

Chinese Negation Morpheme *bù* in Academic Writing*

Chen-Li Kuo (Taoyuan)

Abstract

Negation can be conveyed through various forms in a language. In Chinese, *bù* is the most common negative form, and it can also be a morpheme of a compound word which expresses negation and/or concepts closely associated with negation, e.g. modality and contrast. In order to capture most types of Chinese words carrying these concepts, we investigate Chinese words containing *bù*. This study first classifies *bù*-related words into seven categories according to the word structure and the negative connotation of *bù* as a morpheme retained in the word. The categorisation is then applied to the wordlists of a research article corpus and a conversational corpus. The results show that there are many fewer occurrences of *bù* as a negator in research articles than in conversation, and that there is greater use of four-syllable Chinese idioms (*chéngyǔ*) in research articles than in conversation. In addition, different sets of *bù*-related words are used to express contrast and modality.

1 Introduction

The functions of negation are diverse. In the genres of academic writing, negation can be used as a form of rhetorical device for indirectness (Hinkel 1997), for increasing the level of contradiction between the proposition and the expectation (Jordan 1998), and sometimes to express negative assertions related to factual knowledge (Vincze et al. 2008). In addition, negation is often used to express denial of a specific idea existing in the target academic community (Pagano 1994: 259). As exemplified by Pagano (ibid.), Halliday (1985: xvii) stressed “A text is a semantic unit, not a grammatical one” for pointing out the misconception held by his readers about what “a text” really is. Halliday presupposed that the majority of the academic community he was writing for would assume a text is a grammatical unit. Without having the knowledge of this assumption, the denial will not be fully understood.

As pointed out by Jordan (1998), negative forms provide “great textual and contextual significance” which cannot be conveyed by positive forms. The automatic retrieval of factual knowledge from scientific articles is a challenging task when negation is involved, owing to its diversity of form and function. Previous studies have attempted to understand negation from different points of view. For example, Tottie (1991) studied variation across speech and writing of using negation in English; and Horn (2001) included the views of philosophy and psychology in the discussion of negation in natural languages. Negation can be treated as an

* This research is partially supported by Ministry of Science and Technology, Taiwan, R.O.C. under Grant no. NSC 101-2410-H-182-031.

important feature in various tasks in natural language processing, such as sentiment analysis (Taboada et al. 2011). The Computational Linguistics Journal published a special issue on modality and negation in 2012, in order to draw the attention of the computational linguistics community to extra-propositional aspects in natural languages (Morante/Sporleder 2012). Most of the previous studies regarding negation are on English and only a few focus on academic writing, despite the fact that the use of negation, which could be closely associated with concepts such as hedging and uncertainty, are certainly important to academic writing. The formal writing style expected by the academic community could also affect word choice in situations where negation is required.

The present study investigates the distribution of the most common Chinese negation morpheme *bù* in academic articles written by Taiwanese scholars who speak Taiwan Mandarin as their first language. The results will be compared with the Chinese Spoken Wordlist¹ compiled by Tseng (2013) using the Taiwan Mandarin Conversational corpus (TMC), which consists of 85 unscripted conversations. TMC could be seen as a snapshot of contemporary colloquial language used in Taiwan while the research articles are characterised by high level of formality.

Academic writing in Chinese speaking community exhibits some similarities and differences with the English counterparts. At the discourse level, similar rhetorical moves, i.e. particular rhetorical stages or structures, can be found in both English and Chinese research articles (Loi and Evans 2010). In addition, research articles in both languages feature explicitness, specifying the value of the research, and taking a critical stance. However, Chinese authors tend to present the above features to a lesser degree than English authors (ibid.). For example, there are fewer Chinese research articles containing the move “indicating a gap”, which is a great chance for the author to take a critical stance (ibid.: 2818). Similar to English scholars, Chinese authors also use hedges to be polite or to show less commitment to the truth of the propositions. Hu and Cao (2011) found that English-medium journals feature more hedges than Chinese-medium journals. In contrast, while boosters and intensifiers are used in both languages, they occur more often in articles written by Chinese authors than in articles written by English authors (Hinkel 2005). Hyland’s (1998) study showed that there are more hedges than boosters in English research articles: 14.6 hedges per thousand words verses 5.88 boosters per thousand words. Hu and Cao’s (2011) study also revealed that English authors use significantly more hedges than boosters while Chinese articles contain more boosters than hedges in average. Hinkel (2005) argued that while exaggeration and overstatement are restrained in Anglo-American academic writing, therefore the use of boosters is confined, Confucius rhetorical tradition actually encourage this type of language devices in persuasive writing. While deliberative debates are important in western rhetoric, Confucius concerned the means of “influencing people’s behaviour and moving them to action through exemplary conduct rather than through speeches” (Ding 2007: 150). Consequently, advocating a certain idea has replaced debating and deliberating to a certain extent in Chinese research articles. Since the Chinese author “is assumed to have authority, credibility, and knowledge”, as noted

¹ The wordlist can be downloaded from Tseng’s page: www.ling.sinica.edu.tw/v3-3-1_en.asp-auserid=20.htm [17.02.2016].

by Hinkel (1999), excess use of hedges to convey cautiousness toward the statements appears to be unnecessary.

The objective of this research is to investigate the differences in using *bù*-related words between professional academic writing and much less formal spontaneous spoken Chinese. The contrast between the two might be interesting, since unscripted spoken language and academic writing are from the two ends of the spectrum in terms of production opportunities and formality: spoken language is the most common form of language production while research articles are written by only a small number of native speakers; the level of formality expected from academic writing is high, and the spoken language is often less formal than the written language. The findings can be converted to pedagogical guidelines for both native and foreign students who intend to learn academic writing in Chinese, for example, suggestions on what types of linguistic devices are accepted in colloquial language but less appropriate in academic writing. The contrast between a wordlist generated from unscripted conversations and a wordlist derived research articles is similar to the contrast between academic vocabulary and high frequent words distinguished by scholars like Coxhead (2000) and Nation (2001) who argued that academic vocabulary is difficult for language learners across academic disciplines (see also Rheindorf in this special issue). We hope to compare and contrast the two wordlists and suggest the type of vocabulary needed to be learned in order to master academic writing in Chinese. Through the understanding of the use of the common negation marker in this particular genre, research articles, the results may also serve as the foundation for developing annotation guidelines required for analysing academic texts and automatic extraction of factual knowledge.

In the next two sections, corpus-based studies relevant to negation, and the function of *bù* will be reviewed. In section 4, the methods and materials used will be discussed. The results will be presented and discussed in section 5.

2 Negation in the academic corpus

Biber et al. (1999) studied the actual use of grammatical features in different registers of English based on the Longman Spoken and Written English corpus (LSWE), with an aim of providing “full descriptions of both the structure and the use of grammatical features in English” (ibid.: 5). Four of the core registers composing the LSWE corpus are conversation, fiction, news and academic prose, and negation is one of the grammatical features being investigated in this study. The academic prose includes book extracts and research articles across 15 disciplines. The findings from the corpus show that negation is much more common in conversation than in writing, and that academic prose has the lowest number of negation occurrences (ibid.:159). Biber et al. attributed the results to the fact that negation is often tied to the verb of a sentence, and conversation has the highest frequency of verbs and therefore a high frequency of negative forms. Another possible explanation is that the study did not take into account forms of negation other than *not*-negation and *no*-negation (including *nobody*, *none*, *nothing*, *nowhere*, *never*, *neither*, and *nor*), and in academic articles, negative connotation might be marked by other negative forms. However, the results echo the writing

guidelines provided by professional websites. For example, YourDictionary² suggests avoiding negative words, such as *cannot*, *do not* and *not*, in academic and business writing. The University of Jyväskylä³ suggests using *little* and *few* to replace *not much* and *not many* in academic writing. Gillett (2015) suggests in UefAP.com⁴ that adding affixes to existing words is common in academic English. By adding negative prefixes such as *un-*, *im-/in-/ir-/il-*, *dis-* or *non-* to words, the chances of using the negative word *not* will be reduced. Frequencies of each variant for negative forms are also different. For example, negating a clause without *do* and with *need*, as in sentence (2.1), is common in academic prose, but rare in conversation.

(2.1) *The details need not concern us here.*

Herrero-Zorita (2013) investigated eight forms of negation (four *not*-negation and four *no*-negation) in the British Academic Spoken English corpus (BASE) and the British Academic Written English corpus (BAWE). BASE contains university lectures and BAWE consists of university student assignments across four disciplines: Arts and Humanities, Social Sciences, Life Sciences and Physical Science. The results also showed there were consistently more usages of negation in spoken English than in written English across the eight negative forms. Only a very limited set of negative forms has been investigated in previous studies. Within this limited set, researchers found that negative words are used differently in spoken English and in academic writing in two aspects: the number of occurrences, and the variety of negative forms. Academic writing tends to include a larger variety of negative forms and a smaller number of negators when compared with the spoken counterpart.

3 *Bù*-related words

Bù 不 is the most general and neutral form of negation among the four negative forms, i. e. *bù* 不, *bié* 別, *méi* 沒, and *méiyǒu* 沒有, that are used to negate verb phrases in Mandarin (Li/Thompson 1981: 415). *Bù* is general and neutral in the sense that it does not convey extra semantic or pragmatic meanings to the utterances like *bié* signalling warnings. In addition, *bù* can be used in a wide range of syntactic occasions, while *bié* is mainly used in imperatives and *méi/méiyǒu* is used to negate the existence of an event or an entity in the past or at present, but not in the future. *Wú* 無 and *fēi* 非 are also negation markers (Ross/Ma 2006) but both are less common than *bù*. *Wú* is often used as a morpheme to signify the sense of without or suffix *-less* in a noun. *Fēi* can be used to negate nouns or to replace *bùshì* 不是 to negate grammatical constituents other than verb phrases. In addition to serving as a standalone negative word like *not* or with another morpheme to amplify the strength of negation like *juébù* 絕不 ‘never/definitely not’, *bù* can be part of a compound word in which the negative

² YourDictionary (n. d.). Negative Words to Avoid in Writing. <http://grammar.yourdictionary.com/style-and-usage/negative-words-to-avoid-in-writing.html> [17.02.2016].

³ University of Jyväskylä (n. d.). Language for Academic Writing. www.jyu.fi/hum/laitokset/kielet/oppiaineet_kls/englanti/studies/thesis-and-academic-writing/language-for-academic-writing/Language%20for%20Academic%20Writing [17.02.2016].

⁴ Gillett, Andy (2015). “Academic writing: Features of academic writing”. In Gillett, Andy (ed.): Using English for Academic Purposes: A Guide for Students in Higher Education. www.uefap.com/writing/feature/complex_lexcomp.htm [17.02.2016].

connotation may be preserved, although the scope of negation is not clausal. Some of the compound words act like the non-standard negation described in van der Auwera (2010). Li and Thompson (1981: 46–48) argued that the degree of relatedness between the meaning of a compound word and the meaning of its component morphemes can vary from no apparent semantic connection to directly related or identical meanings shared between the word and the combination of its parts. Only very few compound words in Mandarin have little or no semantic connection with their parts. Nonetheless, the connections between the parts and the whole are not always transparent. The following passages examine the morphological and semantic roles of *bù* in *bù*-related words.

3.1 Contrastive conjunctions

Several Chinese conjunctions contain the morpheme *bù*, and these conjunctions often signify contrast. For example, *búguò*⁵ 不過 means ‘but/on the other hand’; *bùrán* 不然 as a conjunction functions as ‘otherwise’. *Bù* seems to have lost the nature of negation in these words. However, if the closeness of two seemingly different concepts, i. e. negation and contrast, is recognised as in Biber et al. (1999: 82), we can say that the negative function of *bù* is somehow preserved in these conjunctions.

3.2 Negative prefix

Unsurprisingly, *bù* often functions as a prefix which negates the following morpheme, for example *bùtóng* 不同 ‘different’ literally means ‘not-same’. Giora (2006) boldly argued that in most cases “anything negatives can do affirmatives can do just as well”. Giora referred to the possible replacement relationship between negation and affirmation at the pragmatic level. Nevertheless, if we examine the Chinese compound words starting with the negator *bù*, it is possible to find synonyms that do not contain explicit negation markers. For example, *bùtóng* 不同 is arguably a synonym of *yì* 異 or *xiāngyì* 相異 (lit. ‘mutually-different’): the forms of the two words appear to be very different and might not always be interchangeable. *Xiāngyì* is much less frequent: it does not appear in the Taiwan Mandarin Conversation corpus (TMC), which contains 405,435 words and 42 hours of speech recordings (Tseng 2013: 4). It will be interesting to investigate whether there is a tendency to avoid words containing *bù* in Chinese academic writing, as writers of English academic articles reduce the use of *not*-negation and *no*-negation.

Some words in this group express modality, for example *bùfáng* 不妨 ‘no harm to’, *bùbì* 不必 ‘need not’, *bùyòng* 不用 ‘need not’, and *bùjiàndé* 不見得 ‘not necessarily’. As Hsieh (2003) pointed out, *bù* in these words is either bounded with other morphemes or will be unable to express modality if it is separate from other morphemes in the words.

⁵ The tone of *bu* changes to rising if the following morpheme carries a falling tone.

3.3 Potential infix

Bù can be an infix in resultative verb compounds as in (3.1). *Xiěchū* 寫出 is a resultative verb compound (RVC) where “the second element signals some results of the action or process conveyed by the first element” (Li/Thompson 1981: 54–55). The first part in this compound is the action *xiě* 寫 ‘write’, and the second part is *chū* 出 ‘out’. *Bù* 不 or *dé* 得 (3.2) can be used as potential infixes of RVC to signal whether the result is possible (Li/Thompson 1981: 38–39). In this case, *bù* acts as a negator.

(3.1) *Xiě-bù-chū* 寫不出: write-not-out = cannot write down anything

(3.2) *Xiě-dé-chū* 寫得出: write-obtain-out = can write down something

3.4 A-not-A pattern

A way of forming a Chinese disjunctive question is to combine the affirmative predicate and the negative counterpart in one sentence. (3.3) is an example.

(3.3) 她 漂不漂亮?

Tā piào-bù-piàoliang

She pretty-not-pretty

Is she pretty?

This type of question is called A-not-A (Ernst 1994: 241). While an A-not-A constituent is mainly used in questions, it could also be used in a clause with functions similar to the *whether*-leading clause in English.

3.5 Double negative

A compound word can contain two negative morphemes in Chinese. For example, *wúbù* 無不 ‘all/without exception’ literally means ‘no-not’; *bùdébù* 不得不 ‘have to’ literally means ‘not-ought-not’. The two negative elements in one compound word cancel each other out and resolve to a positive meaning with an emphatic effect.

3.6 Idiomatic and semantically less connected compounds

In this group of compound words, it is difficult to derive the meaning of a word from its constituents. (3.4) is an example. This type of four-syllable idiomatic compounds, *chéngyǔ* 成語 (lit. ‘set phrase’), are very common in spoken and written Chinese. Many of them can be traced back to classical Chinese literature. In the formal education system in Chinese speaking regions, students are expected to be able to use a substantial number of Chinese idioms properly (Yang et al. 2006: 755). To establish the connection between the meaning of these compounds and *bù* as one of their elements will be extremely difficult.

(3.4) 三不五時

Sān-bù-wǔ-shí

Three-not-five-time

From time to time

There are other types of fixed expressions that are not *chéngyǔ*. Some words contain fewer than four syllables with a word structure similar to the negative prefix or A-not-A pattern, but the function of the morpheme *bù* is different. (3.5) and (3.6) are examples.

(3.5) 不等 (postposition)

Bù-děng

Not-equal

in between: located in the middle of two things

(3.6) 動不動

Dòng-bú-dòng

Move-not-move

On every occasion

Table 3.1 is a summary of each *bù*-related word group. Excluding the contrastive conjunction group, *bù* often becomes a constituent of an adverb or a verb. In some groups, *bù* can be part of a noun, an adjective or a determiner. Of the seven groups, only the *bù* morphemes in the idiomatic and semantically less connected compounds do not function as a negator.

Group	Word structure	Word class	Connotation of <i>bù</i>	Examples
Negator <i>bù</i>	(Morpheme A)+ <i>bù</i>	adverb, noun, particle	Negator	<i>bù</i> 不 not, <i>juébù</i> 絕不 never
Contrastive conjunction	(Morpheme A)+ <i>bù</i> +morpheme B	conjunction	Contrast	<i>bùjǐn</i> 不僅 not only that, <i>búguò</i> 不過 but, <i>búlùn</i> 不論 whether, <i>bùguǎn</i> 不管 no matter
Negative prefix	(Morpheme A)+ <i>bù</i> +morpheme B	adverb, adjective, verb, noun, determiner	Antonyms of the following morpheme	<i>búdà</i> 不大 not very, <i>búzài</i> 不再 no longer, <i>bùzú</i> 不足 insufficient, <i>bùkěnéng</i> 不可能 impossible
Potential infix	Insert <i>bù</i> in a resultative verb compound	adverb, verb	negating the compound	<i>láibùjí</i> 來不及 too late, <i>fēnbùkāi</i> 分不開 inseparable, <i>chàbùduō</i> 差不多 not far-off
A-not-A pattern	A-not-A pattern	adverb, verb	A or not A	<i>huìbúhuì</i> 會不會 will or won't, <i>hǎobùhǎo</i> 好不好 ok or not, <i>gāibùgāi</i> 該不該 should or

				shouldn't
Idiomatic and semantically less connected compounds	Random number of morphemes	adverb, adjective, verb, noun, postposition	Semantically less connected to the negation connotation	<i>bùxiāngshàngxià</i> 不相上下 comparable, <i>duìbùqǐ</i> 對不起 sorry, <i>liǎobùqǐ</i> 了不起 marvellous
Double negative	Two negative markers	adverb	Positively stressed	<i>wúbù</i> 無不 all, <i>bùdébù</i> 不得不 have to

Table 3.1: Groups of *bù*-related words

4 Methods

The aim of the study is to investigate the use of *bù*-related words in research articles and inspect differences in word choice between academic writing and unscripted conversations. In order to investigate the use of *bù*-related words in academic writing, we first compiled a corpus consisting of German linguistics research articles written in Chinese. After extracting the main body of each article, the CKIP⁶ word segmentation system was used to add part-of-speech labels to the corpus. In order to be able to trace back to the article containing a given *bù*-related word at a later stage, we compiled an inverted index for all words containing *bù*. Inverted index is a type of index data structure, which enables efficient data retrieval. In this study, we used inverted index for storing the mapping of words and the documents (see Section 4.4 for details). Through the mapping, the documents containing the word in question can be assessed within a fraction of time. The last stage was to manually categorise all the *bù*-related words into seven groups according to the discussion in section 3. After categorising *bù*-related words in the research article corpus and the conversational corpus TMC, we compared the frequency of each group in both corpora. The whole procedure is illustrated in Figure 4.1.

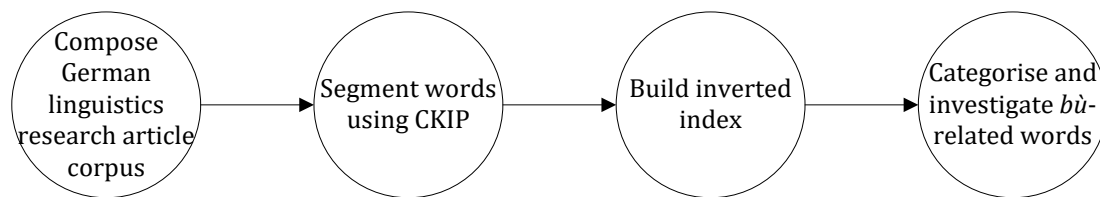


Figure 4.1: Research procedure

The distributions of *bù*-related words in research article will be compared with the distributions of that in unscripted conversations. Tseng's (2013) Chinese Spoken Wordlist which is generated from Taiwan Mandarin Conversational corpus (TMC) will be used for the comparison. The wordlist and the conversational corpus will be introduced in section 4.1. In section 4.2, our corpus will be introduced. Compound words are under investigation in this study and the word boundaries depend on the principles of word segmentation adopted by the word segmentation system. Section 4.3 introduces word segmentation principles adopted by

⁶ CKIP stands for Chinese Knowledge and Information Processing. It is a research group based in Academia Sinica, Taiwan.

CKIP and discusses how they are connected with different *bù*-related word group. Section 4.4 introduces the inverted index.

4.1 Taiwan Mandarin Conversational corpus and Tseng's Chinese Spoken Wordlist

Taiwan Mandarin Conversational corpus (TMC) is based on several spoken corpora collected by the Institute of Linguistics at Academia Sinica in Taiwan. The corpus contains 85 conversations of three different scenarios: conversations on unspecified topics between strangers; conversations on selected news or events between friends and relatives; and Map Task dialogues (Tseng 2013: 3–4). Tseng's Chinese Spoken Wordlist is generated from TMC. The wordlist contains 16,681 unique words, which is the make-up of 405,435 Chinese words in TMC. Information available from the wordlist includes word class label, pronunciation, number of syllabus in a word, frequency, and accumulated frequency in percentage. A word class label, i.e. part-of-speech tag, was attached to each word after applying CKIP word segmentation and POS tagging system to the transcripts of the conversations in TMC (ibid.: 6). The words are listed in the order of frequency. According to the wordlist, *bù* 不 as a standalone adverb occurs 6,677 times in TMC, and there are 243 unique words in TMC that have *bù* as one of the characters.

4.2 Materials: German Linguistic Research Article Corpus (GLRA Corpus)

The corpus is composed of 30 journal articles published in Taiwan written by 20 different German linguistics scholars who teach in Taiwan and speak Chinese as their first language. Topics are mainly associated with German language teaching and learning, including teaching methods, intercultural competence, teaching materials, and vocabulary acquisition. We limited the scope of the topic so that the number of word types used would be realistic to one particular community. TMC represents words that are used by an average Mandarin speaker in Taiwan. Since most of *bù*-related words are content words, if a huge variety of topics were included, the comparison between GLRA and TMC would not be fair in terms of word types. Although the topic is limited to German language teaching and learning, we can assume that the results of word distribution can represent articles in other closely related research areas. Language-related journals in Taiwan accept articles written in foreign languages, and a large proportion of the articles are written in the foreign languages that can be mastered by the authors. As a result, there are only a limited number of Chinese articles discussing German language teaching and learning. We selected articles published after 2003 that are available from the Airiti library, a popular online library subscribed to by many universities in Taiwan. In addition to the research topics, the selection criteria included the length of the article. We discarded articles that were too short, i.e. with a main body less than 6,500 characters. If one author had more than three articles available, we only selected two or three so that the personal style would not dominate the features of the whole corpus.

After deciding which articles should be included, we deleted abstracts, affiliations, footnotes, figures, tables, references and appendixes before applying CKIP word segmentation and POS tagging system to attach a part-of-speech label to each word. POS tagger provides word class information, including nouns, determiners, verbs, and prepositions. The average length of the main body of the articles is 5,754 words, excluding punctuation and foreign words. The total

number of words included in the GLRA corpus is 172,617, and the number of words included in the TMC corpus is 405,435. Both numbers were calculated based on the word segmentation output from the same word segmentation system. Note that different Chinese word segmentation systems might result to different word segmentations even when the input text is identical. The following section describes the principles adopted by the CKIP word segmentation system.

4.3 Word segmentation principles

As discussed in section 3, *bù* as a word is an explicit negation marker. Unlike alphabetic writing where spaces are reliable delimiters marking the boundaries of words, Chinese adopts logographic writing where each word is made up of one or more characters and word boundaries are not explicitly marked. The negator *bù* can be a standalone function word in many cases; it can also be part of a word where *bù* acts as a morpheme, similar to *un-* in uneasy. In some cases, *bù* is part of an idiom or a fixed expression where the meaning of the expression is difficult to derive from the combination of the individual characters. Even when *bù* is treated as part of a word, its negative connotation is often preserved. For this reason, a word segmentation tool is required for extracting *bù* and the relevant words.

The CKIP word segmentation system⁷ developed by Academia Sinica in Taiwan will be used for the purpose of segment Chinese sentences in the GLRA corpus into words, and the same system was applied to TMC where Tseng's (2013) Chinese Spoken Wordlist is generated from. Different word segmentation and POS tagging systems might result to different segmentation and different word class labelling. CKIP word segmentation system was chosen because the wordlist generated from the corpus composed for this study need to be comparable to Tseng's wordlist. That is to say, using the same word segmentation system means the same segmentation rules will be applied to two different corpora, and therefore the consistency can be guaranteed. There are variations in Chinese character sets and Chinese character encoding methods used in different Chinese speaking regions. Since the GLRA corpus contains Taiwan Mandarin articles written in traditional Chinese, using a tool developed in Taiwan will avoid errors occur when converting between traditional Chinese character sets used in Taiwan and simplified Chinese character sets used in some Chinese speaking regions. In addition, the errors contributed by the lexical and grammatical differences between Taiwan Mandarin and other varieties of Mandarin will be limited.

According to Oxford Dictionaries (n/a), a word is “a single distinct meaningful element of speech or writing”. In the Chinese writing system, it is sometimes difficult to decide what a word is, since a word can consist of one or more characters, and the meanings of individual characters might contribute to the meaning of the word, as discussed in section 3. There will be cases where native speakers do not agree on the word boundary. The CKIP word segmentation system adopts two word-segmentation principles and four guidelines (Huang et al. 1996: 1046-1047). The two principles (ibid.: 1046) are:

- a) A string whose meaning cannot be derived by the sum of its components should be treated as a segmentation unit.

⁷ The system can be accessed online from here: <http://ckipsvr.iis.sinica.edu.tw/> [10.02.2016].

b) A string whose structural composition is not determined by the grammatical requirements of its components, or a string which has a grammatical category other than the one predicted by its structural composition, should be treated as a segmentation unit.

The four guidelines (ibid.: 1047) are as follows:

- a) Bound morphemes should be attached to neighbouring words to form a segmentation unit when possible.
- b) A string of characters that has a high frequency in the language or high co-occurrence frequency among the components should be treated as a segmentation unit when possible.
- c) Strings separated by overt segmentation markers should be segmented.
- d) Strings with complex internal structures should be segmented when possible.

According to principle a), when the combination of *bù* and other characters occurs together and the antonym cannot be derived from deleting *bù*, the combination will be treated as a word. Idioms, contrastive conjunctions and some of the words in negative prefix discussed in section 3 are compound words as a result of this principle. According to CKIP (n/a: 4), when *bù* is followed by a modal, the combination will be treated as an adverb. Some words in the negative prefix are of this type. Words other than *bù* in the negator group are segmented as words for their high co-occurrence frequency in the language. *Bù* in potential infix and A-not-A pattern are treated as a bound morpheme which should be attached to neighbouring words. However, if both As in an A-not-A pattern are complete words, the system will separate them according to guideline c) and treat *bù* as an overt segmentation marker. For example, (4.1) will be treated as three words, and (4.2) will be treated as one word, since *xǐ* itself is not equal to *xǐhuān*.

(4.1) *Xǐhuān-bù-xǐhuān* 喜歡不喜歡 like-not-like

(4.2) *Xǐ-bù-xǐhuān* 喜不喜歡 like-not-like

4.4 Inverted index

An inverted index is commonly used for information retrieval tasks (Manning et al. 2008: 6–10), and it is useful for tracing a particular word back to one of the documents in the corpus. An inverted index consists of a dictionary and postings, as illustrated in Figure 4.2.

Dictionary		Postings (Document ID.)							
不 (ADV)	→	1	2	3	4	5	6	7	...
不同 (Vi)	→	1	3	4	5	6	7	8	...
不能 (ADV)	→	3	5	7	8	9	10	11	...
不僅 (C)	→	3	4	7	10	11	12	13	...
...	→								

Figure 4.2: Two parts of the inverted index: the dictionary is composed of words containing *bù*. Each posting list records documents containing the corresponding word.

In this study, the dictionary consists of *bù*-related word in the GLRA corpus. If we are interested in how a particular word is used in context, we can look up the postings and examine the corresponding documents. We intended to compare the distributions of *bù*-

related words in TMC corpus and in GLRA. However, we do not have direct access to TMC corpus, and can only acquire Taiwan Mandarin Spoken Wordlist generated from TMC. Hence, we were unable to attain the original utterance containing a particular word in TMC corpus. Instead, we justify the usage of words in conversations based on the native speaker's intuitive judgment.

5 Results and discussion

The percentage of negator *bù* in the conversational corpus is significantly higher than in the research articles, as shown in Table 5.1. The same table also shows that more than half of the *bù*-related word occurrences in conversation are the negator *bù*, which is not the case for research articles. The results are similar to Biber et al.'s (1999: 159), where *not/n't* occurs almost eight times more frequently than other negative forms in conversation, and the number in conversation is five times that in academic articles. In addition, Biber et al.'s data shows that *not/n't* occurs more often than the total of other negative forms. The reason may be that the forms of negation discussed by Biber et al. were limited to *not*-negation and *no*-negation, and the negative words marked by affixes were not counted. Our data shows that there are more other *bù*-related words in academic articles than the frequency of the negator *bù* in the same corpus. Although the size of the GLRA corpus is less than half the TMC corpus, the number of types of *bù*-related words is similar: 210 in GLRA and 274 in TMC. Only 95 types of *bù*-related words appear in both corpora, meaning that 145 types appear only in GLRA and 179 only in TMC. Of the words appear only in one of the corpora, only a few words such as *bùdòngchǎn* 不動產 'real estate' and *bùdìngcí* 不定詞 'infinitive' are associated with specific topics. Most of the words can fall into two of the vocabulary groups defined by Nation (2001): high frequency words and academic vocabulary. The following sub-sections will examine the distribution of each group of *bù*-related words.

Corpus	Types of <i>bù</i> -related words	Frequency of Negator <i>bù</i>	Frequency of other <i>bù</i> -related words	Total running words (tokens)
GLRA Corpus	210	768 (0.44%)	1340 (0.78%)	172,617 (100%)
TMC Corpus	274	6,680 (1.65%)	4,327 (1.07%)	405,435 (100%)

Table 5.1: *Bù*-related words in the two corpora

5.1 Contrastive conjunctions

The frequency of this group is similar between the two corpora, as shown in Table 5.2. Eight types of contrastive conjunctions appear in both corpora. *Búruò* 不若 'no match for' appears only in GLRA, and *bùrán* 不然 'otherwise', *yàobù* 要不 'otherwise', *yàobúshì* 要不是 'if not', *zhībúguò* 只不過 'but', appear only in TMC. Terms like *dàn* 但 'but' and *fǒuzé* 否則 'if not' are frequent words in academic writing to replace frequent contrastive conjunctions used in conversations such as *búguò* 不過 'but' and *bùrán* 不然 'otherwise'. There are no common

contrastive conjunctions used only in research articles; however, there are some words from this group frequent in conversations, but unsuitable for academic writing.

Corpus	Type	Frequency of contrastive conj.	Conj. tokens/total running words
GLRA Corpus	9	204	0.12%
TMC Corpus	12	639	0.16%

Table 5.2: Contrastive conjunctions in the two corpora

5.2 Negative prefix

The distribution of the negative prefix group in both corpora is similar, as shown in Table 5.3; only 61 types occur in both corpora, leaving a large part of compounds appearing in only one of the corpora.

Corpus	Type	Frequency of prefix	Prefix tokens/total running words
GLRA Corpus	106	1,003	0.58%
TMC Corpus	106	2,764	0.68%

Table 5.3: Negative prefix in the two corpora

Different types of words in this group might result from the topic under discussion. For example, ethical issue might be difficult to fit into the field of applied linguistics, so compounds like *bùdàodé* 不道德 ‘immoral’, *bùrěndào* 不人道 ‘inhuman’ are unlikely to appear in the GLRA. In addition, the level of formality might play a part in the results. For example, *bùduìjìn* 不對勁 ‘not right’ will be considered too informal for academic writing, and *bùtóngyú* 不同於 ‘different from’ is rare in unscripted speech. *Bùjìn* 不盡 ‘not all’ often co-occurs with *xiāngtóng* 相同 ‘same’ in academic writing as in (5.1), and does not appear in TMC corpus. *Bùjìn* mark the feature of hedging in academic writing: while the author of (5.1) could have said *bù xiāngtóng* 不相同 ‘not the same’, she chose to accept the possibility that the meanings of those two nouns in discussion could somehow be the same.

(5.1) 這 兩 個 名詞 意思 不盡 相同 (Chen 2010: 73)

Zhè liǎng gè míngcí yìsi bùjìn xiāngtóng

These two MEASURE noun meaning not-all same

The meanings of these two nouns are not exactly identical

There are more boosters and hedges in this category which appear only in research articles, but not in conversations. Of the 42 negative prefix words unique to academic writing, nine are boosters and only five are hedges. As shown in Table 5.4, there are a range of boosters in this category being used by more than one author, but only two hedges are repeatedly used.

Hedges				
Word	Translation	Category	Document frequency*	Term frequency**
<i>bùjìn</i> 不盡	not all	adverb	9	14

<i>bùmíng</i> 不明	Unclear	verb	7	7
<i>bùquán</i> 不全	not all	verb	1	1
<i>bùshī</i> 不失	can be considered as	verb	1	1
<i>bùshīwéi</i> 不失為	can be considered as	verb	1	1
Boosters				
<i>bùxū</i> 不需	no need	adverb	6	8
<i>búwàihū</i> 不外乎	nothing more than	verb	3	5
<i>búwài</i> 不外	nothing more than	adverb	3	3
<i>bùmǎn</i> 不免	inevitably	adverb	2	2
<i>bùdān</i> 不單	not just/not only/not merely	adverb	2	2
<i>búxiàyú</i> 不下於	no less than	verb	1	1
<i>bùdāndān</i> 不單單	not just/not only/not merely	adverb	1	1
<i>bùfù</i> 不復	no longer	adverb	1	1
<i>bùxū</i> 不須	no need	adverb	1	1

*Document frequency refers to the number of articles containing this word.

**Term frequency refers to the number of times the word occurs in the whole corpus.

Table 5.4: Negative prefix used as hedges or boosters appearing only in research articles

When we further categorise negative prefix according to the word classes of the words, interesting patterns emerge. As shown in Table 5.5, most compounds in this category are adverbs and verbs, and more than two thirds of the occurrences in the TMC corpus are adverbs. As discussed in section 4.3, a word is labelled as an adverb when *bù* is followed by a modal. Most of these adverbs in the TMC corpus express modality described by Lyons (1977: 787–849). In other words, while modality is an important linguistic device to articulate arguments in academic writing in English and German, the use of modality could be also important to Chinese academic writing. This is the reason why a variety of negation and modal verb combinations are being used in the research articles.

Category	Corpus	Type	frequency	Proportion
adverbs	GLRA Corpus	38	383	0.22%
	TMC Corpus	45	1994	0.49%
verbs	GLRA Corpus	58	581	0.34%
	TMC Corpus	70	747	0.18%
nouns	GLRA Corpus	11	22	0.01%
	TMC Corpus	2	2	<0.001%
non-predicate adjectives & determiner	GLRA Corpus	3	17	0.01%
	TMC Corpus	3	21	0.01%

Table 5.5: Negative prefix in the two corpora by category

Of the adverb group, *búhuì* 不會 ‘will not’ occurs 832 times in the TMC corpus but only 59 times in the research articles. *Búhuì* 不會 has at least three possible senses in utterances: (1) the possibility is low, as in (5.2); (2) lack of a certain ability, as in (5.3); (3) in response to gratitude, equivalent to *you are welcome* or *no problem*. The third meaning is used particularly in Taiwan but not in other Chinese-speaking areas, and also deviates from written Chinese. Although it should not be labelled as an adverb, the CKIP could have labelled the word as an adverb by mistake.

(5.2) 明天 不會 放晴
Míngtiān *búhuì* *fàngqíng*
 Tomorrow will not clear up
 The weather will not clear up tomorrow.

(5.3) 這 店員 不會 招待 客人
Zhè diànyuán búhuì *zhāodài kèrén*
 This clerk do-not-know-how serve customers
 This clerk does not know how to serve customers

Bùnéng 不能 ‘cannot’ is the most frequent word occurring in research articles in this group (84 times), and it is also the third most frequent in the TMC corpus (230 times). It has two meanings: (1) unable to; (2) not permitted. The first meaning overlaps with *búhuì* 不會. When the ability connotation is expressed and the subject is animate, both *búhuì* and *bùnéng* could be used; if the subject is inanimate, *huì* 會 cannot be used. (Zhu 1996: 197). This might be the reason why *búhuì* is more frequent in conversation than in research articles. Sentences containing animate subjects are infrequent in research articles. The other two frequent adverbs appearing in the TMC corpus are *búyào* 不要 ‘do not want to do something/don’t’ and *búyòng* 不用 ‘need not’ which appear 324 and 225 times in TMC respectively, and only 9 and 8 times in the GLRA corpus. Both words can be used in comments or to express a person’s not wanting something. *Búyòng* is typically used in imperatives (Xiao and McEnery 2008). The other two frequent adverbs in GLRA are *bùkě* 不可 ‘may not/must not’ (36 times) and *búzài* 不再 ‘no longer’ (33 times). *Bùkě* 不可 occurs 11 times in TMC. The synonym *bùkěyǐ* 不可以 occurs 21 times in TMC and only 7 times in GLRA. This shows that there is a tendency for the authors of research articles to prefer short words whenever possible in Chinese. *Bùkě* and *búzài* are often used as boosters to emphasise a certain point in an argument. (5.4) and (5.5) are examples extracted from the GLRA corpus.

(5.4) 文化 學習 與 語言 學習 密不可分、不可 切割。(Chen 2010: 80)
Wénhuà xuéxí yǔ yǔyán xuéxí mìbùkěfēn, búkě qiēgē.
 culture learning and language learning inseparable, cannot separate
 Culture learning and language learning are inseparable – they cannot be separated.

(5.5) 研究翻譯的方向不再只限於語言內結構問題

(Shieh 2010: 153)

Yánjiū fānyì de fāngxiàng búzài zhǐ xiànyú yǔyán nèi jiégòu wèntí
 Research translation DE direction no-longer only limit-to language inside structure
 issues

The directions of translation research are no longer limited to issues regarding language structures.

Table 5.5 also shows that there are more nouns in research articles than in conversation. Chinese authors of research articles may have adopted a nominalisation strategy similar to English authors of research articles, as described in Tang and John (1999: S33). The group of words appearing in GLRA includes *bùtóngchù* 不同處 ‘difference’, *bùzúxìng* 不足性 ‘insufficiency’, *bùāngǎn* 不安感 ‘anxiety’.

5.3 Potential infix

Bù as a potential infix is rare in academic writing since this type of construction tends towards the colloquial. As shown in Table 5.6, there are very few occurrences of potential infix in academic writing. 14 of the 30 research articles contain this type of *bù*-related words, and of 29 occurrences, eight are from the same article, and another eight are quotations used as language examples. The results suggest that *bù*-related words with *bù* as a potential infix are too casual to be used in academic writing. Only nine of the types appear in both corpora, suggesting that the use of this group of words also depends on the topic under discussion.

Corpus	Type	Frequency of <i>bù</i> infix	<i>bù</i> infix tokens/total running words
GLRA Corpus	24	29	0.02%
TMC Corpus	66	396	0.10%

Table 5.6: Potential infix in the two corpora

5.4 A-not-A pattern

Since interrogatives are not common in research articles, it is reasonable that there are many more A-not-A patterns in conversation than in academic articles, as shown in Table 5.7. From the low frequency of this group in the GLRA corpus, we can assume that the A-not-A pattern is not popular in *whether*-leading clauses in research articles. Five of the types appearing in GLRA also appear in TMC.

Corpus	Type	Frequency of A-not-A	A-not-A tokens/total running words
GLRA Corpus	6	9	0.01%
TMC Corpus	25	371	0.09%

Table 5.7: A-not-A pattern in the two corpora

5.5 Double negative

Double negative compound words are rare in both corpora, as shown in Table 5.8. There are only a few word types of double negative compounds, and the occasions for needing to use one are uncommon. Of the twelve occurrences in the GLRA corpus, one is wrongly segmented by CKIP, and two are quotations. Double negative compounds could be used as

boosters in research articles. Only five authors in the GLRA corpus used words in this category. The results indicate that the use of double negative in academic writing is not an essential device in Chinese academic writing. The overemphasised connotation might not be widely accepted in this field, although other types of boosters are common in a Chinese research article.

Corpus	Type	Frequency double negative	Double neg. Tokens/total running words
GLRA Corpus	3	12	0.01%
TMC Corpus	2	5	< 0.01%

Table 5.8: Double negative in the two corpora

5.6 Idiomatic and semantically less connected compounds

Only nine compounds of this group occur in both corpora. Research articles contain a large number of four-syllable Chinese idioms (*chéngyǔ*) and very few other fixed expressions in this group, as shown in Table 5.9. On the other hand, four-syllable Chinese idioms (*chéngyǔ*) are less common than other expressions in the conversational corpus. The idioms are proverbs or dead metaphors based on traditional structures and patterns. To memorise those set phrases are an important part of the Chinese language education in Taiwan. The results are in line with the findings of previous studies, for example González et al. (2001). González et al. found that even when Chinese authors write in English, set phrases like *stars moving and constellations changing* were used to convey the meaning of the passing of time (ibid.: 435). These types of set phrases are often used to convey abstract ideas, and are frequently used in Chinese academic writing.

Corpus	Type			Frequency			Tokens/total running words		
	<i>chéngyǔ</i>	other	Total	<i>chéngyǔ</i>	other	Total	<i>chéngyǔ</i>	other	total
GLRA	48	8	56	72	11	83	0.04%	0.01%	0.05%
TMC	39	22	61	55	97	152	0.01%	0.02%	0.04%

Table 5.9: Idiomatic and semantically less connected compounds in the two corpora

The results exhibited in this section suggest that there is a tendency towards avoiding the use of *bù* as a standalone negator and as an infix in Chinese academic writing. However, using compound words with *bù* as a negative prefix is not uncommon, bearing in mind that this group of words has different degrees of formality and that some of the compound words frequently used in conversation might not be good choices for academic writing. From the comparison of the two corpora, we can also assume that in a publishable Chinese academic article, at least in the field of applied linguistics, use of *chéngyǔ* is important.

6 Conclusion

This article categorises word compounds containing *bù* and compares the distribution of *bù*-related words between academic writing and the spoken language in Taiwan. The categorisation is based on word structure and the level of negative connotation which the morpheme *bù* retains in the word. All the *bù*-related words are grouped into seven categories: “negator *bù*”, “contrastive conjunctions”, “negative prefix”, “potential infix”, “A-not-A

pattern”, “double negative”, and “idiomatic and semantically less connected compounds”. The results show that *bù* as a negator is much more frequent in the conversational corpus than in research articles. This finding is consistent with previous studies where English is the language under discussion. Potential infix and the A-not-A pattern are rare in research articles. On the other hand, four-syllable Chinese idioms (*chéngyǔ*) are much more common in research articles than in conversation. Within the negative prefix group, words expressing modality are more frequent in the research articles than in conversations.

Corpus	Contrastive conjunctions	Negative prefix (frequency ≥ 3 in TMC or ≥ 2 in GLRA)
GLRA Corpus	<i>búruò</i> 不若 no match for	<i>bùjìn</i> 不盡 (adv.) not all, <i>bùtóngyú</i> 不同於(v.) different from, <i>bùxū</i> 不需 (adv.) no need, <i>bùmíng</i> 不明 (v.) unclear, <i>bùjiǔ</i> 不久(n.) not long, <i>búwàihū</i> 不外乎(v.) nothing more than, <i>bùduìchèn</i> 不對稱 (v.) asymmetric, <i>búgù</i> 不顧 (v.) disregard <i>bùwài</i> 不外 (adv.) nothing more than, <i>bùyí</i> 不宜 (v.) unsuitable, <i>bùxúncháng</i> 不尋常(v.) unusual, <i>bùcéng</i> 不曾 (adv.) not ever, <i>bùmiǎn</i> 不免 (adv.) have to, <i>bùdān</i> 不單 (adv.) not only, <i>bùxiǔ</i> 不朽 (v.) immortal, <i>bùzú</i> 不足 (n.) insufficiency,
TMC Corpus	<i>bùrán</i> 不然 otherwise, <i>yàobù</i> 要不 otherwise, <i>yàobúshì</i> 要不是 if not, <i>zhībúguò</i> 只不過 but	<i>bùjǐngqì</i> 不景氣 (v.) depression, <i>bùhǎoyìsi</i> 不好意思 (v.) embarrassed, <i>bùnénggòu</i> 不能夠 (adv.) cannot, <i>bújiàn</i> 不見 (v.) disappear, <i>bùzhǔn</i> 不准 (adv.) do not permit, <i>zàibùrán</i> 再不然 (adv.) if not so, <i>bùpíng</i> 不平 (v.) uneven or injustice, <i>zhībúguò</i> 只不過 (adv.) only, <i>búxià</i> 不下 (v.) not less than, <i>bùkěn</i> 不肯 (adv) unwilling, <i>bùbǐ</i> 不比 (v.) unlike, <i>bùlǐ</i> 不理 (v.) ignore, <i>búdàodé</i> 不道德 (v.) immoral

Table 6.1: Contrastive conjunctions and negative prefix wordlist in the two corpora

Although the proportions of *contrastive conjunctions* in the two corpora are similar, there are words commonly used in one corpus and absent from the other. A similar conclusion can be drawn for the *negative prefix* group. Table 6.1 lists words that are frequent in one corpus but absent from the other. Note that many of the words in the list express modality, and are not linked to the topics under discussion. Most words appear only in GLRA corpus in Table 6.1 can be considered as “academic vocabulary”. The concept that there is a set of vocabulary need to be learned for special purposes is promoted by language teachers who teach English language for academic purposes like Coxhead (2000). Within the wordlist derived from GLRA corpus in Table 6.1, there are more boosters than hedges. The distribution is in line

with the findings from previous studies: advocating a certain idea tends to be more important than debating the truth in Chinese academic writing.

The results exemplify the distinction between the spoken language used by every speaker and academic writing used by a small proportion of the population. Those who embark on writing research articles in Chinese should be aware of the features of Chinese academic writing. For example, one might expect more *chéngyǔ* in a research article than in conversation, and words that are frequent in conversation might be unsuitable for academic writing. In addition, in situations where a simple negator *bù* can be used in conversation, there should be another way to express the idea without using negative words in academic writing.

A better understanding of negative elements will benefit the field of natural language processing. This article shows that not every *bù*-related word in Chinese is negative or can be treated equally. The negator *bù* and *bù* as a potential infix are both negators; on the other hand, contrastive conjunctions containing *bù* signal contrast without negating the following constituents.

This article only investigates *bù* at the lexical level. Use of negation at the *n*-gram and clause levels should be explored in the future, in order to acquire a full picture of negation in Chinese.

References

- Biber, Douglas et al. (1999): *Longman Grammar of Spoken and Written English*. Essex: Pearson Education Limited.
- Chen, Ying-hui (2010): “Exploring Students’ Cultural Concepts When Teaching German as a Foreign Language: A Task-Based Teaching Method”. *Journal of Humanities and Social Sciences* 6/1: 71–86. (In Chinese)
- CKIP group. (n.d.): *Chinese Word Segmentation Standards for Information Processing*. Taipei: Academia Sinica. (In Chinese) http://rocling.iis.sinica.edu.tw/CKIP/paper/wordsegment_standard.pdf [17.02.2016].
- Coxhead, Averil (2000): “A New Academic Word List”. *TESOL Quarterly* 34/2: 213–238.
- Ding, Huiling (2007): “Confucius’s Virtue-Centered Rhetoric: A Case Study of Mixed Research Methods in Comparative Rhetoric”. *Rhetoric Review* 26/2: 142–159.
- Ernst, Thomas (1994): “Conditions on Chinese A-not-A Questions”. *Journal of East Asian Linguistics* 3/3: 241–264.
- Giora, Rachel (2006): “Anything Negatives Can Do Affirmatives Can Do Just As Well, Except For Some Metaphors”. *Journal of Pragmatics* 38: 981–1014.
- González, Virginia/Chen, Chia-Yin/Sanchez, Claudia (2001): “Cultural Thinking and Discourse Organizational Patterns Influencing Writing Skills in a Chinese English-as-a-Foreign-Language (EFL) Learner”. *Bilingual Research Journal* 25/4: 417–442.
- Halliday, Michael Alexander Kirkwood (1985): *An Introduction to Functional Grammar*. London: Edward Arnold.
- Herrero-Zorita, Carlos (2013): “A Statistical Study of the Usage of No-Negation and Not-Negation in Spoken Academic English”. *Procedia-Social and Behavioral Sciences* 95: 482–489.
- Hinkel, Eli (1997): “Indirectness in L1 and L2 Academic Writing”. *Journal of Pragmatics* 27: 361–386.

- Hinkel, Eli (1999): "Objectivity and Credibility in L1 and L2 Academic Writing". In: Hinkel, Eli (ed.): *Culture in Second Language Teaching and Learning*. Cambridge, Cambridge University Press: 90–108.
- Hinkel, Eli (2005): "Hedging, Inflating, and Persuading in L2 Academic Writing". *Applied Language Learning*, 15/1,2: 29–53.
- Horn, Laurence R. (2001): *A Natural History of Negation*. Stanford: CSLI Publications. Originally published: Chicago: University of Chicago Press, 1989.
- Hsieh, Chia-Ling (2003): "Chinese Modal Verbs and Modal Adverbs: A Semantic Definition and Categorization". *Proceedings of the Seventh World Chinese Teaching Conference 1: Linguistic Analysis*, Taipei/Taiwan: 55–73. (In Chinese)
- Hu, Guangwei/Cao, Feng (2011): "Hedging and Boosting in Abstracts of Applied Linguistics Articles: A Comparative Study of English- and Chinese-Medium Journals". *Journal of Pragmatics* 43: 2795–2809.
- Huang, Chu-Ren/Chen, Keh-jiann/Chang, Li-Li (1996): "Segmentation Standard for Chinese Natural Language Processing". In *Proceedings of the 16th Conference on Computational Linguistics 2*, Copenhagen/Denmark: 1045–1048.
- Jordan, Michael (1998): "The Power of Negation in English: Text, Context and Relevance". *Journal of Pragmatics* 29: 705–752.
- Hyland, Ken (1998): "Boosting, Hedging and the Negotiation of Academic Knowledge". *Text* 18/3: 349–382.
- Li, Charles N./Thompson, Sandra A. (1981): *Mandarin Chinese: A Functional Reference Grammar*. Berkeley: University of California Press.
- Loi, Chek Kim/Evans, Moyra Sweetnam (2010): "Cultural Differences in the Organization of Research Article Introductions from the Field of Educational Psychology: English and Chinese". *Journal of Pragmatics* 42: 2814–2825.
- Lyons, John (1977): *Semantics 2*. Cambridge: Cambridge University Press.
- Manning, Christopher D./Raghavan, Prabhakar/Schütze, Hinrich (2008): *Introduction to Information Retrieval*. New York: Cambridge University Press.
- Monrante, Roser/Sporleder, Caroline (2012): "Modality and Negation: An Introduction to the Special Issue". *Computational Linguistics* 38/2: 223–260.
- Nation, Paul (2001): *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Oxford Dictionaries (n.d.): word: *English Dictionary*. Oxford University Press. www.oxforddictionaries.com/definition/english/word [17.02.2016].
- Pagano, Adriana (1994): "Negative in Written Text". In: Coulthard, Malcolm (ed): *Advances in Written Text Analysis*. London, Routledge: 250–265.
- Ross, Claudia/Ma, Sheng Jing-heng (2006): *Modern Mandarin Chinese Grammar. A Practical Guide*. London: Routledge.
- Shieh, Shu-Mei (2010): "Translation and Cross-Cultural Communication". *Journal of European Languages* 3: 147–163. (In Chinese)
- Taboada et al. (2011): "Lexicon-Based Methods for Sentiment Analysis". *Computational Linguistics* 37/2: 267–307.
- Tang, Ramona/John, Suganthi (1999): "The 'I' in Identity: Exploring Writer Identity in Student Academic Writing Through the First Person Pronoun". *English for Specific Purposes* 18: 23–39.

- Tottie, Gunnel (1991): *Negation in English Speech and Writing. A Study in Variation*. London: Academic Press Limited.
- Tseng, Shu-Chuan (2013): “Lexical Coverage in Taiwan Mandarin Conversation”. *International Journal for Computation Linguistics and Chinese Language Processing*. 18/1 :1–18.
- van der Auwera, Johan (2010): “On the Diachrony of Negation”. In: Horn, Laurence R. (ed.): *The Expression of Negation*. Berlin, de Gruyter: 73–110.
- Vincze, Veronika et al. (2008): “The BioScope Corpus: Biomedical Texts Annotated for Uncertainty, Negation and Their Scopes”. *BMC Bioinformatics* 9 (Suppl 11): 9.
- Xiao, Richard Z./McEnery, Anthony M. (2008): “Negation in Chinese: A Corpus-based Study”. *Journal of Chinese Linguistics* 36/2: 274–330.
- Yang, Yu/Read, Stephen J./Miller, Lynn C. (2006): “A Taxonomy of Situations from Chinese Idioms”. *Journal of Research in Personality* 40: 750–778.
- Zhu, Yongsheng (1996): “Modality and Modulation in Chinese”. In: Berry, Margaret et al. (eds.): *Meaning and Form: Systemic Functions Interpretations*. Norwood, New Jersey, Ablex Publishing Corporation: 183–209.