

Marking Trustworthiness with Near Synonyms: A Corpus-based Study of “Renwei” and “Yiwei” in Chinese

Bei Li

Chu-Ren Huang

Si Chen

Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University
Hong Kong SAR

benita-bei.li@connect.polyu.hk churen.huang@polyu.edu.hk sarah.chen@polyu.edu.hk

Abstract

We conducted a corpus-based study on near synonymous cognitive verbs “renwei 認為” and “yiwei 以為” with a similar meaning of “to think” in Chinese. The motivation of this study is that information trustworthiness in propositions varies with the use of “renwei” or “yiwei”, which concerns the issue of speakers’ stances or attitudes. We examined the epistemic modality attributed to “renwei” and “yiwei” in written discourse, by testing their different functions of evidential marking, negative forms and hedging device. Results showed that “renwei” was associated with specialized and professional subjects and consequently demonstrated its relatively high degree of evidentiality. Then, specific negative polarity shifters exclusively collocating with “yiwei” denoted larger scope of negation of entire propositions. Lastly, only “renwei” functioned as a hedge to mitigate claims lacking full commitment in the political discourse. Therefore, the evidence of data proved that the epistemic marker “renwei” was associated with a higher degree of information trustworthiness, contributing to the annotation system in NLP for the detection of information credibility.

Keywords: information trustworthiness; near-synonyms; epistemic modality

1 Introduction

Epistemic modality is an important topic on semantics, functioning in judging and evaluating the propositional content with regard to the degree of certainty, possibility or necessity in our communication. Speakers or writers use epistemic markers, like cognitive verbs (e.g. “think”/“believe”) or

modal auxiliaries (e.g. “may”/“could”), to express their attitudes or stances-taking toward the propositions. Recently the linguistically decoded information involving cognition has been an important topic to be faced in natural language processing (NLP). For example, in the statement “I thought he always supported me”, the speaker’s intention is to emphasize that “he” does not support the speaker now through the use of the past tense of the epistemic verb “think”. Consequently, the trustworthiness of information seems complicated to decode in NLP with regard to the speaker’s attitudes and stances. The present study investigates one type of epistemic markers, the cognitive verbs in Chinese, to analyze the credibility of information epistemic verbs attributed using a corpus-based approach.

Moreover, the use of near synonymous epistemic verbs makes it more complicated for text detection or information retrieval in Chinese. In the present study, the cognitive verbs “renwei” (認為) and “yiwei” (以為) will be comparatively analyzed to detect the information trustworthiness contained in the statements they constituted respectively. The pair of near synonyms “renwei” and “yiwei” has a similar meaning of “to think” with the function of marking speakers’ epistemic modality. Specifically, “ren4” (認) is a verb referring to the act of identification or recognition, and “wei2” (為) is a verb denoting the action or the state of being (apply to both “renwei” 認為 and “yiwei” 以為). The preposition “yi3” (以) of “yiwei” means “with” or “by”. The grammatical constructions consisting of “renwei” or “yiwei” are similar to the epistemic parenthetical “(I) think” in English. The subject and a complement-taking

psych verb (e.g. “renwei” or “yiwei”) form the matrix clause which induces the coming complement clause.

What motivates the investigation on this pair of near synonyms is the variations in conversational implicatures raised by these two epistemic markers in a Chinese context. In sentence (1), the mood and attitude toward the complemented content are neutral to a certain extent, whereas sentence (2) illustrates the negation of complementizer regarding the stance-taking of the addresser. As mentioned above, sentence (2) highlights the fact that “he” does not support “me” now.

(1) 他認為社會福利問題關係重大。

“He **thinks** that social welfare issues are of great importance.”

(2) 我以為他始終支持我的立場。

“I **thought** he always supported me.”

Therefore the present study aims at examining the information trustworthiness in propositions attributed to cognitive verbs. The functions of evidential marking, negative forms and hedging device in epistemic modality are three strategies adopted to test the differences between the two epistemic markers “renwei” and “yiwei”.

Evidentiality: As a sub-category of epistemic modality, the essence of evidentiality is the way that speakers inherently encode the information sources of a clause (Chafe, 1986). There was no agreement upon the category of evidentiality, as several scholars regarded evidentiality as a grammatical structure (e.g. Willett, 1988; Palmer, 2001), while some others expanded into any evidential expressions referring to speakers’ judgments by linguistic coding of the validity of information (e.g. Chafe, 1986; Su et al., 2010). Following the latter broad sense, studies on evidentiality dealt with the epistemological status of information or evidence at addressers’ disposal (De Haan, 2005; Nuyts, 2006).

In speech communication, speakers often use evidential markers to emphasize the explicit source of information, such as “It’s said...”, “People say...”, and “according to...”, in order to improve the credibility of information to a certain degree. According

to Rubin et al. (2006), the lexical verb “think” together with “seem” and “sound” were grouped into the moderate degree of certainty as evidential markers. However, we need to explore whether Chinese lexical verbs “renwei” and “yiwei” have the function of evidential markers, and if any, what factors may affect the trustworthiness they expressed.

Negative forms: Negation is a complicated issue in natural language with the involvement of logic, semantics and pragmatics. (Horn, 1989, Speranza and Horn, 2010). A typical form of negation is realized by negative function words like “not”. For example, propositions like “Sam didn’t smile” are contradictory with “Sam smiled” (Israel, 2004). Besides, negative statements could also be triggered by negative polarity shifters (NPIs), in which the scale-reversing context licenses without explicit negation (Penka and Zeijlstra, 2010). For instance, in the statement “The car failed to climb the hill”, the NPI “fail” reverses the direction of the proposition.

A series of studies on negation and polarity by Horn and his colleagues (e.g. Horn, 1972; 1989; 2001; 2009; Horn and Kato, 2000) investigated the variety of negation scopes and the sophisticated entailment inferences of propositions designated by NPIs and their licensed context. Horn’s polarity theory about negative strengthening illustrated that both litotes and affixal negation were stronger than the non-conventionalized strengthening. That is, the proposition (p) “She is not happy today” has a stronger negative attitude than the proposition ($\neg p$) “It is false that she is happy today”. The proposition (p') “She is unhappy today” also entailed a stronger negation than $\neg p$.

However, whether the negative forms of Chinese epistemic near synonymous verbs have diversities in their implicatures and the truth of propositions is still remained to be explored. What’s more, the topic of polarity shifters is still challenging in Chinese NLP tasks due to the sheer number as well as their invalidation of automatic approaches (Xu and Huang, 2015).

Hedges: Hedging devices as a rhetorical strategy can be realized by particular lexical items, specific grammatical structures or prosodic variations of utterances in conversation. Speakers use hedges to refer to the fuzziness, vagueness, indirectness or

approximation of information they are conveying. In a broad sense, the term hedge is related to all kinds of linguistic means expressing the lack of full commitment (Fraser, 2010).

In previous literature on hedging devices, phrases like “I think” “I guess” and “I wonder” in English were grouped into the type of plausibility shields (Markannen and Schröder, 2010), and the current study followed this categorization. Pragmatic purposes were one important function of hedges which mainly involved the mitigation of claims, showing politeness to listeners, avoiding the criticism of prediction and etc. (Lakoff, 1975; Hyland, 1996; Brown and Levinson, 1987; Taweel et al., 2011).

Additionally, studies on epistemic parentheses as hedges involved a variety of discourses such as juristic judgments (KOŹBIAŁ, 2020) and political speech (Taweel et al., 2011). However, the majority of literature concerned with the conversations in spoken language, little study has paid attention to the genre of newspapers especially on specific discourse by a corpus-based approach.

Therefore, the purpose of the current study is to test the information trustworthiness between the pair of epistemic markers “renwei” and “yiwei” in written discourse. It was first conducted on the perspectives of evidentiality and negative forms to differentiate the credibility of epistemic modality for the near synonyms. Further, the study analyzed the near synonyms in political discourse by corpus to examine the rhetorical strategy of hedging devices for “renwei” and “yiwei”. Consequently, the research questions of the current study are:

- (1) Whether the information trustworthiness is differently contained in statements attributed to the markers “renwei” and “yiwei” in Chinese newspapers?
- (2) If so, which one is more trustworthy, and which one is more opinionated?
- (3) What are their differences in various strategies of epistemic modality?

2 Methods

2.1 The Chinese Word Sketch Approach

The methodology adopted was a corpus-based approach to the comparative analysis of epistemic markers “renwei” and “yiwei” using two datasets. The first one was Chinese Gigaword2 Corpus including gigaword_xin (xin), gigaword_cna(cna) and gigaword_zbn(zbn), collected from the newspapers of Mainland China, Taiwan, and Singapore respectively. Specifically, gigaword_xin (xin) included the texts of journalism formed by more than 200 million words from the Xinhua News Agency of Beijing from the period 1991 to 2002. For gigaword_cna (cna), more than 380 million words constructed the source data of journalism from the Central News Agency of Taiwan from 1990 to 2002. The newspapers of Lianhe Zaobao of Singapore contributed to the data of gigaword_zbn (zbn).

The tool used to process this dataset was the online system Chinese Word Sketch (available at <http://wordsketch.ling.sinica.edu.tw/cws/>). Three functions of the online system were utilized for the current study. First, “Concordance” provided a new query of the word or phrase and displayed the overview of entries containing keywords. This function also played a role in setting specific “left context” and “right context” to the keywords for the search of surroundings.

Then, the function of “Word Sketch Difference” demonstrated both the similarities and differences of two lexical items in terms of their grammatical structures and collocation patterns. By this approach, two keywords could be input simultaneously to obtain the “Common Patterns” and “Only Patterns” collocations for the pair of near synonyms.

Lastly, “Word Sketch” was the third function involved for detailed descriptions of grammatical relations and collocation patterns for a specific lexical item. The values of frequency and salience with respect to specific collocations contributed most to the following findings.

2.2 The Sinica Corpus Approach

Sinica Corpus (Academia Sinica Balanced Corpus of Modern Chinese) provided the dataset of Chinese newspapers on politics for the study. Data were processed by the online tool Sinica Corpus (version 4.0 available <http://asbc.iis.sinica.edu.tw/>) developed by Chen and Huang (2016) on the basis of the dataset in Chen et al. (1996). This dataset collected a total of 1,396,133 sentences, 11,245,330 word tokens, extracted from Chinese text of various articles in Taiwan from 1981 to 2007.

The topics of Sinica Corpus involved philosophy (8%), science(8%), society(38%), art(5%), life(28%) and literature(13%). Moreover, it functioned in narrowing the scope of the query in terms of the genres of a data source. Currently, Sinica Corpus was adopted to contrast the functions of epistemic markers in hedging devices (Section 3.4) for the genre of newspapers on political topics.

To sum up, this study utilized two approaches, the Chinese Word Sketch and Sinica Corpus, to compare the grammatical features as well as the collocation patterns of the pair of near synonyms “renwei” and “yiwei”. The next section of results and analysis concerns the information trustworthiness tested by the evidentiality, negative forms and hedging device in epistemic modality.

3 Results and Analysis

3.1 Overall distributions of “renwei” and “yiwei”

For the dataset we adopted here, there were a total of 745051 entries found for the keyword “renwei” and 22286 entries for “yiwei”. However, the relatively less frequent distribution for the lexical item “yiwei” still involved a number of invalid data.

Because there are two types of combinations consisting of the two Chinese characters 以 “yi3” and 為 “wei2”/“wei4” but have different meanings from “to think”. One combination is the polyphony 為 (“wei4” here) followed by 以 “yi3” constructing the phrase “以為 yiwei” with the meaning of “to do something for...”. The other kind of combination is the ellipsis of the phrase “以之為 yi3 zhi1 wei2” which means “to regard something as...”. While the

current study only considered the lexical item 以為 “yiwei” which possesses a similar syntactic meaning of “to think” with 認為 “renwei”, and excluded the other two formations.

Obviously, “renwei” and “yiwei” showed dramatic differences in the use of frequency in the written discourse. Despite that they shared similar semantic meanings, a tendency to use in written discourse was demonstrated for “renwei” instead of “yiwei”. The following sub-sections will analyze the factors leading to this tendency and effects on their information trustworthiness.

3.2 Differences in Evidentiality

	Collocations Frequency		Collocations Salience	
	renwei	yiwei	renwei	yiwei
Subject	228268	4433	21.9	14
他	37895	337	62.2	27.8
專家	10238	12	53.5	4.5
一般人	212	98	35.1	51.1
我	5676	295	51.0	41.6
人士	9004	28	44.6	7.6
他們	6488	127	43.0	23.8
她	3808	111	40.4	25.9
我們	2848	99	36.6	25.1
人們	647	122	27.9	36.5
大家	944	151	28.8	36.4
學者	2132	6	36.4	3.8
筆者	75	20	31.7	33.8

Table 1. “Common Patterns” collocations of subjects for “renwei” and “yiwei” by “Word Sketch Difference” function

By the function of “Word Sketch Difference”, the overall and comparative descriptions on grammatical relations and collocations demonstrated the subject, modifier and sent subject of keywords with the

frequency and salience. Table 1 shows the “Common Patterns” of “Subject” collocated with the keywords “renwei” and “yiwei”. The first column is the list of subjects collocated with the two keywords, followed by the frequency and salience of each collocation pattern in the second and third columns, in which there are two sub-columns sequentially displaying the details for “renwei” and “yiwei”.

Based on “Common Patterns”, although “renwei” and “yiwei” shared the subjects of pronouns like “他/她” (he/she), “他們” (they) and “我” (I) with relatively approximate salience, “renwei” showed an overwhelmingly frequent distribution of subjects from professional fields, such as experts (專家) and personages (人士) in Table 1. For instance, the salience for the collocations of experts (專家) is 53.5 for “renwei” whereas it is only 4.5 for “yiwei”.

Moreover, different distributions of subjects were apparently approved in the “Only Patterns” for the keyword “renwei” by the function of Word Sketch. Tables 2 lists the “Subject” of “Only Patterns” which exclusively collocated with “renwei” rather than “yiwei”.

	Collocations Frequency	Collocations Salience
Subject	228268	21.9
觀察家	1704	60.9
分析家	2180	59.7
輿論	2677	58.3
與會者	820	49.4
分析師	803	44.8
經濟學家	762	44.0
行家	218	38.9
科學家	1215	35.2

Table 2. “Only Patterns” collocations of subjects for “renwei” by “Word Sketch” function

In Table 2, people of professional fields with specialized skills, as subjects, collocated with the epistemic verb “renwei” with a quite high salience value, such as the observer (觀察家) (salience of 60.9), the

analyst (分析家) (salience of 59.7) and the economist (經濟學家) (salience of 44.0). Two exemplified sentences (3) (4) and their translations were listed below.

(3) 但西方觀察家認為，實際數字遠超過此數。

“But the Western observer **thought** that the actual number far exceeded this number.”

(4) 經濟學家一致認為，新協定成立之初，必然會有人失業。

“Economists **agreed** that there would inevitably be people who were unemployed when the new pact is created.”

Results indicated the differences in the credibility of evidential markers expressed by the cognitive verbs “renwei” and “yiwei”. That is, the information source from professionals was associated with a high certainty of evidentiality. In journalism written discourse, the proposition predicated by “renwei” could express the writers’ or speakers’ attitudes of trust toward the complemented information. While the findings of “yiwei” demonstrated the absence of functions of evidential marking and expressing information trustworthiness from the addressers.

3.3 Differences in Negative Forms

Modifiers	No. of Entries		MI		Only Patterns	
	ren wei	yiwei	ren wei	yi wei	ren wei	yiwei
不 “bu”	456	9	61.9	20.1	Yes	NA
原 “yuan”	NA	6	NA	25.8	NA	Yes

Table 3. Distributions of NPIs “bu” and “yuan” collocating with “renwei” and “yiwei” by “Word Sketch Difference” function

The function of “Word Sketch Difference” also demonstrated the distinct tendency in collocating with negative polarity items (NPIs) among “renwei” and “yiwei”. Table 3 involved two NPIs, 不 bu4, “no”, and 原 yuan2 “barely”, modified “renwei” and “yiwei” with diverse distributions of frequency. Specifically, the construction like “bu” followed by

“renwei” is the “Only Pattern” while the exclusivity is not available (NA) for the condition of “yiwei”. Similarly, the form consisting of “yuan” and “yiwei” is the “Only Pattern” and this modifier is not the “Only Pattern” for “renwei”.

Results indicated that differences in information trustworthiness may be triggered by different NPIs in terms of the scope of negation, epistemic modality and speakers’ attitudes. On the one hand, the NPI like “bu”, a function word of negation, triggered a negative context where the epistemic word “renwei” was negated by the speaker as the sentence (5a) exemplified. The speaker denied the immediate scope of “bu”, the verb “renwei”, instead of the subject or object of a sentence, and therefore the constituent negation canceled the epistemic value of the proposition. Whereas the NPI “bu” rarely collocated with “yiwei” as the sentence (5b) was peculiar in communication.

(5a) 目前我們不認為畫像中有任何人是嫌犯。

(5b) * 目前我們不以為畫像中有任何人是嫌犯。

“At the moment we don't **think** anyone in the portrait is a suspect.”

On the other hand, results of the polarity shifter “yuan” denoted a distinct degree of information trustworthiness from that of the NPI “bu” expressed. As Table 3 showed, “yuan” exclusively collocated with “yiwei” like sentence (6a) exemplified, while this polarity shifter failed to collocate with “renwei” as the sentence (6b) illustrated. The phrase “yuan” followed by “yiwei” did not create an explicit negative context but still denoted implicatures on negative semantic meanings of the entire proposition. That is, the adverbial NPI “yuan” expressed the negative meaning through the comparison of tenses, in which the speaker’s present attitude changed and opposed to his or her own thought in the past. The polarity shifter “yuan” consequently reversed the direction of commitments instead of negating the constituents of clauses like the negative form “bu” followed by “renwei” in example (5a). Hence, the trustworthiness of information expressed by the combination “yuan” and “yiwei” was discriminated from the negative form “bu” and “renwei” due to their varieties in the scope of negation.

(6a) 人們原以為，只要大家遵守某些規定，這些發展就可在同一制度下彼此相輔相成。

(6b) *人們原認為，只要大家遵守某些規定，這些發展就可在同一制度下彼此相輔相成。

“People originally **thought** that these developments can complement each other under the same system as long as everyone complies with certain regulations.”

Apart from NPIs “bu” and “yuan”, “renwei” and “yiwei” also demonstrated differences in negative imperative sentences in terms of the NPIs “buyao” and “bie”. Table 4 presented the frequency distributions of these two NPIs collocating with “renwei” and “yiwei” respectively.

Modifiers	No. of Entries		MI		Only Patterns	
	ren wei	yiwei	ren wei	yiwei	ren wei	yiwei
不要 “buyao”	16	48	24.6	56.5	NA	NA
別 “bie”	NA	15	NA	43.7	NA	Yes

Table 4. Distributions of NPIs “buyao” and “bie” collocating with “renwei” and “yiwei” by “Word Sketch Difference” function

For the NPI “buyao”, though it could modify both “renwei” and “yiwei”, the latter one showed overwhelmingly frequent use of collocating with this modifier in a negative imperative sentence. As for the NPI “bie”, however, collocating with the verb “yiwei” was its “Only Pattern” against “renwei”. We further conducted new individual queries for the collocations of “renwei” and “yiwei” in which “bie” was the left context of the keywords “renwei” or “yiwei” respectively. There were 211 entries found for “yiwei” whereas only 9 entries for “renwei”, even though the overall distribution (section 3.1) of the latter keyword was considerably much more than that of the former one. As sentence (7) illustrated, the speaker expressed strong negative attitudes towards the propositions by using the NPI “bie” in imperative sentences.

(7) 如果您現在仍吃減肥菜，但沒有不適症狀，別以為自己沒事。

“If you are still on a diet but don’t have symptoms of discomfort, don't **think** you are fine.”

3.4 Differences in Hedging Device

By Sinica Corpus, the present research firstly set newspapers as the searching range which was consistent with the type of Chinese Gigaword2 Corpus used in the Chinese Word Sketch approach (adopted from Section3.1. to Section3.3). Then the political discourse was analyzed to contrast the performances of hedging strategy realized by the markers “renwei” and “yiwei”. Table 5 lists the topic distributions of the keywords, and “renwei” was much more frequently used than “yiwei” in nearly all topics.

Topics	“renwei”	“yiwei”
philosophy	61	14
science	64	1
society	2669	274
art	43	7
life	533	109
literature	96	66
Total	3466	471

Table 5. Topic distributions of “renwei” and “yiwei” in Sinica Corpus

In the topic of society, six sub-topics of 政黨 “political party”, 內政 “domestic affairs”, 軍事 “military affairs”, 政治現象 “political phenomena”, 國際關係 “international relations” and 國家政策 “national policy” were set to locate the genre into political discourse in Sinica Corpus. Table 6 shows the frequency of keywords used in sub-topics of political discourse. The small number of entries (33) indicated that the marker “yiwei” was rarely used by addressers in newspapers especially on the political topic. Therefore the following part will analyze the role of hedging device for “renwei” by cases.

Sub-topic	“renwei”	“yiwei”
Political Discourse	689	33

Table 6. Politic sub-topics distributions of “renwei” and “yiwei” in Sinica Corpus

The epistemic marker “renwei” demonstrated the pragmatic function of hedging devices to express fuzzy information under the consideration of speakers’ stances or attitudes. For example, speakers may mitigate their claims by showing the uncertainty of the statement like the sentence (8) indicated. Together with the possibility, the epistemic marker “renwei” emphasized that the speaker was unconfident in the realization of peace.

(8) 特使普利馬可夫透過一名通譯說：我依然感到樂觀，我認為和平有希望達成。

"I remain optimistic, and I **think** there is hope for peace," Special Envoy Primakov said through a translator."

Also, the marker “renwei” used as hedging reflected the power relation in the political domain. The addressers of governmental agencies were entailed high power relation involving governing authority. In newspapers on politics, therefore, the use of “renwei” was a way to mitigate their strong commands, like sentence (9) exemplified.

(9)即使政府認為電子化政府與電子商務勢在必行，也大可透過市場競爭的模式，讓業者各自推動其智慧卡。

“Even if the Government **considers** that electronic government and e-commerce are imperative, it can also enable operators to promote their smart cards through the market competition model.”

Moreover, the epistemic marker “renwei” tended to co-occur with other grammatical structures to hedge the attitudes toward either the facts or opinions of statements together. As for findings of the current study, the modal verbs, conditional clauses or possibilities were all used together with the marker of epistemic modality.

4 Discussion and Conclusions

This study has adopted a corpus-based method to compare the information trustworthiness between two epistemic modalities attributed to “renwei” and “yiwei”. The results obtained provide the literature with multiple perspectives to retrieve the information credibility from cognitive verbs.

First, on the aspect of evidentiality, statements attributed to the marker “renwei” are relatively trustworthy because the source of information is more authoritative or professional. It seems partially consistent with previous findings (González et al. 2017) that uncertainty epistemic markers and direct evidential were preferred to be used in oral conversation, while written discourse was associated with a relatively high degree of certainty. However, further comparative studies on spoken Chinese are needed before we conclude that the information conveyed by “yiwei” is more opinionated. Because the datasets adopted here were collected from newspapers, the written discourse, and the distribution of “yiwei” was dramatically less frequent than that of “renwei”.

Moreover, the relatively low degree of information trustworthiness attributed to “yiwei” could also be accounted for the polarity theory that negative attitudes towards entire propositions were associated with specific NPIs (e.g. “yuan”) exclusively collocating with “yiwei”. As Riemer and Dittmer (2016) discussed on computational treatments, the interaction between modality and negation was also proved in text understanding in the current study.

Then, this research found only “renwei” used as hedging in the political discourse. Its functions of hedging the certainty of information and mitigating strong statements were proved in Chinese political discourse. The effects of power relation on differentiating near synonyms were proved among epistemic verbs, which was also applied to the previous study on other Chinese near synonyms verbs (Wang and Huang, 2018).

Furthermore, the methodology may also contribute to the comparative analysis of near synonyms with a new perspective. That is, the variations in epistemic modality were thoroughly investigated in terms of multiple strategies, which took the factor of speakers’ cognitive status into consideration. This contribution perfectly meets the requirement of more precise processing of natural language rather than neglect the attitudes or implicatures of speakers.

In conclusion, the present study provided a new approach to detect the information trustworthiness of text understanding attributed to Chinese epistemic verbs. The near synonymous cognitive verbs each has its own inherent epistemic commitment;

yet a speaker can further enhance the (un) trustworthiness with both linguistic (adjuncts, (im) personal subjects, etc.) and extra-linguistic (power-relations, meta-information of speaker identities) cues. The empirical evidence may contribute to future annotation projects, especially for studies on Chinese near synonyms in NLP, as the ability to distinguish and annotate diverse information credibility has been necessary due to the overwhelming information nowadays.

Acknowledgments

We thank Dr. Xuefeng Gao for his invaluable contributions, and three anonymous reviewers for their constructive comments.

References

- Brown, Penelope, and Stephen C. Levinson. *Politeness: Some universals in language usage*. Vol. 4. Cambridge university press, 1987.
- Chafe, Wallace. "Evidentiality in English conversation and academic writing." *Evidentiality: The linguistic coding of epistemology* 20 (1986): 261-272.
- Chen, Keh-jiann, Chu-Ren Huang, Li-ping Chang, and Hui-Li Hsu. 1996. Sinica Corpus: Design Methodology for Balanced Corpora. In B.-S. Park and J.B. Kim. Eds. Proceeding of the 11th Pacific Asia Conference on Language, Information and Computation. Seoul: Kyung Hee University. pp.167-176. <https://www.aclweb.org/anthology/Y96-1018.pdf>
- Chen, Keh-Jiann, and Chu-Ren Huang. Academia Sinica Balanced Corpus. In Rint Sybesma (Ed.), *Encyclopedia of Chinese Language and Linguistics*. Brill. 2016.
- De Haan, Ferdinand. "Encoding speaker perspective: Evidentials." *Linguistic diversity and language theories* 379 (2005): 397.
- Fraser, Bruce. "Hedging in political discourse." *OKULSKA, U., CAP, P., Perspectives in Politics and Discourse, Capitolo* 8 (2010).
- González, Montserrat, Paolo Roseano, Joan Borràs-Comes, and Pilar Prieto. "Epistemic and evidential marking in discourse: Effects of register and debatability." *Lingua* 186 (2017): 68-87.
- Horn, Laurence. "On the semantic property of logical operators in English." *Published by Indiana University Linguistics Club, University of California Los Angeles* (1972).
- Horn, Laurence. "A natural history of negation." (1989).

- Horn, Laurence R. A natural history of negation. CSLI. 2001.
- Horn, Laurence R. "WJ-40: Implicature, truth, and meaning." *International review of pragmatics* 1.1 (2009): 3-34.
- Horn, Laurence R., and Yasuhiko Kato, eds. *Negation and polarity: Syntactic and semantic perspectives*. OUP Oxford, 2000.
- Hyland, Ken. "Writing without conviction? Hedging in science research articles." *Applied linguistics* 17.4 (1996): 433-454.
- Israel, Michael. "The pragmatics of polarity." *The handbook of pragmatics* (2004): 701-723.
- Koźbiał, Dariusz. "Epistemic Modality: A Corpus-Based Analysis of Epistemic Markers in EU and Polish Judgments." *Comparative Legilinguistics* 41.1 (2020): 39-70.
- Lakoff, George. "Hedges: A study in meaning criteria and the logic of fuzzy concepts." *Contemporary research in philosophical logic and linguistic semantics*. Springer, Dordrecht, 1975. 221-271.
- Markkanen, Raija, and Hartmut Schröder, eds. *Hedging and discourse: Approaches to the analysis of a pragmatic phenomenon in academic texts*. Vol. 24. Walter de Gruyter, 2010.
- Nuyts, Jan. "Modality: Overview and linguistic issues in Frawley." *The expression of Modality—Berlin—New York: Mouton de Gruyter* (2006).
- Palmer, Frank Robert. *Mood and modality*. Cambridge University Press, 2001.
- Penka, Doris, and Hedde Zeijlstra. "Negation and polarity: an introduction." *Natural Language & Linguistic Theory* 28.4 (2010): 771-786.
- Riemer, Manuel, and Livia Dittmer. "An introduction to the special issue." *Ecopsychology* 8.3 (2016): 163-166.
- Rubin, Victoria L., Elizabeth D. Liddy, and Noriko Kando. "Certainty identification in texts: Categorization model and manual tagging results." *Computing attitude and affect in text: Theory and applications*. Springer, Dordrecht, 2006. 61-76.
- Taweel, Abeer Q., Emad M. Saidat, Hussein A. Rafayah, and Ahmad M. Saidat. "Hedging in Political Discourse." *Linguistics Journal* 5.1 (2011).
- Speranza, John L., and Laurence R. Horn. "A brief history of negation." *Journal of Applied Logic* 8.3 (2010): 277-301.
- Su, Qi, Chu-Ren Huang, and Helen Kaiyun Chen. "Evidentiality for text trustworthiness detection." *Proceedings of the 2010 Workshop on NLP and Linguistics: Finding the Common Ground*. 2010.
- Xu, Ge, and Churen Huang. "Mining Chinese Polarity Shifters." *Workshop on Chinese Lexical Semantics*. Springer, Cham, 2015.
- Wang, Xiaowen, and Chu-Ren Huang. "From Near Synonyms to Power Relation Variations in Communication: A Cross-Strait Comparison of “Guli” and “Mianli”." *Workshop on Chinese Lexical Semantics*. Springer, Cham, 2018.
- Willett, Thomas. "A cross-linguistic survey of the grammaticization of evidentiality." *Studies in Language. International Journal sponsored by the Foundation “Foundations of Language”* 12.1 (1988): 51-97.