpus (i.e., ADN), we have utilized the rule-based concordance line analysis to find out the SP tendencies of each node word under different topics and examined the relation between TOPIC and SEMANTIC PROSODY via chi-square test. Based on the results, we have concluded that topic has a moderately strong effect on the SP of a node word which has a mixed SP in general-domain genre (general); however, it has a weak effect on the SP of a node word with a strong positive/negative SP in general. Therefore, our results support the moderate hypothesis. We further propose the notion of *topic prosody*, which is at the lower level of *register prosody*. Moreover, we have applied the semantic network analysis in order to discover the semantic features of the prototypical collocates of a node word under certain topics. These semantic features may explain the rising of positive/negative SP tendency under those topics.

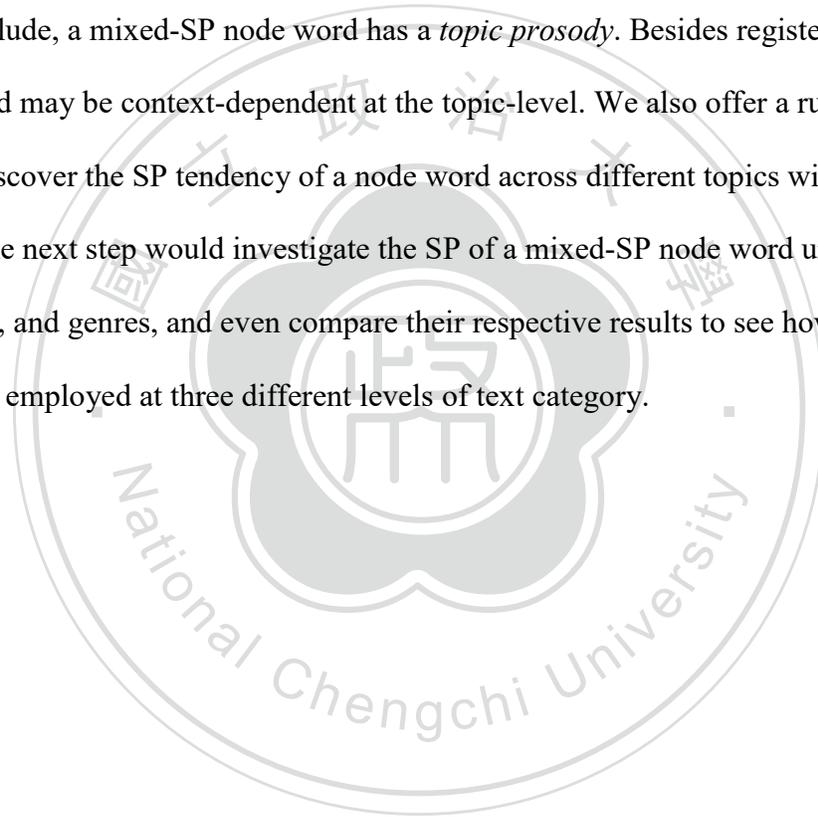
The notion of *topic prosody* suggests that topic has effect on the SP tendency of a node word. However, as noted before, only a node word with a mixed SP in general may showcase *topic prosody*. Compared to *register prosody* where a node word may have a positive SP in one register but a negative SP in another, *topic prosody* indicates the observable and significant change of the SP tendency of a node under one topic with respect to the overall SP tendency of that node word across different topics within a corpus. *Topic prosody* also implies that topic may

be one of the factors that contributes to the mixture of the SP of a node word, such as *chansheng* ‘cause’ and *dailai* ‘bring about’ reported in the study of Xiao and McEnery (2006). On the other hand, a node word with the property of *topic prosody* may indicate such word functions as a sentiment resonator under different topics. Thus, we conclude that the SP of a given node word may be modulated by different text categorizations at the level of either register (Hunston, 2007; O'Halloran, 2007) or topic (cf. Partington, 2017).

There are some limitations in our study. First, the arbitrary decision of one chunk before and after a node word (1:1 chunk-based window size) as a span for a concordance is uncommon comparing to the concordance line analysis in previous studies. Meanwhile, the chunk unit, based on punctuation and symbols, is not necessarily valid. Second, the reference sentiment dictionary (i.e., ANTUSD) has limited entries of words and thus may not provide enough evaluative information for the sentiment determination of some concordances. Third, the way to automatically determine the sentiment of the concordances of each node word is not optimal. The rule-based method may not be able to account for all the instances of concordances due to the variability of Chinese, leading to erroneous sentiment classification. Fourth, in the network analysis, the model of word embedding, i.e., GloVe, may need to be validated for its psycholinguistic importance. Fifth, as seen in the examples of *niangcheng* under topic **lifestyle**, we are not able to single out the target meaning, *cause*, from the node word. The involvement of different senses of the node word in our analysis data might influence the credibility of the SP distribution of that node word under each topic. Also, it was noted that different senses of a polysemous word may have different SPs (Bednarek, 2008; Louw, 1993). Future study on the relation between TOPIC and SEMANTIC PROSODY needs to address on issue of polysemy by applying word sense disambiguation approach so that the “noises” from other senses of the word

may be reduced. Sixth, since we did not use the sentiment analysis to evaluate the positive/negative tendency of each article of a topic, the overall sentiment trend of that topic is still unclear. To exactly know the overall sentiment disposition of a topic may give us a much clearer picture regarding the connection of the topic to the SP of a node word. Thus, future study can employ sentiment analysis to discover the overall sentiment tendency of a topic and examine its link to the typical contexts a node word emerges.

To conclude, a mixed-SP node word has a *topic prosody*. Besides register, the SP of a given node word may be context-dependent at the topic-level. We also offer a rule-based method to efficiently discover the SP tendency of a node word across different topics within a corpus. We hope that the next step would investigate the SP of a mixed-SP node word under different topics, registers, and genres, and even compare their respective results to see how such a word may be flexibly employed at three different levels of text category.



References

- Almende, B.V., Benoit, Thieurmel, & Titouan, Robert. (2018). visNetwork: Network visualization using 'vis. js' library. R package version 2.0.5. Retrieved from <https://CRAN.R-project.org/package=visNetwork>
- Bednarek, Monika. (2008). Semantic preference and semantic prosody re-examined. *Corpus Linguistics and Linguistic Theory*, 4, 119-139. doi:10.1515/CLLT.2008.006
- Blei, David M, Ng, Andrew Y, & Jordan, Michael I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3, 993-1022. doi:10.1162/jmlr.2003.3.4-5.993
- Cohen, Stanley, & Young, Jock. (1981). *The manufacture of news: Social problems, deviance and the mass media*. Newbury Park, CA: Sage Pubns.
- Ellis, Nick C, & Ogden, Dave C. (2017). Thinking about multiword constructions: Usage-based approaches to acquisition and processing. *Topics in Cognitive Science*, 9, 604-620. doi:10.1111/tops.12256
- Ellis, Nick C, Römer, Ute, & O'Donnell, Matthew Brook. (2016). *Usage-based approaches to language acquisition and processing: Cognitive and corpus investigations of construction grammar*. Hoboken, NJ: John Wiley & Sons Limited.
- Evert, Stefan. (2008). Corpora and collocations. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics. An international handbook* (Vol. 2, pp. 1212-1248). Berlin, Germany: de Gruyter.
- Freeman, Linton C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, 35-41. doi:10.2307/3033543

- Gablasova, Dana , Brezina , Vaclav, & McEnery, Tony. (2017). Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence. *Language Learning*, 67, 155-179. doi:10.1111/lang.12225
- Gries, Stefan Th. (2013). 50-something years of work on collocations. *International Journal of Corpus Linguistics*, 18, 137-166. doi:10.1075/ijcl.18.1.09gri
- Griffiths, Thomas L, & Steyvers, Mark. (2004). Finding scientific topics. *Proceedings of the National academy of Sciences*, 101, 5228-5235. doi:10.1073/pnas.0307752101
- Halliday, M.A.K., & Hasan, Ruqaiya. (1989). *Language, context, and text: Aspects of language in a social-semiotic perspective*. Oxford, UK: Oxford University Press.
- Hoey, Michael. (1991). *Pattern of lexis in text*. Oxford, UK: Oxford University Press.
- Huang, Chu-Ren, & Hsieh, Shu-Kai. (2010). Infrastructure for cross-lingual knowledge representation-towards multilingualism in linguistic studies. *Taiwan NSC-granted Research Project (NSC 96-2411-H-003-061-MY3)*.
- Huang, Ting-Hao, Chen, Yun-Nung, & Kong, Lingpeng. (2015). *Acbima: Advanced Chinese bi-character word morphological analyzer*. Paper presented at the Proceedings of the Eighth SIGHAN Workshop on Chinese Language Processing.
- Huang, Ting-Hao, Ku, Lun-Wei, & Chen, Hsin-Hsi. (2010). *Predicting morphological types of Chinese bi-character words by machine learning approaches*. Paper presented at the Proceedings of LREC.
- Hunston, Susan. (1995). A corpus study of some English verbs of attribution. *Functions of Language*, 2, 133-158. doi:10.1075/fol.2.2.02hun

- Hunston, Susan. (2004). Counting the uncountable: problems of identifying evaluation in a text and in a corpus. In A. Partington, J. Morley, & L. Haarman (Eds.), *Corpora and Discourse* (Vol. 9, pp. 157-188). Bern, CH: Peter Lang.
- Hunston, Susan. (2007). Semantic prosody revisited. *International Journal of Corpus Linguistics*, 12, 249-268. doi:10.1075/ijcl.12.2.09hun
- Khedr, Ayman E, Salama, S.E., & Yaseen, Nagwa. (2017). Predicting stock market behavior using data mining technique and news sentiment analysis. *International Journal of Intelligent Systems and Applications*, 9, 22. doi:10.5815/ijisa.2017.07.03
- Krestel, Ralf, Fankhauser, Peter, & Nejd, Wolfgang. (2009). *Latent dirichlet allocation for tag recommendation*. Paper presented at the Proceedings of the third ACM conference on Recommender systems.
- Ku, Lun-Wei, Liang, Yu-Ting, & Chen, Hsin-Hsi. (2006). *Opinion extraction, summarization and tracking in news and blog corpora*. Paper presented at the Proceedings of AAAI.
- Li, Meixia, & Jiao, Aihui. (2017). The study of semantic prosody of Chinese logical resultative formulae: A corpus-assisted discourse analysis approach. *GSTF Journal of Law and Social Sciences (JLSS)*, 2.
- Louw, Bill. (1993). Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. *Text and technology: In honour of John Sinclair*, 240, 251.
- Louw, Bill. (2000). Contextual prosodic theory: Bringing semantic prosodies to life. *Words in context: A tribute to John Sinclair on his retirement*, 48-94.
- Manning, Christopher D., & Schütze, Hinrich. (1999). *Foundations of statistical natural language processing*. Cambridge, MA: MIT Press.

- Morley, John, & Partington, Alan. (2009). A few frequently asked questions about semantic—or evaluative—prosody. *International Journal of Corpus Linguistics*, 14, 139-158. doi:10.1075/ijcl.14.2.01mor
- O'Halloran, Kieran. (2007). Critical discourse analysis and the corpus-informed interpretation of metaphor at the register level. *Applied Linguistics*, 28, 1-24. doi:10.1093/applin/aml046
- Pang, Bo, & Lee, Lillian. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2, 1-135. doi:10.1561/15000000011
- Partington, Alan. (1998). *Patterns and meanings: Using corpora for English language research and teaching* (Vol. 2). Amsterdam and Philadelphia: John Benjamins.
- Partington, Alan. (2004). "Utterly content in each other's company": Semantic prosody and semantic preference. *International Journal of Corpus Linguistics*, 9, 131-156. doi:10.1075/ijcl.9.1.07par
- Partington, Alan. (2017). Evaluative clash, evaluative cohesion and how we actually read evaluation in texts. *Journal of Pragmatics*, 117, 190-203. doi:10.1016/j.pragma.2017.06.008
- Pennington, Jeffrey, Socher, Richard, & Manning, Christopher. (2014). *Glove: Global vectors for word representation*. Paper presented at the Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP).
- Rosch, Eleanor. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology. General*, 104, 192. doi:10.1037/0096-3445.104.3.192
- Sinclair, John. (1991). *Corpus, concordance, collocation*. London, UK: Oxford University Press.
- Sinclair, John. (2004). The search for units of meaning *Trust the Text* (pp. 34-58). London, UK: Routledge.

- Stefanowitsch, Anatol, & Gries, Stefan Th. (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, 8, 209-243. doi:10.1075/ijcl.8.2.03ste
- Stubbs, Michael. (1995). Collocations and semantic profiles: On the cause of the trouble with quantitative studies. *Functions of Language*, 2, 23-55. doi:10.1075/fol.2.1.03stu
- Stubbs, Michael. (2001a). On inference theories and code theories: Corpus evidence for semantic schemas. *Text - Interdisciplinary Journal for the Study of Discourse*, 21, 437-465. doi:10.1515/text.2001.007
- Stubbs, Michael. (2001b). *Words and phrases: Corpus studies of lexical semantics*. Oxford, UK: Blackwell.
- Wang, Shih-Ming, & Ku, Lun-Wei. (2016). *ANTUSD: A large Chinese sentiment dictionary*. Paper presented at the LREC.
- Wei, Naixing, & Li, Xiaohong. (2014). Exploring semantic preference and semantic prosody across English and Chinese: Their roles for cross-linguistic equivalence. *Corpus Linguistics and Linguistic Theory*, 10, 103-138. doi:10.1515/cllt-2013-0018
- Xiao, Richard, & McEnery, Tony. (2006). Collocation, semantic prosody, and near synonymy: A cross-linguistic perspective. *Applied Linguistics*, 27, 103-129. doi:10.1093/applin/ami045

Appendix

Appendix A

List of negators

未 *wei* 'not'

並未 *bingwei* 'not'

並不會 *bingbuhui* 'not'

沒有 *meiyou* 'no'

不 *bu* 'no'

沒 *mei* 'no'

不會 *buhui* 'cannot'

不用 *buyong* 'no need to'

也不會 *yebuhui* 'not'

無法 *wufa* 'cannot'

Appendix B

List of PREVENTION words

以免 *yimian* 'so as not to'

避免 *bimian* 'avoid'

以避免 *yibimian* 'to avoid'

免得 *miande* 'lest'

預防 *yufan* 'prevent'

防止 *fanzhi* 'prevent'

以防 *yifan* 'prevent'



Appendix C

Top 15 relevant words under topic society

交通 事件	毒品 犯罪	刑事	校園	詐騙	暴力 事件	消防	法律 糾紛	情色 事件	家庭 糾紛
駕駛	查獲	歹徒	學生	詐騙	砍	火勢	檢方	少女	自殺
肇事	販售	竊盜	女童	公司	刀	起火	賭客	女子	死者
撞上	毒品	嫌犯	男童	帳戶	談判	消防局	檢察官	猥褻	妻子
車禍	漁船	強盜	學校	投資	行兇	火警	議員	下體	狗
貨車	走私	行竊	老師	銀行	檳榔	濃煙	法官	侵	丈夫
車道	販毒	竊賊	同學	匯款	被打	火場	判決	賣淫	遺書
行經	查緝	犯案	校方	詐欺	砸	撲滅	處分	性交	兒子
司機	破獲	搶	家長	獎金	衝突	爆炸	法院	狼	燒炭
乘客	販賣	皮包	輔導	會員	攻擊	逃生	立委	性交易	醫師
轎車	海洛因	搶劫	孩子	金額	口角	住戶	裁定	偷拍	專線
煞車	船長	搶奪	校長	股票	圍毆	消防人 員	死刑	強姦	症
騎士	偽鈔	搶走	國小	被騙	砍殺	搜救	司法	自主	生命線
行駛	漁工	偷	就讀	人頭	打人	消防隊	羈押	援交	女嬰
撞	保育	持槍	校園	受騙	西施	大火	賭場	裸照	男嬰
路段	搖頭丸	贓車	導師	廠商	菜刀	消防	上訴	網友	死因

Appendix D

Top 15 relevant words under topic entertainment

精品	表演	法律 糾紛	保養	生活 娛樂	醫療	感情 八卦	影視	家庭	電影
設計	演唱會	離婚	肌膚	飾 值得一	醫院	婚禮	節目	媽媽	最佳
品牌	歌迷	律師	保養	看	治療	戀情	主持	女兒	電影
時尚	專輯	告	下載	圖	醫生	交往	豬哥	爸爸	獎
設計師	唱	法院	眼霜	人類	罹患	男友	新聞	兒子	主演
搭配	演唱	豪宅	內衣	料理	症	女友	華視	小孩	入圍
材質	音樂	官司	產品	客廳	住院	分手	綜藝	買	上映
表	開唱	聲明	保溼	恐怖	檢查	約會	霞	爸媽	日劇
服裝	唱片	吸毒	免費	動物	開刀	婚宴	佼佼	懷孕	票房
腕錶	巨蛋	指出	細紋	地理	車禍	婚紗	陶子	老婆	影展
風格	歌	法律	眼袋	場景	病情	緋聞	中視	孩子	得獎
攝影	歌曲	法官	暗沉	景點	病房	目擊	亮	胎	飾演
優雅	新專輯	指控	詳情請	刺青	手術	特勤	罵	東西	主角
款式	粉絲	大麻	精華	基哥	受傷	求婚	林佑威	媽	劇
配件	舞臺	涉嫌	美白	蠟像	憂鬱	記者	臺視	婆婆	導演
皮革	門票	財產	皮膚	地球	送醫	結婚	胡瓜	送	執導

Appendix E

Top 15 relevant words under topic international

競選	軍事	投資 貿易	刑案	生技 與太 空	名人	網路	事故	娛樂	日常
選舉	美軍	企業	警方	科學家	王子	例	爆炸	圖	污染
黨	飛彈	公司	男子	太空	女王	網友	乘客	動物園	食品
民主	軍事	集團	嫌犯	研究	婚姻	意思	地震	答案	環保
投票	核武	臺商	女子	感染	王室	好	受傷	表演	餐廳
國會	會談	投資	報警	手術	女性	看	機場	動物	火山
候選人	基地	臺幣	歹徒	病毒	女星	短片	事故	遊戲規	食物
總統	制裁	銀行	殺人	患者	婚禮	覺得	炸彈	則	興建
大選	軍	人民幣	男童	飛行	蜜雪兒	收聽	救援	比賽	鑽石
議員	空襲	員工	逮捕	地球	畢業	動詞	造成	貓	牛肉
選民	部隊	用戶	謀殺	火星	結婚	請	班機	紙	建築
蕊	談判	董事長	女童	基因	獎	去	列車	狗	不同之
當選	部署	微軟	少女	人類	電影	例如	搜救	處	水庫
示威	主權	金額	殺害	太空人	精子	瀏覽	發生	眼力	排放
民調	戰爭	駭客	兒子	病患	年齡	教學	受困	考考	完工
競選	戰略	資產	鄰居	醫師	繼承	想	火車	展出	豪華
								選手	攝取

Appendix F

Top 15 relevant words under topic sports

籃球	體壇 近況	簽賭	台灣 職棒	棒球	競賽	賽事	健身	網球	商業 贊助
籃板	春訓	客	不好	安打	奧運	回合	動作	強	球迷
助攻	球團	來源	變化	局	車手	洞	身材	網球	老婆
防守	復健	資料	有點	全壘打	車隊	標準	肌肉	種子	簽名
火箭	名單	賠率	球路	打點	金牌	獎金	伸展	女單	球衣
熱火	選秀	結算	高志	支	項目	低於	雙手	單打	支持
進攻	月薪	薪資	好	敲出	賽道	推杆	舞蹈	男單	轉播
騎士	手術	賭金	嘟嘟	敲	賽車	高球	有氧	法網	廣告
命中率	報到	擂臺	感覺	擊出	跆拳道	巡迴賽	瑜伽	進球	事件
中鋒	登錄	年薪	不錯	三振	亞運	名人賽	公分	皇馬	商品
小牛	簽約	蘋果	揮棒	率	銀牌	首回合	身體	直落	兒子
得分	韌帶	加封	球速	二壘	奪金	球道	功能	溫網	現場
禁區	經典	虛擬	變速	轟	站	體育臺	雙腳	健全	媒體
						高於標			
啤	旅外	為準	綱	保送	銅牌	準	體重	晉級	代言
活塞	人選	上限	致遠	失分	全國	碼	身高	首盤	加油
外線	釋出	受邀	很好	打擊	獎牌	高球賽	三圍	女網賽	女兒

Appendix G

Top 15 relevant words under topic finance

外幣 匯率	銀行	社會 新鮮 人	保險	銷售	餐飲 百貨	企業 併購	房地 產	金融 股市	電子 產品
經濟 經濟成	銀行	運	基金	產能	百貨	股東	土地	股價	手機
長	銀	創業	保單	純益	統一	董事	建設	外資	電腦
升息	信用卡	老闆	人壽	毛利率	店	股權	房價	個股	智慧
美元	分行	客人	保費	出貨	餐飲	董事會	豪宅	大盤	數位
日圓	貸款	工作	壽險	單季	飯店	員工	大樓	金融股	宏碁
央行	卡	金錢	保險	庫存	遊戲	增資	建築	反彈	平板
兌	存款	朋友	險	去年同	飲料	席	住宅	買超	微軟
通膨	刷卡	收藏	保障	盈餘	門市	案	工程	題材	機種
寬鬆	房貸	貼心	投保	面板	開幕	私募	園區	成交量	蘋果
升值	放款	覺得	報酬	調漲	通路	股東會	新藥	行情	電視
經濟學									
家	臺銀	攝影	保戶	定單	精品	購併	發電	回檔	相機
出口	行庫	畢業	型基金	單月	展	改選	興建	籌碼	螢幕
貨幣	優惠	吃	理賠	淡季	購物	雙方	環保	整理	裝置
匯率	手續費	喜歡	壽險業	報價	連鎖	認購	臺電	買盤	軟體
聯準	髮卡	媽媽	保險公 司	利用率	販	收購	房屋	基本面	處理器

Appendix H

Top 15 relevant words under topic lifestyle

攝影	醫療 照護	服裝	休閒	旅遊	食材	占卜	汽車	商品 特惠	美食
眼	患者	洋裝	蘋果花	建築	超市	愛情	遊戲	即日起	滋味
相機	治療	褲	週刊	公園	材料	指數	馬力	商品	無休
手機	症狀	衣	月餅	步道	頂	感情	玩家	折起	口味
眼線	成分	鞋	團	農場	平均	開運	車款	優惠	香氣
			吃喝玩						
記憶體	皮膚	背心	樂	搭乘	調味料	對方	車型	特賣	吃
處理器	肌膚	襯衫	媽媽	觀光	鹽	孩子	系統	折	料理
影	保溼	項鍊	網友	抵達	大匙	事情	扭力	特價	菜色
睫毛	藥物	T恤	評審	園區	一般	占卜	動力	現金	醬汁
彩妝師	過敏	裙	糉子	客運	愛	投資	原廠	下殺	開出
鏡頭	身體	上衣	社團	臺幣	備用	運勢	變速箱	折扣	味道
功能	改善	剪裁	生日	民宿	作法	財運	新車	售完	吃來
剪刀	肌肉	針織	更多	前往	拌勻	事業	全球	好康	品嚐
機身	檢查	印花	樣樣	航空	適量	命理	性能	為止	餐廳
眼窩	感染	短褲	先生	山	小匙	運	最大	消費	濃郁
眼	患者	洋裝	蘋果花	建築	超市	愛情	遊戲	即日起	滋味