The Use of Perspective Markers and Connectives in Expressing Subjectivity: Evidence from Collocational Analyses

Yipu Wei   WEIYIPU@PKU.EDU.CN
School of Chinese as a Second Language, Peking University
Yiheyuan Road 5, 100871, Beijing, China

Jacqueline Evers-Vermeul   J.EVERS@UU.NL
Utrecht Institute of Linguistics OTS, Utrecht University
Trans 10, 3512 JK, Utrecht, The Netherlands

Ted J.M. Sanders   T.J.M.SANDERS@UU.NL
Utrecht Institute of Linguistics OTS, Utrecht University
Trans 10, 3512 JK, Utrecht, The Netherlands

Abstract

This study explores how subjectivity is expressed in coherence relations, by means of a distinctive collocational analysis on two Chinese causal connectives: the specific subjective *kejian* 'so', used in subjective argument-claim relations, and the underspecified *suoyi* 'so', which can be used in both subjective argument-claim and objective cause-consequence relations. On the basis of both Horn's pragmatic *Relation* and *Quality* principles and the Uniform Information Density Theory, we hypothesized that the presence of other linguistic elements expressing subjectivity in a discourse segment should be related to the degree of subjectivity encoded by the connective. In line with this hypothesis, the association scores showed that *suoyi* is more frequently combined with perspective markers expressing epistemic stance: cognition verbs and modal verbs. *Kejian*, which already expresses epistemic stance, co-occurred more often with perspective markers related to attitudinal stance, such as markers of expectedness and importance. The paper also pays attention to similarities and differences in collocation patterns across contexts and genres.

Keywords: collocational analysis, connectives, perspective marking, stance, subjectivity

1 Introduction

In everyday communication, speakers and writers often express their conclusions and feelings. For instance, instead of merely reporting objective causal relations between events in the real world, as in (1a), they frequently utter subjective relations, which involve someone's reasoning (Langacker 1990, Pander Maat & Sanders 2000, Verhagen 2005), as illustrated in (1b). Subjective relations are not observable in the real world; one needs to take into account another person's (e.g., the speaker's or another agent's) perspective (Sanders et al. 2009, 2012) to process the reasoning, and thus one needs to track the source of information. In other words, subjective relations concern the degree of

involvement of a locutionary agent or a *Subject of Consciousness* (Finegan 1995, Lyons 1977, Sanders et al. 2009).

(1)
    a. This restaurant is decorated with several art works of Mondriaan, *so* it attracts lots of fans of Modern art.
    b. This restaurant is decorated with several art works of Mondriaan, *so* its owner must be a fan of Modern art.

In order to communicate in a coherent way, speakers choose words to express the relations between consecutive discourse segments (Sanders et al. 1993: 94, cf. also Sanders & Spooren 2007, Schilperoord & Verhagen 1998). For instance, they can use connectives such as *so* and *therefore* to provide the reader with information on the type of coherence relation to be established, in this case a causal one (Britton 1994, Graesser & McNamara 2011, Mak & Sanders 2010, van Silfhout et al. 2014, 2015). Such information facilitates the reading process. It triggers faster processing of information immediately following the connective (Cain & Nash 2011, Cozijn et al. 2011, Sanders & Noordman 2000, van Silfhout et al. 2014, 2015) compared to the processing of that same information in unmarked relations.

As examples (1a) and (1b) illustrate, English *so* can be used in objective and subjective causal relations. It only marks the causal nature of the relation, and does not indicate the degree of subjectivity of the relation. However, certain connectives in other languages do code information about subjectivity. For example, some connectives are only used for objective relations, such as Dutch *daardoor* 'as a result' and Chinese *yin'er* 'as a result', as is illustrated in the Dutch (2a) respectively Chinese (3a) translations of (1a). By contrast, the Dutch connectives *want* 'because' and *dus* 'so' (Degand & Pander Maat 2003, Sanders & Spooren 2015, Spooren et al. 2010, Stukker & Sanders 2008, Verhagen 2005), and Mandarin Chinese *kejian* 'so' prototypically express subjective coherence relations (Li et al. 2013). This is illustrated by the Dutch (2b) respectively Chinese (3b) counterparts of the subjective relation in (1b). Just like English *so* in example (1a) and (1b), some connectives in other languages leave the subjectivity information underspecified, i.e. they can be used for both subjective and objective relations (e.g. Chinese *suoyi* 'so' in example (3a) and (3b)).

(2) Dutch
    a. Dit restaurant is versierd met diverse kunstwerken van Mondriaan, *daardoor* trekt het veel fans van moderne kunst.
    b. Dit restaurant is versierd met diverse kunstwerken van Mondriaan, *dus* de eigenaar moet wel een fan zijn van moderne kunst.
(3) Chinese
    a. Zhe jia canguan zhuangshi zhe hao ji fu Mengteli'an de huazuo, *yin'er/ suoyi* ta xiyin le henduo xiandai yishu mi.
    This CL restaurant decorate ASP(IPFV) CL Mondrian MOD painting, *as a result/ so* 3SG attract ASP(PFV) many modern art fan.
    b. Zhe jia canguan zhuangshi zhe hao ji fu Mengteli'an de huazuo, *kejian/ suoyi* ta de zhuren keneng shi yi ge xiandai yishu mi.
    This CL restaurant decorate ASP(IPFV) CL Mondrian MOD painting, *in conclusion/ so* 3SG MOD owner probably be a CL modern art fan.

The degree of subjectivity expressed by connectives is found to affect the processing of coherence relations. For instance, the Dutch subjective connective *want* 'because' leads to longer processing times directly after the connective compared to the Dutch objective connective *omdat* 'because' (Canestrelli et al. 2013). Such processing effects can be attributed to the difficulty of interpreting

subjectivity: the reader needs to track the source of information to interpret subjectivity. Specific subjective connectives such as *want* 'because' instruct the reader at an early stage that there is a coherence relation, and that the relation is subjective, before the entire sentence is processed. In terms of the information density, subjective connectives encode more information compared to underspecified connectives.

As the choice of connectives in examples (1) to (3) and the accompanying processing results illustrate, speakers and writers continuously have to decide how informative they should be in order to provide sufficient cues for others to comprehend them. At the same time, they should also avoid being too wordy. This tension has been systematically described by Horn's framework for pragmatic inference: his *Q* (Quality) *Principle* describes the need to 'make your contribution sufficient', the *R* (Relation) *Principle* describes the need to 'make your contribution necessary' (Horn 1984: 13). According to Horn (1984), speakers should find a balance between the speaker-based economy (saving the speaker's production efforts) and the hearer-based economy (saving the hearer's processing efforts).

A highly similar point has been made by the *Uniform Information Density Theory* (UID), which is about the speakers' strategy of choosing between alternative linguistic forms at several levels of linguistic representations: phonetic, syntactic, pragmatic, etc. (Frank & Jaeger 2008, Jaeger 2010, Levy & Jaeger 2007). The UID suggests that speakers modulate their word choice according to the amount of information in the utterance: full linguistic forms are more often used at the point where the content conveyed by the form is unexpected in its context, i.e. the point with a low probability and a high information density (for details, see Frank & Jaeger 2008). For instance, connectives can be omitted if the information they convey is highly predictable given other linguistic cues in the context (Asr & Demberg 2015). Through such modulation of word choices, the density of information of the utterance is kept at a uniform level – a roughly equal amount of information at each unit of the sentence (Levy & Jaeger 2007). The UID theory echoes Horn's pragmatic theory in the sense that both theories predict a modulated process of word selection to optimize communication.

In terms of discourse relations and connectives, these theoretical discussions raise the question as to which information is exactly conveyed by connectives, and how that information may become predictable given other cues in the context. Hence, it is worthwhile to explore, as is done in the current paper, which linguistic markers also provide information on the degree of subjectivity of a relation, and would thereby allow for a division of labor between connectives and segment-internal elements (see Hoek 2018; Hoek et al., 2018). If other markers already indicate the degree of subjectivity, this will reduce the need of information on subjectivity to be expressed at the connective. This seems to be the case for expressions such as *probably*, *surprisingly* and *according to Peter*, which are addressed as markers of stance (Biber et al. 1999, Conrad & Biber 2000), evaluation markers (Bednarek 2006, 2009, Thompson & Hunston 2000), or appraisals (Eggins & Slade 1997, Martin 2000). Conrad and Biber (2000) suggest three sub-types of stance markers (see Bednarek 2006, Bednarek 2009 and Thompson & Hunston 2000 for similar classifications):

  i.   Epistemic stance, which indicates how certain the speaker or writer is, or where the information comes from (e.g. *probably*, *according to the President*).
  ii. Attitudinal stance, which indicates feelings or judgements about what is said or written (e.g. *surprisingly*, *unfortunately*).
  iii. Style stance, which indicates how something is said or written (e.g. *honestly*, *briefly*.)

(Conrad & Biber 2000: 57)

Stance markers introduce the viewpoint of the speaker or other agents, and hence can be termed as *perspective markers* (Sanders & Redeker 1996). Perspective markers expressing epistemic stance show overlap with specific subjective connectives. Both indicate subjective reasoning, either from the speaker or from a character. Canestrelli et al. (2013) and Traxler et al. (1997) found that the

processing effects of connectives are influenced by epistemic stance markers: by adding *volgens Peter* 'according to Peter' to the first clause connected in a subjective relation, as in example (2c), the extra processing time associated with the subjective connective *want* 'because' disappears.

(2)

    c. *Volgens* Peter is de eigenaar van dit restaurant een fan van moderne kunst, *want* het restaurant is versierd met diverse kunstwerken van Mondriaan.
    According to Peter the owner of this restaurant is a fan of Modern art, *because* the restaurant is decorated with several art works of Mondrian.

In terms of Horn's pragmatic theory, the reader/hearer has obtained sufficient information about the degree of subjectivity by the introduction of epistemic perspective markers. Upon encountering the subjective connective the reader/hearer does not have to establish an entirely new subjective mental representation, but rather only has to make a link to an already established mental representation introduced by the perspective marker in the first clause. In other words, epistemic stance markers in the first clause make it clear that the first clause is a claim and thereby create the expectation that the next clause will be an argument for this claim.

The empirical findings of Canestrelli et al. (2013) and Traxler et al. (1997) suggest an overlap between specific subjective connectives and perspective markers in their function of instructing readers on the degree of subjectivity of the relation. The question is whether this holds true for perspective markers in general, including all types of stance markers, or only pertains to markers of epistemic stance. Epistemic stance markers explicate the dimension of reliability/certainty and evidentiality, which directly introduces a source of information. However, attitudinal stance markers and style stance markers introduce a source in an indirect way: by indicating attitudes, feelings and styles of writing/speaking that can be attributed to a source. Although all three types of stance markers presuppose a source of information, they differ in the way in which this source of information is involved. How these perspective markers overlap with connectives marking different degrees of subjectivity may shed light on the relation between subjectivity and perspective marking.

In this paper, we investigate this issue in natural language data. Starting from the assumption that language users will tend to avoid a doubling of information in terms of marking subjectivity in discourse relations, we may expect authors/speakers to observe some pragmatic strategies (e.g. apply Horn's *R principle* or try to produce an information flow with a *Uniform Information Density*) to achieve a successful communication (both sufficient and necessary). Avoiding repetition of information in the same dimension fits the *R principle* as well as the UID. Therefore, in natural language data we may expect connectives marking different degrees of subjectivity to vary in their co-occurrence patterns with perspective markers.

In corpus linguistics, the method of collocational analysis (Evert 2008; Gries & Stefanowitsch 2004) provides insightful information on the context of given linguistic elements. It measures the association strengths between words or expressions, and produces a list of important collocates in attraction or repulsion with a target word. Collocational analysis can advance our knowledge about the properties of a connective on the basis of its contextual features. We therefore conducted a corpus-based study using collocational analyses to examine the use of connectives and perspective markers in discourse, aiming to answer the following research questions:

1) Do connectives of different subjectivity degrees differ in their types of collocates?
2) More specifically, do connectives differ in the types of perspective markers they co-occur with?

We focused on two Chinese causal connectives for which we could derive hypotheses from the literature. *Kejian* 'so' is mostly used in the epistemic domain (Li et al., 2013), indicating that the causal reasoning arises from someone's mind; it encodes the epistemic stance apart from its

discourse function of causally connecting two segments. Such subjectivity information is underspecified with the generic connective *suoyi* 'so', which can be used in both objective and subjective relations (Li et al. 2013). On the basis of Horn's theory of speaker economy, *kejian* can be expected to co-occur less with perspective markers of the epistemic stance than *suoyi*. Since neither the specific subjective *kejian* nor the generic connective *suoyi* encode attitudinal or style stance, no differences in collocation tendencies are expected between the connectives for the other two types of perspective markers.

2   Method

We conducted a series of distinctive collocates analyses on the two Chinese causal connectives *suoyi* 'so' and *kejian* 'so', with the aim to investigate the contextual features of the two connectives. Regular collocational analyses allow researchers to calculate association strengths between target words and their collocates. Distinctive collocates analyses (Church et al., 1991) are a specific type of collocational analysis: they allow for a direct comparison of the contexts of two semantically similar words (a word pair), identifying collocates that prefer to appear in the context of one word over the other word from the pair. With this type of analysis, words with high association scores are not associated with the target word in a general sense, but only if they are attracted more to this target word than to a reference context (i.e., in this study the alternative connective). This type of analyses has become especially popular for lexical alternatives in specific constructions (i.e., distinctive collexeme analysis or distinctive collostructional analysis, see Gries & Stefanowitsch 2004; Stefanowitsch & Gries 2003). In the current study, we use this method to identify words that tend to 'sit' in the context of *suoyi* more often than in the context of *kejian* and vice versa, paying special attention to linguistic elements expressing subjectivity.

2.1   Sample of texts

We used a balanced modern Chinese corpus: the CCL corpus (Zhan et al. 2003), which covers a variety of written texts: fiction, newspapers, conferences, translated literature, blogs, etc. The total size of the CCL corpus is 581,794,456 characters.

We only investigated actual texts, which lead to the exclusion of dictionaries, and to make sure all the texts were homogeneous in terms of mode (written), we excluded the sources of oral texts (spoken), and TV (written to be spoken), etc. From the remainder of the corpus, we selected texts from three types of genres: narrative genres on the one hand, and informative and argumentative genres on the other. Narrative genres included literature, drama, biographies and fiction magazines; informative and argumentative genres included newspapers, legal documents, academic works of natural science and social sciences, governmental reports and other texts labeled as practical writing. The argumentative and informative texts were collapsed as the 'non-narrative genre', because of the low number of argumentative texts available in the CCL corpus.

From the afore-mentioned parts of CCL, we then generated two raw datasets: text files containing all the sentences with the words *suoyi* or *kejian*, with a search scope of 200 characters to the left and 200 to the right. This scope was much wider than the length of a sentence so that we would have enough contexts for the analysis on the intended discourse unit.

In line with the parameters of collocation (Gries 2013), we first decided to investigate words as the linguistic units of collocates. Because natural Chinese texts do not have spaces between words, we used the Chinese word segmentation tool *NLPIR-ICTCLAS* (Zhang et al., 2003; tag: ICT_POS_MAP_SECOND) to separate the word boundaries of characters in the text. In this segmentation system, white spaces were added between words, and words were tagged based on their semantic types. Meanwhile, punctuations such as commas, full stops, parentheses, colons were also marked with tags. The word segmentation tool thereby generated segmented and annotated texts for later analysis.

In terms of the distance between collocates, a collocate did not need to be directly adjacent to the connective. Any words appearing within one clause before or one clause after the connectives were considered collocates. Instead of adapting an arbitrary number of words as the context, we set the context of the target word in such a way that it was meaningful at the discourse level: discourse clauses were taken as the units for analysis.

2.2 Sample of connective fragments

From the two segmented datasets of all sentences containing *suoyi* or *kejian*, we compiled a sample of connective fragments. This step was necessary, because *suoyi* does not only occur as a connective, but can also be used in an inversion construction *zhisuoyi* 'why there is a consequence of'. For the word *kejian*, we can observe a clear grammaticalization process in progress (Liu & Yao 2011; Q. Zhang 2012). There are cases in which *kejian* is used as a verb, sometimes resulting in modified constructions such as *qingxi kejian* 'clearly can see', and there are cases where *kejian* is clearly a connective, or where the use of *kejian* is ambiguous. In order to exclude the clear verbal cases of *kejian* and all the inversion constructions *zhisuoyi*, we conducted a two-step screening process in our sampling. First, we restricted the sample of target items to cases preceded by a punctuation marker (namely comma, full stop, question mark, exclamation mark, semicolon, or ellipsis) in the software *AntConc_3.4.4.0* (Anthony 2016). This screening process filtered out verbal uses of *kejian* such as *qingxi kejian* 'clearly can see', as well as cases of *kejian* which are preceded by prepositional phrases such as *youci kejian* 'from this can see'. After the rough automatic screening process, 67,147 sentences with *suoyi* and 3,902 sentences with *kejian* were included for further analyses.

We then manually checked the remaining sentences marked by *kejian*, in order to exclude all other verbal instances of *kejian*. The verbal status of *kejian* could easily be derived from the absence of the main verb in the clause headed by *kejian*. For example, in (4), interpreting *kejian* as a connective with the meaning 'so' would only leave a noun phrase as the remainder of the second clause: *the status of German cars in the minds of Chinese*. By contrast, interpreting *kejian* as a verb 'can see', results in a grammatical clause, because in Chinese, the subject can be dropped. Hence, only full sentences such as (5) were included in the analyses of the connective use of *kejian*.

(4) Deguo chan de dazhong, Aodi and Benchi zhanyou hen da de bili, *kejian* deguo chan de qiche zai zhongguoren xinmuzhong de diwei.
Germany produce MOD Volkswagen, Audi and Benz occupy very big MOD proportion, *kejian 'from this can see'/*kejian 'so'* Germany produce MOD car in Chinese mind MOD status.
The German products Volkswagen, Audi and Benz take a big proportion (of Chinese market), *from this we can see/*so* the status of German cars in the mind of Chinese people.

(5) Yi ge neng zhide yi tou niu de jiaqian, *kejian* nashihou shiliu zai woguo haishi xihan wu.
One CL can worth one CL cow MOD price, *kejian 'so'* that-time pomegranate in our-country still-is rare thing.
One (pomegranate) was worth the price of a cow, *so* pomegranate was still very rare in our country at that time (in Ancient China).

All in all, the automatic and manual screening process excluded 20,096 cases of *suoyi* and 10,900 cases of *kejian*. Table 1 shows the resulting distribution of *suoyi* and *kejian* in the narrative and non-narrative texts in the sample.

Connectives	Narrative texts		Non-narrative texts	
	Retrieved from CCL	Used for analysis (%)	Retrieved from CCL	Used for analysis (%)
suoyi	34,641	29,077 (83.94%)	52,445	37,913 (72.29%)
kejian	2,494	752 (30.15%)	11,688	2,530 (21.65%)
Total	37,135	29,829 (80.33%)	64,133	40,443 (63.06%)

Table 1. Distributions (and percentage of actually used cases) of *suoyi* and *kejian* in two genres

2.3　Three sets of distinctive collocates analyses

The actual collocate analyses were conducted using the software *R* (R Core Team 2015) with the R package *mclm_0.1* (Speelman 2018). The method of distinctive collocates analysis was applied three times. We first applied it to a context of one clause before and one clause after the connective, irrespective of genre. To obtain a proper context containing exactly one clause before and one clause after the connective, we automatically searched the closest punctuation markers (including comma, full stop, question mark, exclamation mark, semicolon, and ellipsis) around the target (a connective preceded by a punctuation marker). With this first analysis, we obtained a general picture of the words in collocation with one connective compared to the other. Second, we explored the collocates of the two connectives in their preceding context and following context separately, so that contextual features could be located more precisely.

However, the distinctive collocates of *suoyi* versus *kejian* may be different depending on the genre they appear in, because the narrative genre is supposed to be more descriptive (e.g., describing events and actions), while the non-narrative genre is expected to be more argumentative. Therefore, in the third analysis, we took genre into account, distinguishing the collocational patterns in the narrative genre on the one hand, and in the informative and argumentative genres on the other.

The attraction and repulsion strength between a given word and the target connectives are measured by association scores (Evert 2008, Gries 2013), which are calculated on the basis of observed frequencies (O_{11}, O_{12}, O_{21}, O_{22}) and expected frequencies (E_{11}, E_{12}, E_{21}, E_{22}) in a contingency table (Table 2). For each word (target word) in the corpus that appeared at least once in the context of *kejian* or *suoyi*, the *mclm* R package computes O_{11} (i.e. target word instances in the target context), O_{12} (non-target words in the target context), O_{21} (target word instances in non-target contexts), and O_{22} values (occurrence of non-target words in non-target contexts), as well as the corresponding expected frequencies.

	Observed frequencies			Expected frequencies	
	Target-word	**Non-target word**	**Totals**	**Target-word**	**Non-target word**
suoyi-context	O_{11}	O_{12}	R_1	$E_{11}=R_1C_1/N$	$E_{12}=R_1C_2/N$
kejian-context	O_{21}	O_{22}	R_2	$E_{21}=R_2C_1/N$	$E22=R_2C_2/N$
Totals	C_1	C_2	N		

Table 2. Contingency table (adapted from Evert 2008: 1231)

In the current study, we selected G_2 (the log-likelihood measure, $2\sum_{ij} O_{ij}\, log(O_{ij}/E_{ij})$, Evert 2008), a statistical measure that is one of the most frequently used measures in collocational analyses. It is robust for differences in sample size, and compares observed frequencies and expected frequencies for each of the words taking into account the amount of evidence. We selected the top 100 items ranked according to G_2 values (see the Appendix). Since G_2 reports association strengths without indication of their direction, the top 100 collocates contains both words in strong attraction with the target word *suoyi* (i.e. in repulsion to *kejian*) and words in strong repulsion to *suoyi* (i.e. in attraction with the reference word *kejian*). The *dir* (direction) values provided by the R package

were used to judge whether a word was attracted to *suoyi* (positive) or repelled by *suoyi* and hence attracted by *kejian* (negative).

The Delta-P value ($O_{11}/(O_{11}+O_{12}) - O_{21}/(O_{21}+O_{22})$, Gries 2013) is an effect size measure and measures the difference between the observed frequency of the target word in one context and that in the other context. In this study, Delta-P value was used as a secondary criterion for the collocates: the words in attraction to *suoyi* all needed to be above the threshold of 0, and the words in repulsion to *suoyi* (the collocates of *kejian*) needed to be below this threshold (<0). The Delta-P measure was applied because it is considered more psycholinguistically realistic, as it takes into account the directionality of the collocation: 'whether the $word_1$ is more predictive of $word_2$ or the other way round' (Gries 2013: 141).

An efficient and common way to interpret the outcomes of a collocational study is to cluster the collocates manually and draw meaningful interpretations based on these clusters (see Gries & Stefanowitsch 2010). In the current study, we were able to identify seven clusters within the top 100 collocates: pronouns, communication verbs, cognition verbs, modal verbs, and three types of perspective markers, namely exclamatory adverbials, expressions of expectation and expressions of importance. The results section will focus on these items; a full list of collocates for each distinctive collocates analysis can be found in the Appendix.

3   Results

In this section, we discuss the results of the three distinctive collocates analyses. Section 3.1 illustrates the general collocation patterns of the two connectives. Section 3.2 compares the collocates of the two connectives in the clause preceding the connective and the clause following the connective. A genre-specific analysis shown in Section 3.3 reveals the collocations in different genres.

3.1   General analysis

The top 100 collocates (either attracted by *suoyi* or attracted by *kejian*) were categorized according to their semantic types. Since our goal was to find out whether language users avoid overlap in the expression of subjectivity in their utterances, we checked the top 100 for linguistic elements that can be related to subjectivity and perspective marking. Table 3 shows the collocates of *suoyi* that are relevant to our discussion, with their observed and expected frequencies and the G_2 scores indicating the distinctiveness of particular collocates in the context of *suoyi* compared to the context of *kejian*.

From the top 100, certain types of words stood out as significant collocates of *suoyi*, the connective that is underspecified in terms of subjectivity. An important cluster is formed by pronouns of all types (singular and plural, 1_{st}, 2_{nd} and 3_{rd} person). On the one hand, pronouns can be linked to objective relations in which actors carry out certain actions for certain reasons. On the other hand, they can be used in subjective relations in which the pronouns refer to the individuals whose perspective is presented. Therefore, we are not sure whether the higher number of occurrences of pronouns in the context of *suoyi* compared to the context of *kejian* should be attributed to a contextual feature of the objective relations that *suoyi* can express, or to the tendency to avoid doubling of subjectivity information in the context of *kejian*.

This is much clearer for the other clusters that are attracted by *suoyi*, but repulsed by *kejian*: communication verbs, cognition verbs and modal verbs. Both communication verbs and cognition verbs can express the epistemic stance of the speaker, to be specific, the evidentiality of the information. Modal verbs indicate the author's/character's degree of certainty towards the proposition, which is also one of the dimensions of epistemic stance. This observation can be accounted for in terms of subjectivity. With *suoyi* in the sentence, the subjectivity information is underspecified. If subjectivity needs to be expressed, cognition verbs (marking evidentiality) and modal verbs (marking certainty) are used to help readers/hearers track the source of information.

Collocates	Frequency (obs. vs exp.)	G_2
Pronouns		
wo 'I/me'	22802: 21785	1189.91
ta 'she/her'	8972: 8602	380.85
ni 'you'(singular)	8177: 7840	345.64
ta 'he/him'	24340: 23734	318.83
women 'we/us'	9212: 8864	314.23
tamen 'they'	7183: 6908	253.13
ziji 'self'	6104: 5905	146.47
nimen 'you'(plural)	1129: 1080	53.62
Communication verbs		
shuo 'say'	12301: 12050	103.46
gaosu 'tell'	856: 818	43.12
Cognition verbs		
xiang 'think'	3990: 3829	159.58
zhidao 'know'	3333: 3199	131.28
renwei 'believe'	2794: 2699	74.07
xiwang 'hope'	1213: 1161	55.65
juede 'feel'	1576: 1516	55.04
pa 'be afraid of'	901: 863	39.06
Modal verbs		
hui 'would'	8270: 8030	152.17
neng 'can'	8883: 8714	64.20
keneng 'may'	2711: 2628	55.84
yinggai 'should'	1526: 1469	51.18
keyi 'can'	4025: 3933	43.46
bixu 'have to'	2123: 2059	42.70

Table 3. Important collocates of *suoyi* from top 100

Cognition and modal verbs would be repetitive for readers/hearers, however, in *kejian* contexts. *Kejian* already implies someone is making the inference (normally, the speaker), so the use of cognition verbs and modal verbs would be a repetition of information on subjectivity.

The strong association between *suoyi* and the communication verb *shuo* 'say' is indicative of the pattern we try to establish. An alternative explanation would be that this collocation is due to the high frequency of the expression *suoyi shuo* 'so (I) say'. This expression has been segmented as two separate words by *NLPIR-ICTCLAS*, but in combination, it functions as a discourse marker that expresses the epistemic stance of the speaker. However, the cases in which *suoyi* and *shuo* are not intervened by any other linguistic elements, only account for 6.36% of the data (782 out of 12301 instances). This leaves many instances in which the communication verbs contributed to the expression of the epistemic stance of the speaker, as in example (6). Still, our data show that communication verbs were not exclusively used in epistemic contexts; they could also be used for reporting an objective description of real-world events, as in example (7). Therefore, we cannot be sure of the reason for the collocation of communication verbs and *suoyi*. This collocation pattern could be due to the speaker/author's strategy to avoid repetition of subjectivity information in subjective relations, just as for the cases with cognition verbs. Alternatively, communication verbs could be a feature of the context typical of the objective relations expressed by *suoyi*.

(6) Ta meiyou wudao yiyang meili de dongzuo, wangwang dongzuo de kaishi jius shi dadou de jieshu, *suoyi* Li Xiaolong **shuo**, Jiequandao juedui bu shiyi biaoyan.
It (Jiequandao 'Jeet Kune Do', a type of Chinese Kong Fu) NEG:have dance alike MOD motion, often motion MOD start just COP fight MOD end, *CONJ* NAME *said*, Jiequandao absolutely NEG suit performance.
It (Jeet Kune Do) does not have beautiful motions like a dance; the start of a motion is often the end of a fight, *so* Li Xiaolong *said* Jeet Kune Do is absolutely not suitable for performances.

(7) Ta **shuo** ta jiu shi changqi jianchi xialai, *suoyi* bai toufa zhijin dou bi ta de jiemeimen shao de duo.
3SGF said 3SGF just COP long:time insist down, *CONJ* till:now all compare 3SGF MOD sisters less MOD much.
She said that she just kept (the good habit) for a long time, *so* up till now she has much less white hair compared to her sisters.

Some of the words in the top 100 list were repelled by *suoyi* and should therefore be seen as distinctive for *kejian* instead of *suoyi*, as illustrated in Table 4. As mentioned in Section 2, we included all kinds of indications of subjectivity, irrespective of their grammatical categories. The noun *jiazhi* 'value' was clustered with the adjective *zhongyao* 'important', because *jiazhi* 'value' is often associated with evaluations that are made from a person's perspective.

Collocates	Frequency (obs. vs exp.)	G_2
Exclamatory adverbials		
duome 'how much'	147: 235	263.63
hedeng 'how much' (literary)	33: 75	164.21
Expressions of importance		
jiazhi 'value'	768: 843	84.06
zhongyao 'important'	1423: 1489	42.90
Expressions of expectation		
jing 'surprisingly'	245: 272	33.95

Table 4. Important collocates of *kejian* from top 100

The exclamatory adverbials, expressions of expectation and expressions of importance can be related to subjectivity: they indicate that someone's feeling or evaluation is involved, and that the hearer/reader is not merely dealing with a description of real-world facts. These collocational patterns indicate that language users do not necessarily avoid a doubling of information, as both *kejian* and these collocates express *that* subjectivity is involved. However, from this list of collocates of *kejian*, it can also be derived that language users do pay attention to the type of subjectivity information, in other words *how* the perspective of a speaker/character is involved. While the important collocates of *suoyi* (cognition verbs, communication verbs and modal verbs) could be related to epistemic stance marking, the important collocates of *kejian* – expectation markers and importance markers – can be related to attitudinal stance marking. Hence, there is no doubling of epistemic stance marking information, the crucial type of subjectivity expressed by the connective *kejian*.

3.2 Collocational analysis on different clauses

Given the general information on the contextual features in the analysis across clauses and genres, we obtained a basic understanding of the types of collocates that appear in the context of *kejian* and *suoyi*. However, we do not know from the overall analysis where these collocates appeared exactly

– do they appear in the clause preceding the connective, or do they appear in the clause following the connective? By precisely identifying the locations of different types of collocates, we can be more informed on how language users combine different linguistic cues to express subjectivity in discourse. Moreover, for further psycholinguistic experiments, collocation distributions by clause provide insights into how linguistic stimuli should be designed to closely reflect authentic linguistic data. The current section therefore elaborates on the distribution of collocates in different clauses. Table 5 and Table 6 show the collocates of *suoyi* and *kejian* we derived from the top 100 in preceding clauses and in following clauses.

Collocates	Preceding clause		Following clauses	
	Frequency (obs. vs exp.)	G₂	Frequency (obs. vs exp.)	G₂
Pronouns				
wo 'I/me'	12056: 11547	497.78	10777: 10277	740.49
ta 'she/her'	5023: 4816	193.22	3959: 3796	192.48
ni 'you'(singular)	4056: 3878	182.67	4128: 3974	155.00
ta 'he/him'	13497: 13154	171.33	10870: 10609	145.81
women 'we/us'	4721: 4534	166.11	4505: 4349	142.19
tamen 'they'	3826: 3685	113.05	3372: 3240	144.12
ziji 'self'	3501: 3382	86.09	2617: 2537	61.23
nimen 'you'(plural)	--	--	601: 575	32.74
Communication verbs				
shuo 'say'	--	--	4720: 4549	166.63
jiao 'call'	--	--	866: 824	65.78
ting 'listen'	--	--	542: 516	38.84
chengwei 'be stated as'	--	--	442: 421	33.77
wen 'ask'	--	--	413: 393	30.92
Cognition verbs				
mingbai 'understand'	372: 353	24.87	--	--
zhidao 'know'	2452: 2342	117.21	--	--
xiang 'think'	2168: 2072	98.91	1825: 1761	59.68
renwei 'believe'	1898: 1833	46.96	903: 870	32.49
pa 'be afraid of'	666: 636	32.87	--	--
liaojie 'understand'	560: 533	31.11	--	--
juede 'feel'	903: 870	26.87	675: 648	29.90
xiwang 'hope'	--	--	713: 680	46.61
gan 'dare'	--	--	521: 497	32.69
Modal verbs				
hui 'would'	4535: 4404	76.03	3747: 3639	76.24
keneng 'may'	1562: 1507	40.97	--	--
yinggai 'should'	501: 478	25.46	--	--
yiding 'must'	898: 858	41.32	--	--
xiande 'seem'	--	--	273: 259	29.38
neng 'can'	--	--	4621: 4513	57.71
bixu 'have to'	--	--	1452: 1410	29.55

Table 5. Important collocates of *suoyi* in preceding and following clauses

Most of the general collocation patterns also held in the analysis per clause except for the communication verbs. In both preceding and following clauses, pronouns, cognition verbs, modal verbs co-occurred with *suoyi*. Most of these perspective markers may serve as the supplement of

subjectivity information supplied by *suoyi*, regardless of whether they appear before or after the connective. Examples (8) and (9) illustrate the combined use of *suoyi* 'so' and the perspective marker *renwei* 'believe', which can appear in both the clause before and the clause after the connective.

(8) Rensheng nande you tongtongkuaikuai xiangshou de rizi, ni cuoguo le, jianglai lao le shi, xiang xiangshou dou meiyou nengli, yanli bugou, meiyou yachi, tingjue you buhao, ni xiang qu xiangshou yixia, ye libucongxin, *suoyi* wo **renwei** yinggai chen nianqing de shihou, jishi xingle.
Life difficult have joyful enjoy MOD day, you miss ASP(PFV), future old ASP(PFV) time, want enjoy even NEG:have ability, eye NEG:enough, NEG:have tooth, hearing also NEG:good, you want go enjoy a:bit, also incapable, *CONJ* I **believe** should when young MOD time, in:time enjoy:life.
It's difficult to have joyful days to enjoy. If you miss them, you cannot enjoy them anymore when you get old: with failing eyesight, few teeth, and defective hearing, it is impossible to enjoy even for a little bit, *so* I **believe** (we) should enjoy life at youth.

(9) Ta zi **renwei** wancheng qingzang gaoyuan de senlin hangkong kance renwu shi ta yiburongci de zeren, *suoyi* ta bu gu ziji de shenti.
3SGM self **believe** complete Qinghai-Tibet Plateau aviation survey task COP 3SGM unshirkable duty, *CONJ* 3SGM NEG care self MOD health.
He **believes** that completing the aviation survey task in Qinghai-Tibet Plateau is his unshirkable duty, *so* he does not care about his own health.

An important difference with the general collocation pattern is that the communication verbs appeared as important collocates of *suoyi* only in the clauses following this connective. This means that for the co-occurrence with such reportative verbs, no significant difference between *suoyi* and *kejian* can be found in the first clause.

A clear-cut difference between the collocates of *kejian* in preceding and following clauses is suggested in Table 6. Exclamatory adverbials and expressions of importance were only distinctive for *kejian* in the clauses following this connective. This finding may be due to the tendency to express an evaluation in the second clause in a forward causal relation: the evaluation of importance is expressed in the second clause based on the events/phenomena described in the first clause, as is illustrated in (10).

(10) Huaiyun hou jiaolü bu'an de muqin geng rongyi nanchan he shengchu yichang de haizi, *kejian* yunqi zhong zhuyi xinli weisheng shi **duome zhongyao**.
Pregnant after anxious disturbed MOD mother more easy dystocia and deliver abnormal MOD infant, *CONJ* pregnancy middle pay:attention:to mental health is **so:much important**.
Mothers who are anxious and disturbed after pregnancy are more likely to suffer dystocia and deliver abnormal infants, *so* paying attention to mental health is **very important** during pregnancy.

Expressions of expectation only appeared as important collocates of *kejian* in the preceding clause. These linguistic elements express an attitude of the speaker towards the situation described in the first clause, such as in example (11): the author is surprised by the fact that Wang Jian, a general, won the battles both in the south and in the north.

Collocates	Preceding clause		Following clauses	
	Frequency (obs. vs exp.)	G₂	Frequency (obs. vs exp.)	G₂
Exclamatory adverbials				
duome 'how much'	--	--	67: 153	343.15
hedeng 'how much' (literary)	--	--	16: 60	207.58
xiangdang 'considerably'	--	--	269: 303	49.06
Expressions of expectation				
jing 'surprisingly'	110: 143	69.67	--	--
juran 'unexpectedly'	44: 59	34.71	--	--
jingran 'surprisingly'	30: 41	27.07	--	--
Expressions of importance				
zhongyao 'important'	--	--	757: 841	109.15
juzuqingzhong 'crucial'	--	--	3: 9	29.39
jiazhi 'value'	432:463	26,61	336:380	64.20
Communication verbs				
cheng 'state'	231: 265	47.84	--	--
yue 'say' (formal)	42: 57	35.93	--	--
yan 'speak'(formal)	165: 185	25.00	--	--

Table 6. Important collocates of *kejian* in preceding and following clauses

(11) Wang Jian **jingran** neng zai nanbei liang fang de zuozhan zhong dou qusheng, *kejian* qi zai yongbing fangmian yingdang shi shuyu quanfangwei de wujiang.
Wang Jian (a general in Chinese history) **surprisingly** can at south:north two side MOD battle in all win, *kejian* 3SGM at military aspect should COP belong:to extensive MOD general.
Surprisingly Wang Jian won the battles both in the south and in the north, *so* he should be a general with extensive military capabilities.

In contrast to the general collocation pattern, some instances of communication verbs were found as important collocates of *kejian* instead of *suoyi* in the preceding clause. Most of them are formal expressions, which are more characteristic of formal contexts such as informative and argumentative texts. Compared to the findings in Table 5 and 6, communication verbs can be collocates of either *suoyi* or *kejian*, depending on the formalities encoded in different specific communication verbs. Formal communication verbs such as *cheng* 'state' and *yue* 'say' patterned with *kejian*, while informal communication verbs such as *shuo* 'say' patterned with *suoyi*. Therefore, it is not possible to identify a uniform pattern in the co-occurrence of communication verbs in relation to the degree of subjectivity expressed by the connective.

3.3 Collocational analysis on different genres

The results discussed so far may be the result of a confound with the genre preference of the connectives under investigation. *Suoyi* is a generic connective that can be used for all types of genres, while *kejian* is not frequent in narrative texts (cf. Table 1). Moreover, several of the collocate clusters found in Section 3.1 and 3.2 may be a side-effect of genre preferences as well. For example, communication verbs can be expected to appear more in the narrative genre, just like pronouns. Therefore, communication verbs and pronouns may pattern with *suoyi* simply because they all share the preference for the narrative genre. To neutralize the influence of genre as a confounding factor, we further examined the collocation of the two connectives in different genres,

namely narratives and non-narratives. The collocation distributions of these two connectives with other linguistic elements in different genres are summarized in Table 7 and Table 8.

Collocates	Narratives		Non-narratives	
	Frequency (obs. vs exp.)	G_2	Frequency (obs. vs exp.)	G_2
Pronouns				
wo 'I/me'	16329: 16043	241.62	6473: 6104	407.51
women 'we/us'	3503: 3434	69.49	5709: 5424	257.24
tamen 'they'	3744: 3673	67.21	3439: 3279	129.18
ta 'she/her'	7614: 7514	58.08	1358: 1290	62.14
ta 'he/him'	16328: 16196	43.53	8012: 7868	37.47
ni 'you'(singular)	6439: 6368	33.57	1738: 1632	129.07
ziji 'self'	3202: 3155	31.68	2902: 2789	72.55
ta 'it'	--	--	4099: 3979	54.42
nimen 'you'(plural)	--	--	339: 316	34.80
Communication verbs				
shuo 'say'	--	--	6623: 6433	83.64
chengwei 'be stated as'	--	--	444: 418	28.38
gaosu 'tell'	--	--	313: 293	27.29
Cognition verbs				
xiang 'think'	2776: 2733	30.99	1214: 1156	49.56
zhidao 'know'	2504: 2463	30.56	829: 794	24.15
renwei 'believe'	--	--	1928: 1837	76.42
xiwang 'hope'	--	--	713: 673	44.12
juede 'feel'	--	--	497: 470	26.93
Modal verbs				
hui 'would'	4043: 3990	30.82	4227: 4078	84.80
neng 'can'	3258: 3223	16.11	5625: 5477	58.89
bixu 'have to'	545: 533	15.05	1578: 1511	47.52
zhineng 'can only'	240: 233	13.56	--	--
keneng 'may'	876: 860	13.46	1835: 1758	54.43
yinggai 'should'	--	--	898: 853	41.34
keyi 'can'	--	--	2492: 2415	37.27

Table 7. Important collocates of _suoyi_ in different genres

Pronouns were observed as important collocates of _suoyi_ in both types of genres, which indicated that this collocation pattern is not a side-effect of the genre preference of _suoyi_. Cognition verbs also appeared as important collocates of _suoyi_ in both types of genres. Although the exact collocates differ per genre, they all expressed the same cognitive state of knowing and thinking. In addition, modal verbs were still distinctive collocates for _suoyi_ in both narratives and non-narratives. Therefore, we may infer that the collocation of the generic connective _suoyi_ with cognition verbs and modal verbs is not due to genre differences. We did find a difference with the general collocation pattern, however. Contrary to our hypothesis communication verbs were found not to be significant collocates of _suoyi_ in narratives, although they were still important collocates in non-narratives (559 cases (8.44%) of which were instances of _suoyi_ immediately followed by _shuo_). Apparently, _suoyi_ and _kejian_ do not differ in their preference for co-occurring with communication verbs in the narrative genre, but only in the non-narrative genre.

Collocates	Narratives		Non-narratives	
	Frequency (obs. vs exp.)	G₂	Frequency (obs. vs exp.)	G₂
Exclamatory adverbials				
duome 'how much'	102: 124	62.78	45: 112	238.76
hedeng 'how much' (literary)	--	--	15: 53	152.93
Expressions of expectation				
juran 'unexpectedly'	68: 79	25.59	--	--
jing 'surprisingly'	140: 151	16.78	105: 124	25.97
guoran 'as expected'	28: 34	16.69	--	--
Expressions of importance				
jiazhi 'value'			660: 723	57.04

Table 8. Important collocates of *kejian* in different genres

Even though the ratios of observed versus expected frequencies differ from the ones in the general analysis, the top 100 items still display similar collocation patterns for *kejian*. As Table 8 indicates, exclamatory adverbials and expressions of expectations stayed distinctive for *kejian* in both narratives and non-narratives, although there were some differences per item. These perspective markers on the attitudinal stance dimension of expectedness are more associated with *kejian* rather than *suoyi* across genres. Expressions of importance only appeared as important collocates of *kejian* in non-narrative genres.

4   General Conclusion and Discussion

The current study explored whether language users try to avoid doubling of subjectivity information in discourse, specifically in coherence relations. On the basis of distinctive collocates analyses, we examined whether the Chinese connectives *kejian* and *suoyi*, which differ in the degree of subjectivity they express, differed in their types of collocates, and especially if they differed in the types of perspective markers they co-occurred with. In line with our predictions, the degrees of subjectivity encoded in the two connectives was related to the type of linguistic cues in their contexts. In Section 4.1, we will summarize and discuss the general patterns we found; in Section 4.2, we will discuss our main findings per clause (preceding or following the connective) and genre, and in Section 4.3, we discuss the limitations of our study and put forward some suggestions for future research.

4.1   General collocation patterns in line with pragmatic principles and UID

In general, the underspecified connective *suoyi* 'so', which can express both subjective and objective relations, patterned with more occurrences of cognition verbs and modal verbs in comparison to the specific subjective connective *kejian* 'so'. In the context of *kejian*, we found more exclamatory adverbials, expressions of importance and expressions of expectation compared to the context of *suoyi* as a reference level.

The collocation results showed that perspective markers as a general type of linguistic cues marking subjectivity can be used in combination with either of the two causal connectives. However, if perspective markers are specifically categorized into sub-types with regards to various dimensions of subjectivity, different collocation patterns surfaced. *Suoyi* turned out to collocate with epistemic stance markers more often, while *kejian* co-occurred with attitudinal stance markers.

The collocation pattern of epistemic stance markers and *suoyi* is consistent with Horn's pragmatic theory of *Relation principle* (reducing the speaker's production effort) and *Quality principle* (reducing the hearer's comprehension effort). From the perspective of the *R principle*, if subjectivity information on the epistemic stance (including (un)certainly and evidentiality) is already specified in the connective *kejian*, epistemic stance markers in the context of the connective

are redundant, i.e. not efficient from the speaker economy account. *Suoyi*, by contrast, does not provide sufficient information on the epistemic stance, and the use of epistemic stance markers therefore provides valuable information that compensates the lack of subjectivity information in *suoyi*. The *Q principle* is observed and hearers/readers' comprehension process should be facilitated.

The collocation results can also be well explained by the *Uniform Information Density Theory* account. With the two alternative connectives expressing discourse coherence, the presence of epistemic stance markers (e.g. cognition verbs, modal verbs) makes the content of the context highly expectable (high probability and low information), which is why it is more likely to have an underspecified connective, *suoyi* in this case. On the other hand, utterances with fewer occurrences of epistemic stance markers make the content conveyed by the context unexpected (low probability and high information). In this sense, the use of a specific connective is preferred. The prevalence of epistemic stance markers in the context of *suoyi* and their lower co-occurrence with *kejian* fit the need for a uniform information density throughout the sentence in terms of subjectivity. Optimal information density is realized in this way.

However, speakers/authors did not avoid overlap in the expression of subjectivity at all costs. Some attitudinal stance markers such as *jingran* 'surprisingly' and *zhongyao* 'important', which also indicate the involvement of a speaker responsible for an evaluation, occurred as important collocates in the context of *kejian*. Both epistemic stance markers and attitudinal stance markers express *that* a source of information is involved. Apparently, in their use of *kejian*, which also indicates *that* a source of information is involved, speakers and writers do not avoid overlap with that same information provided by attitudinal stance markers. However, *kejian* does not overlap with attitudinal stance markers in the subjectivity dimension it expresses, i.e. in expressing *how* the source of information is involved. The fact that *kejian* patterns with attitudinal but not with epistemic stance markers indicates that language users try to avoid overlap in the expression of these dimensions. The connective *kejian* and epistemic stance markers both indicate how certain the speaker/writer is about the information, while attitudinal stance markers express the attitude or feelings of a person towards the information. In terms of UID, the combination of attitudinal stance markers and *kejian* does not create high information density in the utterance. Taken together, the two observations above help to explain why attitudinal stance markers and *kejian* were found in collocation; they show a kind of agreement of subjectivity at the discourse level, jointly contributing to a subjective context.

4.2 Collocation patterns in different genres and clauses

To test for potential genre influences on the results, we performed collocational analyses on narratives and non-narratives separately. These analyses showed that communication verbs patterned with *suoyi* in the non-narrative genre, but not in narratives. This asymmetry might be due to different usage patterns of communication verbs in these genres. As illustrated in Section 3.1, communication verbs can be used to express an epistemic stance, as in example (6), or as a reportative verb to introduce a description of real-world events, as in example (7). In the non-narrative genre, we would expect a higher frequency of epistemic communication verbs. The fact that communication verbs stood out as important collocates of *suoyi* in this genre is in line with the avoidance of doubling of information (as illustrated in Table 7): *suoyi* has more needs of epistemic markers to strengthen the epistemic nature of the utterance than *kejian*, which encodes such information by itself. In narratives, with their abundance of descriptions of real-world events, however, we would expect a higher number of reportative communication verbs, which do not create a doubling of information with the information provided by *kejian* when they are used to report objective events in one of the clauses connected by *kejian*. This might explain why communication verbs do not stand out as collocates of *suoyi* in narratives, although this explanation needs to be corroborated in future corpus research on the actual usage of different types of communication verbs in narrative and non-narrative genres.

Apart from communication verbs, all other types of collocates of *suoyi* in the general analysis still surfaced as important in both narrative and non-narrative genres. Although the individual collocates of each cluster slightly vary per genre, the collocation between cognition verbs, modal verbs and pronouns with *suoyi* was robust across genres. As for *kejian*, exclamatory adverbials and expressions of expectation also appeared in both narratives and non-narratives as important collocates, which suggests that the perspective markers related to expectations were indeed an important contextual feature of *kejian*.

In order to locate the positions of each type of collocates in causal relations, we analyzed the preceding clauses and the following clauses of connectives separately. Most of the perspective markers as collocates of *suoyi* appeared in both the clauses preceding the connective and the clauses following it, except for communication verbs. The collocates of *kejian*, however, differed in the position they appeared in contexts. Expressions of expectation appeared as important collocates of *kejian* in the preceding clause, which makes sense because in a subjective relation with an argument-claim structure, expressions of expectation such as *jingran* 'surprisingly' in example (11) mark the speaker's surprisal – either about the propositional content of the clause preceding the connective, or about the fact as an argument-claim relation as a whole. Both exclamatory adverbials and expressions of importance tended to appear with *kejian* in the clause following the connective. These perspective markers served as expressions of the speaker's attitude towards the claim presented in the second segment of the relation.

Communication verbs exhibited very different collocation patterns depending on the clause they occurred in. In the preceding clause, formal communication verbs did not surface as important collocates of *suoyi*, but rather patterned with *kejian* more often (example (11)). In the following clause, the tendency was reversed – communication verbs only surfaced as collocates of *suoyi* such as in example (6). As example (12) shows, the communication verb *yue* 'say' co-occurred with *kejian* mainly in very formal texts, in which *kejian* was found more often. Therefore, such collocation pattern could be attributed to an effect of formality. On the basis of the current explorative study, we cannot draw a decisive conclusion on this issue. Further studies on the use of communication verbs in different contexts are needed, especially to find out whether our ideas about the formality and about the objective versus the subjective use of communication verbs can be corroborated.

(12) (Master Linji, a Buddhism master) da **yue**: ruguo yi kou qi bu lai, zhe routi haiyou ganqing ma? *Kejian* qinggan buzai routi shang, er zai lingxing shang.
(Master Linji) answer **say**: if one CL breath NEG come, this body have emotion? *CONJ* emotion NEG at body, but at spirituality on.
(Master Linji) **said** in response: if one doesn't breathe anymore, does the body still have emotions? *So* emotion is not in the body, but rather in the spirit.

4.3   Future studies and conclusion

In closing, we would like to discuss some limitations of this study. First, the current study only comprises a small set of connectives, which enabled us to provide an in-depth analysis. Future studies could extend this set to include other causal connectives. Interesting candidates for further distinctive collocates analyses seem to be *yushi* and *yin'er* (both meaning 'so/therefore'); Li et al. (2013) have shown that these causal connectives differ in the area of *volitionality* (i.e., is the causal relation an intentional one or not?). It would be interesting to see whether collocation patterns vary with this feature as well.

Second, the sentences in causal relations were retrieved directly from the corpus without any manual annotation of the relation type, which would have taken even more time and effort. *Kejian* is mainly used for subjective relations, while *suoyi* is generic (Li et al. 2013). This means that the sample of *suoyi* contexts contained both subjective and objective relations, while the contexts of *kejian* mainly consisted of subjective relations. The unbalanced distribution of relations in the

contexts of the two connectives may be a confounding factor. For instance, the fact that pronouns were distinctive for *suoyi* may be a feature of objective relations, because the descriptions of events and acts in objective relations may involve the use of pronouns. Nonetheless, the major findings such as the fact that modal verbs and cognition verbs are important collocates of *suoyi* are not characteristic of objective relations at all. We would expect stronger distinctive collocation patterns of these expressions with *suoyi* if we had limited the scope of the investigation to subjective relations only. More fine-grained analyses are expected to shed a clearer light on this issue.

Third, collocational analyses only provide rough tendencies in the word use in the context of a target word. It cannot support any decisive inferences, such as the predictability of one word given the other word. To further investigate the relation between connectives and their collocates, one could refer to regression analyses to investigate whether the presence of certain words in the context correlates with the presence of specific connectives, or one could opt for experimental research to investigate the effects of perspective markers on the processing of connectives.

A fourth limitation is that collocational analyses cannot distinguish word forms with the same syntactic tag but with multiple meanings. A relevant case in our corpus concerns modal verbs, some of which can be used in either a deontic or an epistemic way. Although both types of modals can have a subjective/interpersonal function (Lyons, 1977), they differ in the meaning they encode - deontic modals express obligations and permissions, while epistemic modals concerns believes (Foley & Van Valin, 1984; Johnson-Laird & Ragni, 2019; Verstraete, 2001). Our claims about modal verbs as linguistic means to express perspective would pertain to epistemic modals in particular. Examining whether the modal verbs in our corpus are used in the deontic or the epistemic way, however, would require a manual screening process. This seems to be an interesting avenue for further research.

Lastly, for practical reasons we made no distinction between argumentative and informative genres. These two genres have certain features in common in which they differ from narratives. For instance, both argumentative and informative genres have the author as the illocutionary force in most of the cases, while in narrative texts other characters are also frequently involved as the illocutionary force. However, argumentative genres also differ from informative genres in several respects – the use of communication verbs, for instance, may be different between argumentative texts and informative texts. Separate analyses of the argumentative genres and the informative genres can provide a more refined picture.

Despite these limitations, we can conclude that the explorative approach in the current study has produced a number of interesting insights into the way in which subjectivity is expressed in discourse. What is more, this study has illustrated the relation between connectives and perspective markers, demonstrating that this will be a productive area for future research to pursue.

Acknowledgements

Appendix. Top 100 collocates per analysis

1-25		26-50		51-75		76-100	
words	**dir**	**words**	**dir**	**words**	**dir**	**words**	**dir**
wo 'I/me'	1	zhan 'occupy'	-1	juede 'feel'	1	hao 'good'	1
yinwei 'because'	1	laodong 'work'	-1	bingfei 'not'	-1	zai 'again'	1
youyu 'since'	1	cai 'only'	1	nimen 'you'(plural)	1	yizhi 'always'	1
ta 'she/her'	1	yi 'already'	-1	gei 'give'	1	pa 'be afraid of'	1
ni 'you'(singular)	1	jiazhi 'value'	-1	yinggai 'should'	1	zong'e 'total amount'	-1
ta 'he/him'	1	woguo 'our country'	-1	dan 'but'	1	di 'emperor'	-1
women 'we/us'	1	gen 'and'	1	jizai 'record literally'	-1	zengzhang 'increase'	-1
duome 'how much' (excl. mood)	-1	lai 'come'	1	ziben 'capital'	-1	yin 'because'	1
tamen 'they'	1	rang 'let'	1	you 'further'	1	juan 'roll'	-1
jiu 'just'	1	renwei 'believe'	1	danshi 'but'	1	men (plural affix)	1
yao 'want'	1	shihou 'time'	1	zai 'at'	1	zhege 'this one'	1
bu 'no'	1	na 'that'	1	tebie 'especially'	1	wei 'for'	-1
dou 'all'	1	bisai 'race'	1	de (particle)	1	shangnian 'last year'	-1
hedeng 'how much' (literary)(excl. mood)	-1	shengyu jiazhi 'surplus value'	-1	xianzai 'now'	1	yige 'one'	1
xiang 'think'	1	tai 'too'	1	gongren 'worker'	-1	bujin 'not only'	-1
hui 'would'	1	qi 'that'	-1	bili 'proportion'	-1	ju 'according to'	-1
zhi 'of'	-1	de 'of'	1	keyi 'can'	1	mei (negation)	1
ziji 'self'	1	di (adverbial particle)	1	gaosu 'tell'	1	cidian 'dictionary'	-1
meiyou 'have not'	1	neng 'can'	1	zhongyao 'important'	-1	zhunbei 'prepare'	1
zhidao 'know'	1	zhe 'imperfective aspect marker'	1	bixu 'must'	1	kaishi 'begin'	1
qu 'go'	1	zuo 'make'	1	shangpin 'goods'	-1	shengchan 'produce'	-1
hen 'very'	1	yu 'left'	-1	zhi 'reach'	-1	fazhan 'develop'	-1
yuan 'Chinese dollar'	-1	daojiao 'Taoism'	-1	ba (disposal constr.)	1	ren 'people'	1
shuo 'say'	1	keneng 'may'	1	shu 'book'	-1	jing 'surprisingly'(short)	-1
le (perf. aspect marker)	1	xiwang 'hope'	1	jingji 'economy'	-1	zhao 'find'	1

Table 1. Top 100 collocates ranked by G2 (general context)

1-25		26-50		51-75		76-100	
words	dir	words	dir	words	dir	words	dir
yinwei 'because'	1	yu 'left'	-1	dui 'correct'	1	yinyong 'quote'	-1
youyu 'since'	1	juan 'roll'	-1	zhi 'reach'	-1	deng 'wait'	-1
wo 'I/me'	1	jizai 'record literally'	-1	shou 'first'	-1	bijiao 'compare'	1
shi 'is'	1	dan 'but'	1	jin 'now'	-1	chuan 'pass'	-1
de 'of'	1	yijing 'already'	1	juran 'unexpectedly'	-1	gongzuo 'work'	1
ta 'she/her'	1	yin 'because'	1	ju 'sentence'	-1	yan 'how'(literary)	-1
ni 'you' (singular)	1	cheng 'state'	-1	zengzhang 'increase'	-1	jiu 'just'	1
bu 'no'	1	danshi 'but'	1	yishang 'above'	-1	da 'reach'	-1
ta 'he/him'	1	renwei 'believe'	1	na 'that'	1	jingran 'surprisingly'	-1
women 'we/us'	1	zong'e 'total amount'	-1	tongji 'statistics'	-1	juede 'feel'	1
hen 'very'	1	zheng 'straight'	1	san 'three'	-1	shihou 'time'	1
yuan 'Chinese dollar'	-1	shu 'book'	-1	pa 'be afraid of'	1	jiazhi 'value'	-1
meiyou 'have not'	1	yiding 'for sure'	1	zhe 'imp.asp.marker'	1	shi 'family name'	-1
zhan 'occupy'	-1	keneng 'may'	1	chanzhi 'output value'	-1	dongxi 'things'	1
zhidao 'know'	1	buguo 'but'	1	ju 'according to'	-1	wen 'literal'	-1
tamen 'they'	1	gai 'change'	-1	chuban 'first edition'	-1	you 'further'	1
xiang 'think'	1	gen 'and'	1	yige 'one'	1	zai 'at'	1
tai 'too'	1	zhege 'this one'	1	nian 'year'	-1	yinggai 'should'	1
zhi 'of'	-1	ren 'people'	1	shi 'poem'	-1	xing 'gender'	1
dou 'all'	1	xihuan 'like'	1	liaojie 'understand'	1	chutu 'unearthed'	-1
ziji 'self'	1	que 'but'	-1	ji 'collection'	-1	ce 'volume'	-1
hui 'would'	1	bili 'proportion'	-1	jian 'see'	-1	ru 'if'	-1
yao 'want'	1	yue 'say' (literary)	-1	zhe 'this'	1	yin 'print'	-1
jing 'surprisingly'	-1	woguo 'our country'	-1	bisai 'race'	1	yan 'speak'(literary)	-1
wei 'for'	-1	shengyu jiazhi 'surplus value'	-1	yi 'upon'	-1	mingbai 'understand'	1

Table 2. Top 100 collocates in the preceding clause, ranked by G2

1-25		26-50		51-75		76-100	
words	dir	words	dir	words	dir	words	dir
wo 'I/me'	1	qi 'that'	-1	jingji 'economy'	-1	guanxi 'relation'	-1
duome 'how much' (excl. mood)	-1	gei 'give'	1	de (particle)	1	renwei 'believe'	1
shi 'is'	-1	rang 'let'	1	dao 'reach'	1	kaishi 'begin'	1
jiu 'just'	1	jiao 'call'	1	gen 'and'	1	you 'further'	1
hedeng 'how much' (literary, excl. mood)	-1	jiazhi 'value'	-1	shen 'very'	-1	juda 'huge'	-1
ta 'she/her'	1	zhi 'of'	-1	shehui 'society'	-1	wen 'ask'	1
shuo 'say'	1	ziji 'self'	1	ting 'listen'	1	yizhi 'always'	1
yi 'already'	-1	fazhan 'develop'	-1	shen 'deep'	-1	tebie 'especially'	1
ni 'you'(singular)	1	xiang 'think'	1	zuo 'make'	1	juede 'feel'	1
qu 'go'	1	neng 'can'	1	qiye 'company'	-1	zaiyu 'lie in'	-1
ta 'he/him'	1	daojiao 'Taoism'	-1	bing 'and'	-1	bixu 'must'	1
yao 'want'	1	yi 'one'	1	shangpin 'goods'	-1	zai 'again'	1
tamen 'they'	1	shihou 'time'	1	qianli 'potential'	-1	juzuqingzhong 'crucial'	-1
women 'we/us'	1	bian 'convenient'	1	zhao 'find'	1	miansha 'cotton'	-1
bingfei 'not'	-1	ci 'times'	1	zuoyong 'function'	-1	xiande 'seem'	1
zhongyao 'important'	-1	qing 'please'	1	shengyu jiazhi 'surplus value'	-1	zhe 'imp. aspect marker'	1
cai 'only'	1	shi 'time'	1	dang 'at'	1	liang 'good'	-1
lai 'come'	1	bisai 'race'	1	bujin 'not only'	-1	ziben 'capital'	-1
dangshi 'that time'	-1	woguo 'our country'	-1	de 'of'	-1	li 'inside'	1
di (adverbial particle)	1	xiangdang 'considerably'	-1	jiang 'be about to'	1	yongxin 'attentively'	-1
le (perf. asp. marker)	1	bei (passive)	1	yinsu 'reason'	-1	zhidu 'system'	-1
laodong 'work'	-1	xiwang 'hope'	1	diwei 'status'	-1	meiyou 'have not'	1
dou 'all'	1	na 'that'	1	chengwei 'be stated as'	1	dai 'take'	1
ba (disposal constr.)	1	xianzai 'now'	1	nimen 'you'(plural)	1	dui 'correct'	-1
hui 'would'	1	yijing 'already'	-1	gan 'dare'	1	zaoyi 'very early already'	-1

Table 3. Top 100 collocates in the following clause, ranked by G$_2$

1-25		26-50		51-75		76-100	
words	dir	words	dir	words	dir	words	dir
wo 'I/me'	1	meiyou 'have not'	1	zai 'again'	1	ren 'mercy'	-1
yinwei 'because'	1	jie 'session'	-1	gen 'and'	1	men (plural affix)	1
jiu 'just'	1	juran 'unexpectedly'	-1	qin 'diligent'	-1	jie 'all'	-1
women 'we/us'	1	wu 'no'	-1	xi 'to be informed of'	-1	an 'we'(oral)	-1
tamen 'they'	1	cai 'only'	1	liang 'good'	-1	fuhe 'conform to'	-1
zhi 'of'	-1	jian 'see'	-1	mingming 'apparently'	-1	Lin Yuxiang (name)	-1
duome 'how much' (excl. mood)	-1	xiang 'mutural'	-1	bingfei 'not'	-1	zhi 'only'	1
ta 'she/her'	1	de 'of'	1	haocheng 'be known as'	-1	diren 'enermy'	1
yao 'want'	1	ba (disposal construction)	1	shang 'even'	-1	dou 'all'	1
qi 'that'	-1	lajiao 'peper'	-1	Stalin (Name)	-1	he 'and'	1
ta 'he/him'	1	que 'but'	-1	shouji 'collect'	-1	ding 'vessel'	-1
youyu 'since'	1	qu 'go'	1	bujin 'not only'	-1	shenme 'what' (literary)	-1
rang 'let'	1	zou 'walk'	1	jing 'surprisingly'	-1	beibi 'mean'	-1
kejian 'so'	-1	zhe 'imperfective aspect marker'	1	guoran 'as expected'	-1	shengguo 'outrace'	-1
ni 'you'(singular)	1	yuanyi 'would like'	1	ji 'avoid'	-1	yongxin 'attentively'	-1
ziji 'self'	1	gei 'give'	1	neng 'can'	1	xinli 'psychology'	-1
zu 'block'	-1	hourou 'next generation'	-1	shen 'deep'	-1	zhineng 'can only'	1
shi 'family name'	-1	danshi 'but'	1	kandao 'see'	1	keneng 'may'	1
zai 'at'	1	biaodian 'punctuation'	-1	xie 'some'	1	Zhou Enlai (name)	-1
xiang 'think'	1	shen 'very'	-1	zhong 'heavy'	-1	anquanju 'security office'	-1
hui 'would'	1	du 'poison'	-1	fangyang 'dialect'	-1	juejin 'tunnelling'	-1
zhidao 'know'	1	di (adverbial particle)	1	la 'spicy'	-1	Shen Congwen (name)	-1
renyi 'righteousness'	-1	chunqiu 'spring and autumn'	-1	bixu 'must'	1	yongtu 'purpose'	-1
shihou 'time'	1	bu 'no'	1	dao 'reach'	1	wufa 'have no mean'	1
yuan 'source'	-1	lai 'come'	1	zi 'character'	-1	zhi 'reach'	-1

Table 4. Top 100 collocates in narrative genres, ranked by G$_2$

1-25		26-50		51-75		76-100	
words	dir	words	dir	words	dir	words	dir
youyu 'since'	1	laodong 'work'	-1	di 'emperor'	-1	yi 'also'	-1
yinwei 'because'	1	ta 'she/her'	1	qu 'go'	1	xunlian 'train'	1
wo 'I/me'	1	neng 'can'	1	yin 'because'	1	chuban 'first edition'	-1
women 'we/us'	1	shengyu jiazhi 'surplus value'	-1	xing 'gender'	1	juede 'feel'	1
duome 'how much' (exclamatory mood)	-1	jiazhi 'value'	-1	jizai 'record literally'	-1	ziben 'capital'	-1
hedeng 'how much' (literary, excl. mood)	-1	dan 'but'	1	chuan 'pass'	-1	chanzhi 'output value'	-1
yao 'want'	1	keneng 'may'	1	nimen 'you'(plural)	1	chang 'space'	1
dou 'all'	1	ta 'it'	1	bingfei 'not'	-1	miansha 'cotton'	-1
tamen 'they'	1	xiang 'think'	1	canjia 'participate'	1	chongdie 'overlap'	-1
ni 'you'(singular)	1	bixu 'must'	1	gongren 'worker'	-1	xianzai 'now'	1
de 'of'	1	dui 'group'	1	shangnian 'last year'	-1	jing 'surprisingly'	-1
bisai 'race'	1	aoyunhui 'Olympic game'	1	zai 'at'	1	he 'and'	1
bu 'no'	1	daojiao 'Taoism'	-1	kaishi 'begin'	1	yizhi 'always'	1
hui 'would'	1	xiwang 'hope'	1	di (adverbial particle)	1	shu 'book'	-1
yuan 'Chinese dollar'	-1	zuo 'make'	1	zhunbei 'prepare'	1	renhe 'any'	1
shuo 'say'	1	woguo 'our country'	-1	zong'e 'total amount'	-1	ce 'volume'	-1
jiu 'just'	1	bijiao 'compare'	1	ji 'collection'	-1	xianling 'shilling'	-1
zhan 'occupy'	-1	yinggai 'should'	1	juan 'roll'	-1	shengchanziliao 'production means'	-1
hen 'very'	1	feichang 'very much'	1	you 'further'	1	danshi 'but'	1
renwei 'believe'	1	cai 'only'	1	shijian 'time'	1	jinxing 'going on'	1
meiyou 'have not'	1	tebie 'especially'	1	zui 'most'	1	zhidao 'know'	1
ziji 'self'	1	shi 'is'	1	chengwei 'be stated as'	1	fubai 'corruption'	-1
zhi 'of'	-1	tai 'too'	1	zhexie 'these'	1	yiban 'ordinary'	1
yi 'already'	-1	ta 'he/him'	1	bili 'proportion'	-1	bei (passive)	1
yu 'left'	-1	keyi 'can'	1	gaosu 'tell'	1	lai 'come'	1

Table 5. Top 100 collocates in non-narrative genres, ranked by G$_2$

References

Laurence Anthony (2016). AntConc (Version 3.4.4.0) [Computer Software]. Tokyo, Japan, Waseda University. Retrieved from http://www.laurenceanthony.net/.

Fatemeh Torabi Asr and Vera Demberg (2015). Uniform information density at the level of discourse relations: Negation markers and discourse connective omission. In *Proceedings of the 11th International Conference on Computational Semantics*, pages 118-128, Association for Computational Linguistics, London.

Monika Bednarek (2006). Evaluation in media discourse: Analysis of a newspaper corpus. *Journal of Quantitative Linguistics*, 17(3), 253-256.

Monika Bednarek (2009). Dimensions of evaluation: Cognitive and linguistic perspectives. *Pragmatics & Cognition*, 17(1), 146-175.

Douglas Biber, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan (1999). *Longman grammar of spoken and written English*. London, Longman.

Bruce K. Britton (1994). Understanding expository text: Building mental structures to induce insights. In Morton A. Gernsbacher (ed.), *Handbook of psycholinguistics* (1st ed.): 641-674. San Diego, CA, Academic Press.

Kate Cain and Hannah M. Nash (2011). The influence of connectives on young readers' processing and comprehension of text. *Journal of Educational Psychology*, 103(2), 429-441.

Anneloes R. Canestrelli, Willem M. Mak, and Ted J.M. Sanders (2013). Causal connectives in discourse processing: How differences in subjectivity are reflected in eye-movements. *Language and Cognitive Processes*, 28(9), 1394-1413.

Reinier Cozijn, Leo G.M. Noordman, and Wietske Vonk (2011). Propositional integration and world-knowledge inference: Processes in understanding because sentences. *Discourse Processes*, 48(7), 475-500.

Kenneth W. Church, William Gale, Patrick Hanks, and Donald Hindle (1991). Using Statistics in Lexical Analysis. In Uri Zernik (ed.) *Lexical acquisition: Exploiting on-line resources to build up a lexicon*: 115-164. Hillsdale, NJ, Lawrence Erlbaum Associates.

Susan Conrad and Douglas Biber (2000). Adverbial marking of stance in speech and writing. In Susan Hunston and Geoff Thompson (eds.), *Evaluation in text: Authorial stance and the construction of discourse*: 56-73. Oxford, Oxford University Press.

Liesbeth Degand and Henk L.W. Pander Maat (2003). A contrastive study of Dutch and French causal connectives on the speaker involvement scale. In Arie Verhagen and Jereon van de Weijer (eds.), *Usage based approaches to Dutch*: 175-199. Utrecht, LOT.

Suzanne Eggins and Diana Slade (1997). *Analysing Casual Conversation*. London/ Washington, Cassell.

Stefan Evert (2008). Corpora and collocations. In Anke Lüdeling and Merja Kytö (eds.), *Corpus linguistics. An international handbook*: 1212-1248. Berlin, Mouton de Gruyter.

Edward Finegan (1995). Subjectivity and subjectivisation: An introduction. In Dieter Stein and Susan Wright (eds.), *Subjectivity and subjectivisation: Linguistic perspectives*: 1-15. Cambridge, Cambridge University Press.

William Foley and Robert Van Valin (1984). *Functional Syntax and Universal Grammar*. Cambridge: Cambridge University Press.

Austin F. Frank and T. Florian Jaeger (2008). Speaking rationally: Uniform information density as an optimal strategy for language production. In Bradley C. Love, Kateri McRae, and Vladimir M. Sloutsky (eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, pages 933-938, Cognitive Science Society, London.

Arthur C. Graesser and Danielle S. McNamara (2011). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science*, 3(2), 371-398.

Stefan Th. Gries (2013). 50-something years of work on collocations. *International Journal of Corpus Linguistics*, 18(1), 137-166.

Stefan Th. Gries and Anatol Stefanowitsch (2004). Extending collostructional analysis: A corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics*, 9(1), 97-129.

Stefan Th. Gries and Anatol Stefanowitsch (2010). Cluster analysis and the identification of collexeme classes. In Sally Rice and John Newman (eds.), *Empirical and Experimental Methods in Cognitive/Functional Research*: 73-90. Stanford, CSLI Publications.

Laurence R. Horn (1984). Toward a new taxonomy for pragmatic inference. In D. Schiffrin (ed.), *Meaning, form, and use in context: linguistic applications*: 11-42. Washington, D.C., Georgetown University Press.

T. Florian Jaeger (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, 61(1), 23-62.

Jet Hoek (2018). Making sense of discourse: *On discourse segmentation and the linguistic marking of coherence relations.* PhD thesis, Utrecht University. Utrecht, LOT. Available online at https://www.lotpublications.nl/Documents/509_fulltext.pdf.

Jet Hoek, Sandrine Zufferey, Jacqueline Evers-Vermeul, and Ted J.M. Sanders (2018). The linguistic marking of coherence relations: Interactions between connectives and segment-internal elements. *Pragmatics and Cognition*, 25(2), 275-309.

P.N. Johnson-Laird and Marco Ragni (2019). Possibilities as the foundation of reasoning. *Cognition*, 193, 103950.

Ronald W. Langacker (1990). Subjectification. *Cognitive Linguistics*, 1(1), 5-38.

Roger Levy and T. Florian Jaeger (2007). Speakers optimize information density through syntactic reduction. In Bernhard Schölkopf, John Platt, and Thomas Hofmann (eds.), *Advances in neural information processing systems* (1st ed.): 849-856. Cambridge, Mass, MIT Press.

Fang Li, Jacqueline Evers-Vermeul, and Ted J.M. Sanders (2013). Subjectivity and result marking in Mandarin: A corpus-based investigation. *Chinese Language and Discourse*, 4(1), 74-119.

Yahui Liu and Xiaopeng Yao (2011). Kejian de qingtaihua yu guanlianhua – jian lun hanyu liang lei shijueci de yanhua chayi [On the modalitization and relevantization of 'kejian']. *Hanyu Xuebao [Chinese Linguistics],* 4, 27-33.

John Lyons (1977). *Semantics*. Cambridge, Cambridge University Press.

Willem M. Mak and Ted J.M. Sanders (2010). Incremental discourse processing: How coherence relations influence the resolution of pronouns. In Martin Everaert, Tom Lentz, Hannah De Mulder, Øystein Nilsen, and Arjen Zondervan (eds.), *The linguistics enterprise: From knowledge of language to knowledge in linguistics*: 167-182. Amsterdam, John Benjamins.

James R. Martin (2000). Beyond exchange: Appraisal systems in English. In Susan Hunston and Geoff Thompson (eds.), *Evaluation in Text: Authorial Stance and the Construction of Discourse*: 56-73. Oxford, Oxford University Press.

Henk L.W. Pander Maat and Ted J.M. Sanders (2000). Domains of use or subjectivity? The distribution of three Dutch causal connectives explained. In Elizabeth Couper-Kuhlen and Bernd Kortmann (eds.), *Cause, condition, concession and contrast: Cognitive and discourse perspectives:* 57-82. Amsterdam, John Benjamins.

R Core Team. (2015). R: A language and environment for statistical computing (Version 3.1.3) [Computer Software]. Vienna, Austria, R Foundation for Statistical Computing. Retrieved from: http://www.R-project.org/

Ted J.M. Sanders and Leo G.M. Noordman (2000). The role of coherence relations and their linguistic markers in text processing. *Discourse Processes*, 29(1), 37-60.

José Sanders and Gisela Redeker (1996). Perspective and the representation of speech and thought in narrative discourse. In Gilles Fauconnier and Eve E. Sweetser (eds.) *Spaces, worlds, and grammar*: 290–317. Chicago, University of Chicago Press.

Ted J.M. Sanders, José Sanders, and Eve E. Sweetser (2009). Causality, cognition and communication: A mental space analysis of subjectivity in causal connectives. In Ted J.M. Sanders and Eve E. Sweetser (eds.), *Causal categories in discourse and cognition*: 19-60. Berlin, Mouton de Gruyter.

José Sanders, Ted J.M. Sanders, and Eve E. Sweetser (2012). Responsible subjects and discourse causality. How mental spaces and perspective help identifying subjectivity in Dutch backward causal connectives. *Journal of Pragmatics*, 44(2), 191-213.

Ted J.M. Sanders and Wilbert P.M.S. Spooren (2007). Discourse and text structure. In Dirk Geeraerts and Hubert Cuyckens (eds.), *Handbook of Cognitive Linguistics*: 916-941. Oxford, Oxford University Press.

Ted J.M. Sanders and Wilbert P.M.S. Spooren (2015). Causality and subjectivity in discourse: The meaning and use of causal connectives in spontaneous conversation, chat interactions and written text. *Linguistics*, 53(1), 53-92.

Ted J.M. Sanders, Wilbert P.M.S. Spooren, and Leo G.M. Noordman (1993). Coherence relations in a cognitive theory of discourse representation. *Cognitive Linguistics*, 4(2), 93-133.

Joost Schilperoord and Arie Verhagen (1998). Conceptual dependency and the clausal structure of discourse. In J.-P. Koenig (ed.), *Discourse and cognition: Bridging the gap*: 141-163. Chicago, University of Chicago Press.

Wilbert P.M.S. Spooren, Ted J.M. Sanders, Mike Huiskes, and Liesbeth Degand (2010). Subjectivity and causality: A corpus study of spoken language. In Sally Rice and John Newman (eds.), *Empirical and experimental methods in cognitive/functional research*: 241-255. Stanford, CSLI Publications.

Dirk Speelman (2018). *Mastering corpus linguistics methods: A practical introduction with Antconc and R*. Hoboken, NJ: Wiley.

Anatol Stefanowitsch, and Stefan Th. Gries (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, 8(2), 209-243.

Ninke Stukker and Ted J.M. Sanders (2008). Another('s) perspective on subjectivity in causal connectives: A usage-based analysis of volitional causal relations. In Liesbeth Degand, Cathrine Fabricius-Hansen, and Wiebke Ramm (eds.), *Linearisation and segmentation in discourse*: 113-122. Oslo, Oslo University.

Geoff Thompson and Susan Hunston (2000). Evaluation: An introduction. In Susan Hunston and Geoff Thompson (eds.), *Evaluation in text: Authorial stance and the construction of discourse*: 1-27. Oxford, Oxford University Press.

Matthew J. Traxler, Anthony J. Sanford, Joy P. Aked, and Linda M. Moxey (1997). Processing causals and diagnostics in discourse. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 23(1), 88-101.

Gerdineke van Silfhout, Jacqueline Evers-Vermeul, Willem M. Mak, and Ted J.M. Sanders (2014). Connectives and layout as processing signals: How textual features affect students' processing and text representation. *Journal of Educational Psychology*, 106(4), 1036-1048.

Gerdineke van Silfhout, Jacqueline Evers-Vermeul, and Ted J.M. Sanders (2015). Connectives as processing signals: How students benefit in processing narrative and expository texts. *Discourse Processes*, 52(1), 47-76.

Jean-Christophe Verstraete (2001). Subjective and objective modality: Interpersonal and ideational functions in the English modal auxiliary system. *Journal of Pragmatics*, 33(10), 1505-1528.

Arie Verhagen (2005). *Constructions of Intersubjectivity: Discourse, Syntax, and Cognition*. Oxford, Oxford University Press.

Weidong Zhan, Rui Guo, and Yirong Chen (2003), *The CCL Corpus of Chinese Texts: 700 million Chinese Characters, the 11th Century B.C. - present*, Available online at the website of Center for Chinese Linguistics (abbreviated as CCL) of Peking University, http://ccl.pku.edu.cn:8080/ccl_corpus.

Hua-Ping Zhang, Hong-Kui Yu, De-Yi Xiong, and Qun Liu. (2003). HHMM-based Chinese lexical analyzer NLPIR-ICTCLAS. In *Proceeding of the second SIGHAN workshop on Chinese Language Processing*, pages 184-187, Sapporo, Japan.

Qi Zhang (2012). 'Kejian' de yufahua [The grammaticalization of 'kejian']. *Journal of Hubei University of Education*, 29(5), 28-32.