



Hongling Xiao*, Fang Li, Ted J. M. Sanders and
Wilbert P. M. S. Spooren

***Suǒyǐ* ‘so’, they are different: an integrated subjectivity account of Mandarin RESULT connectives in conversation, microblog and newspaper discourse**

<https://doi.org/10.1515/ling-2021-0118>

Received April 25, 2019; accepted September 10, 2020; published online June 21, 2021

Abstract: In this study, we analyze the meaning and use of Mandarin causal connectives *kějiàn* ‘therefore/it can be seen that’, *suǒyǐ* ‘so’, *yīncǐ* ‘for this reason’, and *yúshì* ‘thereupon/as a result’ in terms of causality and subjectivity. We adopt an integrated approach to subjectivity and analyze the subjectivity profile of a causal construction in terms of three features: the propositional attitude of the consequent, the identity of the subject of consciousness (SoC), and the linguistic realization of the SoC. The investigation is based on natural discourse produced in fundamentally distinctive channels, namely, spontaneous conversation, microblogging, and formal writing. Compared to previous studies, the empirical foundation is therefore enlarged and more varied. The results show that these connectives differ systematically from each other with regard to the above three features, and that the differences remain robust across the three discourse types. The relative importance of each feature in characterizing the connectives is also determined. The propositional attitude appears to be the most important subjectivity feature, followed by the linguistic realization of the SoC and the identity of the SoC.

Keywords: causality; discourse types; Mandarin; RESULT connective; subjectivity

***Corresponding author: Hongling Xiao**, Faculty of Foreign Languages and Cultures, Xiamen University, South Siming Rd. 422, Xiamen, 361005, China; and Centre for Language Studies, Radboud Universiteit Nijmegen, Nijmegen, The Netherlands, E-mail: xhling@xmu.edu.cn
Fang Li, School of English Language, Literature, and Culture, Beijing International Studies University, Beijing, China, E-mail: lifang@bisu.edu.cn
Ted J. M. Sanders, Utrecht Institute of Linguistics, Universiteit Utrecht, Utrecht, The Netherlands, E-mail: t.j.m.sanders@uu.nl
Wilbert P. M. S. Spooren, Centre for Language Studies, Radboud Universiteit Nijmegen, Nijmegen, The Netherlands, E-mail: w.spooren@let.ru.nl

1 Introduction

Discourse is not an arbitrary set of utterances but the mental representation of the utterances that are interrelated, or coherent, in Hobbs' (1979) term. For example, people do not generally interpret the two segments in Example (1a) below as presenting two independent events. They identify a causal relation between the two: the first segment "it is going to be a nice day tomorrow" acts as the reason for the following segment "you are lucky". The causal relation can be more explicitly illustrated via (1b) or (1c).

- (1) a. *It is going to be a nice day tomorrow. You are lucky.*
 b. *It is going to be a nice day tomorrow, so you are lucky.*
 c. *You are lucky, because it is going to be a nice day tomorrow.*

A coherence relation is an aspect of meaning of two or more discourse segments that cannot be described in terms of the meaning of the segments in isolation (Sanders et al. 1992: 2). It is characterized in terms of the inferences that can be drawn between discourse segments (Hobbs 1979). Take (2a) for example, where several relations can be inferred between the two segments. It can be additive, as in (2b); it can be causal, as in (2c); it can also be concessive, as in (2d).

- (2) a. *Tom loves winter. He is a professional skier.*
 b. *Tom loves winter, and he is a professional skier.*
 c. *Tom loves winter, because he is a professional skier.*
 d. *Tom loves winter, but then, he is a professional skier.*

Coherence is a cognitive and not a linguistic entity in that creating a coherent interpretation does not necessarily depend on its linguistic realization in the discourse (Hobbs 1979; Knott and Dale 1994; Mann and Thompson 1988; Sanders and Spooren 2001; Sanders et al. 1992). Nonetheless, coherence markers have an important function in helping people build the mental representation of the relationship between discourse segments. In this study, we focus on one of the discourse coherence relations and the expressing coherence markers, namely, the causal relations and causal connectives (with which we mean the coordinating and subordinating conjunctions that express coherence relations at the discourse level).

Empirical studies using either corpus data of natural discourse or experimental methods have found that RESULT connectives such as *so* and *as a result* in several languages differ systematically in their prototypical meaning and use. The differences have been successfully modelled in terms of subjectivity. This holds, for instance, for *daardoor* 'as a result' vs. *daarom* 'that's why' vs. *dus* 'so' in Dutch, *as a result* vs. *for this reason* in English, *de ce fait* 'as a result' vs. *c'est pourquoi*

‘that’s why’ vs. *donc* ‘so’ in French, and *yúshì* ‘as a result’ vs. *kějiàn* ‘therefore’ in Chinese (e.g., Anderson 2016; Li et al. 2013; Pander Maat and Degand 2001; Pander Maat and Sanders 2000, 2001).

The linguistic categories of causal connectives show that language users distinguish between different types of causal relations. Focusing on the meaning and use of four Mandarin RESULT connectives¹ *kějiàn*, *suǒyǐ*, *yīncǐ*, and *yúshì*, the current study aims to investigate the system underlying the linguistic categorization of, as well as the distinctions between, the causal connectives. Though highly similar at both syntactic and semantic levels, these connectives are not freely interchangeable in actual language use. For example, the mutual substitution of *kějiàn* and *yúshì* in Examples (3) and (4) gives rise to serious acceptability problems.² Neither the semantic nature of the two connectives nor the semantics of the two relations allows for such an exchange. *Kějiàn* ‘therefore’, literally meaning ‘it can be seen that’, is intrinsically indicative of a personal perspective and/or motivation when drawing a conclusion. Hence, in (3), the reasoning from the antecedent (P) “we always describe a beauty as having big eyes and a small mouth” to the consequent (Q) “having a pair of big eyes seems to be a necessary feature of a standard beauty” is well acceptable, even in the absence of the epistemic marker *sīhū* ‘seem to’. *Yúshì* can be interpreted as ‘thereupon/as a result’. It usually introduces a consequence or an activity triggered by the situation in P, as can be seen in (4): that “the villagers nearby saw that [...]” leads to their action “came to find *Běiyīng Wēn* one after another, [...]”. When it comes to *suǒyǐ* ‘so/therefore’ and *yīncǐ*, literally meaning ‘for this [reason]’, it is doubtful whether they fit well in (3). Intuitively, the acceptability of *suǒyǐ* and *yīncǐ* is largely increased by the presence of *sīhū* ‘seem to’ in the Q-segment, which helps to soften the impact of the purely subjective flavor of reasoning from P to Q. The use of *suǒyǐ* and *yīncǐ* in (4) leads to two readings of the relation. One is from the perspective of the participants, “the villagers nearby”, who take the intentional action “came to find *Běiyīng Wēn* [...]” because “they saw that [...]”. In this case, the speaker acts as a reporter/narrator of the causal event (i.e., a pair of events that stand in a causal relation to each other). The other is from the perspective of the speaker, who acts as a conceptualizer herself and, based on her observation, concludes that “(the villagers nearby) came to find *Běiyīng Wēn* [...]” because “they saw that [...]”.

1 Other Mandarin RESULT connectives *cóng’ér*, *yīn’ér* and *yǐzhì* are not included in this study because of their low frequencies, especially in spoken and microblog discourse.

2 All the illustrative causal constructions in this paper are from our corpora (see Section 3.1 for detailed information). The symbol \emptyset in the Mandarin text stands for a phonologically null subject (*pro-drop*), which is recovered and put in parentheses () in the translated text. The asterisk * and the question mark ? indicate a problematic and a doubtful use of the connective, respectively.

- (3) P [Wōmen xíngróng měinǚ zǒng shuō dàyǎnjīng,
 1PL describe beautiful:woman always say big:eye
xiǎo zuǐbā,
 small mouth

P ‘We always describe a beauty as having big eyes and a small mouth,’

kějiàn / *yúshì / ?yīncǐ / ?suǒyǐ,
it can be seen/as a result/for this reason/so

- Q [yōngyǒu yīshuāng dàyǎn **sìhū** shì gòuchéng biāozhǔn
 own one:CLF big:eye **seem** COP form standard
měinǚ de bìyào yīnsù.]
 beautiful:woman ATTR necessary factor

Q ‘having a pair of big eyes **seems to** be a necessary feature of a standard beauty.’

- (4) P [Fùjìn cūnmín kàndào yǎngjī yǒulìkétú,
 Nearby villager see raise:chicken have:profit:may:expect
shì zhèngjīngbābǎi de zhìfù lù,]
 COP serious ATTR make:rich way

P ‘The villagers nearby saw that raising chickens was profitable, (and) a serious way to get rich,’

yúshì / *kějiàn / ?yīncǐ / ?suǒyǐ
thereupon/it can be seen/for this reason/so

- Q [∅ fēnfēn zhǎodào Wēn Běiyīng, yāoqiú zuò
 (the villagers) ADV find Wēn Běiyīng require become
de yǎngjīhù.]
 GEN raise:chicken:family

Q ‘(the villagers) came to find *Běiyīng Wēn* one after another, asking for joining the chicken farm as chicken farmers.’

Previous studies have provided valuable inspirations for the interpretation of the above relations. The typical function of *kějiàn* to express the inferential, or epistemic (to use Sweetser’s [1990] term), causal relations and its capability to signal a judgment or to draw a conclusion (Li et al. 2013; Lü 1999: 335; Xing 2001: 40–41) explain why it is the best fitting connective for the relation illustrated in (3). Historically, *kějiàn* has gone through a grammaticalization process from a phrasal verb (“can see”) to a discourse connective (“it can be seen that”) that specializes in expressing *personal attitudes, comments* and *judgments* (Li 2012). Such an expressive property makes it suitable to occur in causal relations

perceived from a heavily personal point of view, without necessarily resorting to explicit epistemic or perspective markers, or to some common ground between the speaker and the addressee. For example, *kějiàn* (and intuitively only *kějiàn*) fits in the reasoning trajectory from P to Q in this example: “P [Those women first make a requirement, that is, he should be a good-tempered man.] *Kějiàn* Q [women in that country are very miserable.] Their first requirement for a man is to be good-tempered.”

Yúshì is characteristic of introducing a dynamic and narrative Q that indicates an action or a change of state triggered by the situation in P; moreover, the cause-consequence relation holding between P and Q tends to be temporally related as well, that is, P happens prior to Q and leads to Q (Guo 2006; Lü 1999: 636; Xing 2001: 527; Zhao 2003). This distinctive profile of *yúshì* helps to explain the interpretation of the coherence relation holding in (4). The multiple possible readings when marked with *suǒyǐ* and *yīncǐ* correspond to the multiple functions of the two connectives in expressing various types of causal relations, be it expositive, narrative, or argumentative (or, in Sweetser’s [1990] terms, content or epistemic; see also Guo [2006]; Li et al. [2013]; Zhao [2003]).

The differences between *kějiàn*, *suǒyǐ*, *yīncǐ*, and *yúshì* seem to be aligned on the subjectivity-objectivity continuum, that is, to introduce the END state of a causal event as a judgment/conclusion, an (intended) action, or sometimes a fact in the world. Up to now, the study of Mandarin RESULT connectives has rarely been conducted systematically from the perspective of subjectivity. An exception is the study by Li et al. (2013), which, however, is restricted to the analysis of written discourse.

Written and spoken discourse differ substantially with respect to linguistic and textual features, given the producing and the receiving processes that differ in writing and speaking. Due to the reader-writer distance in both time and space dimensions, written language is usually more *explicit*, *detached*, and produced in more *integrated* and *well-planned* structures. In contrast, spoken language is less explicit but more *involving* and *interactional*, since both speaker and addressee have access to the immediate context *here and now* (see Chafe 1982, 1984; Clark 1996; and the references cited in; Xiao et al. 2021). Text-based communication between interlocutors on microblogging platforms in social media, such as Twitter, and Facebook, has been described as “*oralized written text*”, as if it is a hybrid of written text and oral speech (Voiskounsky 1997; Yus 2011: 174).

Such linguistic variations in discourse make explicit the speaker’s positioning with regard to the content *and* the speaker-addressee interaction in the communication event (De Smet and Verstraete 2006). It has been found that written language and spoken language differ in terms of subjectivity: the marking of subjectivity via the use of mental state predicates occurs most frequently in

conversational language but hardly in written language (Nuyts 2001: 396). An interesting question to ask is whether the difference in subjectivity will also be reflected in the linguistic expression of causal relations in different types of discourse.

A few cross-medium studies on the categorization of causal connectives have revealed interesting patterns. For example, French *parce que* ‘because’ is used significantly more often to express speech act and epistemic relations in telephone speech than in writing, while the use of *puisque* ‘because’ is constant: in both speech and writing, it is not found in the content domain, and is used more often in the epistemic than in the speech act domain (Zufferey 2012). Dutch *omdat* and *want* ‘because’ and Chinese *jìrán* ‘since’, *yīnwèi* and *yóuyú* ‘because’ differ systematically from each other in terms of subjectivity, and the distinctions remain largely stable across oral speech, written texts and internet-mediated texts (Sanders and Spooren 2015; Xiao et al. 2021).

It is as yet unknown whether or not the subjectivity characteristics of Mandarin RESULT connectives are sensitive to the particular context of different discourse types. To get insight into the issue, it is necessary to extend the empirical foundation of the analysis. Working on natural language corpus data of newspaper articles, spontaneous conversations, and microblog texts, we intend to answer the following questions:

RQ1: a: What is the prototypical meaning and use of the RESULT connectives *kějiàn*, *suǒyǐ*, *yīncǐ*, and *yúshì* in terms of subjectivity in distinctive discourse types?

b: Does the prototypical profile of each connective remain robust or does it vary across the distinctive discourse types?

RQ2: In what way do the different discourse types show variation with respect to subjectivity, as far as causal constructions are concerned?

RQ3: Are the various subjectivity features encoded in causal constructions equally important in characterizing the connectives under study?

We will first sketch the theoretical background of our approach to discourse coherence and subjectivity (Section 2), followed by the methodology section in which we introduce the data used in this study and our model of analysis (Section 3). In Section 4 we present the results, followed by a discussion of the results (Section 5) and the conclusion (Section 6).

2 An integrated approach to subjectivity

The meaning and use of connectives have been accounted for in terms of the categories of the relations they tend to express.³ Sweetser (1990) distinguished for example between content domain, epistemic domain and speech act domain. This approach to coherence is basically a cognitively oriented account that focuses on the conceptual meaning relations between two discourse segments (Sanders et al. 1992). When two events/situations are causally related to each other in the content world, the relation is classified as an instance of content causality, such as (4) in the previous section: that “the villagers” take the real-world action “came to find *Běiyīng Wēn* [...]” is because “they saw [...]”. Example (3) illustrates an instance of epistemic causality, which is construed from the speaker’s point of view: based on her argument “we always describe a beauty as having [...]”, she draws the conclusion “having [...] seems to be [...]”. A relation holds in the speech act domain when the speaker performs and motivates a speech act targeting an addressee,⁴ as exemplified in (5): the speaker suggests “you (i.e., *the addressee*) eat as usual” based on her argument “you only gain weight if [...]”.

- (5) A: *P* [*Wǒ gēn nǐ shuō, bù chīfàn nǐ zhǐ huì zēngféi.*]
 1SG PREP 2SG say NEG eat:meal 2SG only MOD increase:fat
P ‘I tell you what, you simply will gain weight if you do not have normal meal.’
- B: *èng.*
 INJ
 Uh.
- A: *Suǒyǐ* *Q* [*wǒ jiànyì nǐ gāi zhàocháng chīfàn.*]
So 1SG suggest 2SG should as:usual eat:meal
Suǒyǐ *Q* ‘I suggest you eat as usual.’

One thing worth special attention is that the three causal relations exemplified above all involve a *subject of consciousness (SoC)*, “an animate subject, a person,

³ The dichotomous accounts for relation categories are, among others, external/internal (Halliday and Hasan 1976; Martin 1992), subject matter/presentational (Mann and Thompson 1988), objective/subjective (Pander Maat and Degand 2001), semantic/pragmatic (Dijk 1979; Sanders 1997; Sanders et al. 1992), propositional/illocutionary (Sanders and Spooren 1999).

⁴ Speech act is meant here in the classical Austinian sense that an utterance can also be a social act, such as a declaration, a question, a promise, a piece of advice, namely, a performative utterance (Austin 1962; Huang 2006).

whose intentionality is conceptualized as the ultimate source of the causal event, be it an act of reasoning or some “real-world” activity” (Pander Maat and Sanders 2000: 64). By contrast, the interpretation of a causal relation like Example (6) does not require a responsible participant, i.e., there is no *SoC*: the fact that “having alcohol for the first time” leads to the consequence “I feel a bit dizzy”, which is not an intended act. Therefore, over and above Sweetser’s trichotomy, relations in the content domain are further refined into volitional relations, such as Example (5), and non-volitional relations, such as Example (6).

- (6) P [*Jintiān shì Ø píngshēng dìyìcì hējiǔ a*
 Today COP (I) this:life first:time drink:alcohol PRT
wèi, chún shēngpíjiǔ (hòuwèi zhēnxīn nánhē).]
 PRT, pure draft:beer (after:taste real:heart awful)
 P ‘Today, (I) have had alcohol for the first time in (my) life, the pure
 draft beer (honestly, the aftertaste is awful).’
Yúshì Q [*Ø xiànzài yǒudiǎn yūn.*]
As a result (I) now have:bit dizzy.
Yúshì Q ‘(I) feel a bit dizzy now.’

An integrated notion of subjectivity has been proposed to give a more detailed account of the distinctions between causal connectives, which takes into account both the domain-specific nature of the causal relations and whether or not an *SoC* is involved in the causal relation (Pander Maat and Degand 2001; Pander Maat and Sanders 2000, 2001). More specifically, the differences and similarities between connectives have been investigated in view of the distance between the speaker and the *SoC*, or the degree of the *speaker involvement* in the construction of the causal relation. Subjectivity is operationalized in terms of three relational aspects, each as a predictor of subjectivity encoded in the construction. The first aspect is the relational domain that a connective is typically used in, namely, the non-volitional, the volitional, the epistemic, or the speech act domain. The consequent, i.e., the *END* situation of the causal relation plays a vital role in determining the relation types (see Section 3.2.2 for more details). The propositional attitude expressed in the result segment can be an inevitable consequence or an act resulting from an external or uncontrollable force (non-volitional), an intentional activity (volitional), a personal opinion/conclusion (epistemic), or a performative utterance (speech act). The degree of subjectivity indicated in these relations increases in the following order: non-volitional, volitional, and epistemic/speech act relations.

The second aspect that indicates subjectivity is the identity of the *SoC*, which is closely related to the communicative *here and now* or the *Deictic Center of Communication* of the construction of the causal relation (Sanders et al. 2009; Traugott 1989, 1995). When the speaker is responsible for the causal interpretation, she is identical with the *SoC* and what are expressed are the inner thoughts or the intentional acts of the speaker herself. In other words, there is a high degree of speaker involvement. When a third-person character is the *SoC*, both he and the causal event are situated outside the communicative *here and now* of the speaker. In this case, the speaker is presenting the causal relation construed from the character's perspective. Hence, the speaker is distanced from the *SoC*, i.e., the speaker involvement is relatively decreased.

Speech act relations take place in the communicative *here and now*. The *SoC* is necessarily the speaker herself. In epistemic relations, the *SoC* is generally the speaker, but there are also contexts where the words are verbalized by the speaker, while the content is that of a character, for example: “That guy must be a lawyer, she thought”. In volitional content relations, the *SoC* is the agent of a volitional act, be it the speaker or a character. In the case of non-volitional relations, in which one event leading to another as a concrete fact, independent of anyone's intentionality, no *SoC* is involved. Consequently, non-volitional relations express the minimum degree of subjectivity.

Another subjectivity feature concerns the linguistic realization of the *SoC*. Following Langacker (1990: 7–9), the subjective or objective construal of an entity depends on whether it functions as the subject, or the object, of the conceptualization of the situation. Similarly, if an *SoC* is explicitly referred to, it is put “on-stage” and becomes part of the object of the conceptualization; if there is no explicit reference to an *SoC*, i.e., if the *SoC* is “off-stage”, it is the subject of the conceptualization described by the utterance. In other words, a linguistically explicit *SoC* is objectively construed, whereas an implicit *SoC* is subjectively construed.

The assumption of this integrated approach is that it allows for a fine-grained distinction of the communicative function of connectives. This has been supported by several analyses in different languages, among which Mandarin Chinese, using authentic corpus data as well as experimental methods (e.g., Degand and Pander Maat 2003; Pander Maat and Sanders 2001; Sanders and Spooren 2015; Spooren et al. 2010; Stukker et al. 2008; Xiao et al. 2021). In this study, we will employ this approach to the study of four Mandarin RESULT connectives *kějiàn*, *suǒyǐ*, *yīncǐ*, and *yúshì*.

3 Data and model of analysis

3.1 Corpus, data type and collection

Our data stem from three corpora: newspaper articles, spontaneous conversations, and microblog messages.⁵ These materials were selected because they are reasonably representative of written, spoken, and social media discourse in terms of the production and reception processes and potentially the variation of subjectivity.

The newspaper articles were collected from *Rénmín Rìbào* ‘People’s Daily’ (time span: 1991–1996). It is part of the Mandarin Chinese News Text corpus LDC95T13 (Wu 1995). Microblog data were taken from *Sina Weibo* ‘Sina Microblog’.⁶ Akin to the social networking applications of Facebook and Twitter, *Sina Weibo* is the most popular microblogging platform in China. Our data are taken from the posts in the year 2013, which is a sub-corpus of BCC (Xun et al. 2016).

Incorporated in the spoken corpus are conversations over telephone and in TV interview/talk-show programs. The telephone data consist of two parts. One is from LDC2005S15/T32 (Fung et al. 2005), which contains 897 calls between Mandarin speakers from Mainland China (about 150 h’ audio and transcripts of 2,030,730 Chinese characters). The other part is from LDC96S34/T16 (Canavan and Zipperlen 1996; Wheatley 1996), which contains 120 calls between family members from Mainland China (about 18.3 h audio and transcripts of 303,906 Chinese characters). TV interview/talk-show conversations are extracted from four programs: *Lǚyù yǒu yuē* ‘A Date with Luyu’, *Shíhuà shíshuō* ‘Tell It like It Is’, *Qiāngqiāng sānrénxíng* ‘Behind the Headlines’, and *Xīnwén huìkètīng* ‘People in the News’. These data stem from two corpus resources. One is LDC, from which we obtained in total 93 h’ audio and corresponding transcripts of 1,199,777 Chinese characters (Glenn et al. 2013, 2014, 2015, 2016; Walker et al. 2013, 2014, 2015, 2016). The other is MLC (2005), from which we made use of the transcripts for the programs *Lǚyù yǒu*

5 We used the following resources to create our datasets, i.e., LDC: Linguistic Data Consortium, hosted by the University of Pennsylvania. <https://www ldc.upenn.edu/>; BCC: Beijing Language and Culture University Corpus Center (BLCU BCC). <http://bcc.blcu.edu.cn/>; and MLC: Media Language Corpus, developed by the Monitoring and Research Center of National Broadcast Media Language Resources, Communication University of China. <http://ling.cuc.edu.cn/RawPub/>.

6 The microblog data under analysis are (predominantly) from the personal users. The corpus BCC makes no distinction between personal and official accounts. We read through the downloaded examples and left out those seemingly coming from the official accounts. It is a content-based screening method, in which posts containing brand/company/institution names, advertisements, professional subjects from journalism, financial, IT fields, etc. were discarded.

yuē (1,994,244 Chinese characters) and *Qiāngqiāng sānrénxíng* (5,903,852 Chinese characters).

From each corpus, 50 constructions marked with *kějiàn*, *suǒyǐ*, *yīncǐ* and *yúshì*, respectively, were randomly collected for further analysis (600 in total). The tokens *kějiàn* and *yúshì* do not function only as discourse connectives. *Kějiàn* can also be a verbal phrase ‘can see’ or an adjective phrase ‘visible’; *yúshì* very often occurs in phrases such as *děngyú shì* or *xiāngdāngyú shì* ‘be equal to’. For this reason, special attention was paid to make sure that only causal connective uses of *kějiàn* and *yúshì* were collected. Table 1 presents an overview of the complete datasets and the corpus information.

Table 1: Corpus size, discourse type and number of instances for analysis.

	<i>Kějiàn</i>	<i>Suǒyǐ</i>	<i>Yīncǐ</i>	<i>Yúshì</i>	Total
Newspaper (163,393,972)	50	50	50	50	200
Microblog (2,263,211,538)	50	50	50	50	200
Conversations: Telephone (2,334,636) & TV interview/ talk-show (9,097,873)	50	50	50	50	200
Total	150	150	150	150	600

3.2 Model of analysis

In line with the integrated notion of subjectivity introduced in Section 2, we analyzed each example in terms of three subjectivity features and followed a three-step procedure in annotation. The first is to annotate the domain of the causal relation, which is essentially related to the propositional attitude expressed in the consequent, i.e., the END status of the causal event (referred to as *PropAtt* hereafter). The second is to identify the subject responsible for the causal interpretation of the relation, i.e., the identity of the *SoC* (*IdSoC*). The third step is to annotate the linguistic realization of the *SoC* (hereafter *LingReal-SoC*). Each feature contains several categories that differ in their degree of subjectivity (Table 2). Subsequently, the analysis tests whether there are systematic co-occurring patterns between each connective and the subjectivity features, and whether or not the patterns remain robust across corpora.

Table 2: Model of integrated subjectivity.

Subjectivity feature	high.....	Subjectivity degree.....	low	
PropAtt (Domain)	Speech act/Judgment	Mental act	Physical act	Fact
Identity of <i>SoC</i>	Speaker	Character	No <i>SoC</i>	
LingReal- <i>SoC</i>	Implicit	Pro-drop	Explicit	Absent

The basis of this model was built up in previous studies on Dutch causal connectives (Sanders and Spooren 2015; Spooren et al. 2010) and Mandarin causal connectives (Li et al. 2013, 2016; Xiao et al. 2021). The major difference is that we have merged the domain of the relation and the propositional attitude of the result segment due to the correlation between these two features in the model (cf. Table 3).

Table 3: Paraphrase test for domain and the domain-PropAtt correlation.

Domain	PropAtt	Paraphrase test
Speech act	Speech act	The fact that P leads to the <i>speaker here and now</i> asking/suggesting/offering/commanding, etc. the addressee that Q.
Epistemic	Judgment	The fact that P leads to the <i>SoC's conclusion here and now</i> that Q.
Vol. content	Mental/Physical act	The fact that P leads to the <i>SoC's intentional</i> mental/physical act that Q.
Non-vol. content	Fact	The fact that P leads to the <i>fact</i> that Q. No intentionality is involved.

P: the cause segment, Q: the result segment.

3.2.1 Modification to the analytical model

Compared to the previous studies, the current model has left out the relation domain as a subjectivity predictor. The reason for this is that the value of *domain* can be predicted on the basis of the value of *PropAtt*.

Domain and *PropAtt* are conceptually different: the former stands for the category of coherence relation holding between P and Q; the latter concerns only the propositional content of Q. Nevertheless, determining a causal relation domain depends substantially on the property of the END status of the causal relation. This has been well supported in the literature. For example, the paraphrase test (Sanders 1997) is widely used to systematically analyze coherence relations. This test states that a relation is pragmatic (speech act and epistemic domain) if it has a Q-segment that expresses a(n) *advice/claim/conclusion*, whereas semantic relations (content domain) have Q-segments expressing *fact/action*. A similar correlation between the two features is demonstrated in the analysis of the construction “*p, parce que ‘because’ q*” by Groupe Lambda-I (1975) and Lambrecht et al. (2006) referred to in Zufferey (2012). These authors find that in the content domain, the END situation in *p* is something that “must be known by the hearer” (e.g., is part of his/her world knowledge), while it is “not known” in the epistemic and speech act domains (e.g., is a speaker’s conclusion) (Zufferey 2012: 140). They further conclude that *p* in the epistemic domain contains a *conclusion*, and a *speech*

act in the speech act domain (Zufferey 2012). Another reason for discarding *domain* as a predictor is that *PropAtt* plays an essential role in determining the property of the complex sentence in Chinese. Zhao (2003: 27–28) claims that the *jìngtài* ‘static’ or *dòngtài* ‘dynamic’ and the *yǐrán* ‘realis’ or *wèirán* ‘irrealis’ status of a complex sentence depends on the status of the Q-segment being *static* or *dynamic* and being *realis* or *irrealis*.

The *domain-PropAtt* correlation is also manifest in the study on Mandarin causal connectives in written discourse (Li et al. 2013: 93–95). Specifically, a Q-segment of *speech act* and *judgment* figures in the *speech act* and the *epistemic* domain, respectively, whereas the *content* domain (both *volitional* and *non-volitional*) correlates with *fact* (*mental* and *physical*). An approximation of the model is that *intentionality* was not strictly considered as a rule of distinguishing between the cases within either *mental fact* or *physical fact*. This obscures, to some extent, the predictive power of these two categories, since *intentionality* is the key element in a subjectivity account. However, it should be noted that whether or not there is involvement of *intentionality* in the Q-segment is considered in the paraphrase test for the volitional/non-volitional domain, which adds to our argument for leaving out *domain* or *PropAtt* when the correlation occurs.

In the latest version of the integrated subjectivity model (Xiao et al. 2021), *PropAtt* is classified on the basis of both the propositional content of the Q-segment and the involvement of an SoC in it. The direct outcome of this classification is a high correlation between *domain* and *PropAtt*. Given that *PropAtt* stood out as the most decisive variable in the analyses of random forests and conditional inference trees (see Section 4.6 in that study), it seems to make sense to keep *PropAtt* in the model and leave out *domain*.

In short, dropping *domain* as a separate variable makes the model more parsimonious and descriptively adequate. That is why the present study uses the updated three-variable model. Other modifications regarding the categorization of *PropAtt* and *LingReal-SoC* are presented in the respective sections.

3.2.2 Propositional attitude of the consequent

PropAtt is classified into five categories, which is based on the propositional content of the result segment (Q), and whether or not *intentionality* is involved in it. The degree of subjectivity increases successively from *fact*, *physical act*, *mental act*, to *judgment/speech act*.

Q is annotated as *speech act* if it is a performative utterance, namely, an illocutionary or perlocutionary act that is intended for the addressee and initiated

here and now by the speaker (cf. Footnote 4 and Table 3, above). For example, in example (5), repeated as (7) for convenience,⁷ that “you only gain weight if [...]” leads the speaker (I) to suggest you (the addressee) to “eat as usual”.

- (7) A: P [Wǒ gēn nǐ shuō, bù chīfàn nǐ zhǐ huì zēngféi.]
 P ‘I tell you what, you simply will gain weight if you don’t have normal meal.’
 B: En...
 Uh...
 A: **Suǒyǐ** Q [wǒ jiànyì nǐ gāi zhàocháng chīfàn.]
Suǒyǐ Q ‘I suggest you eat as usual.’

Q is annotated as *judgment* when it expresses feelings, opinions, attitudes, beliefs, evaluations and the like. In other words, the interpretation of what is expressed has to refer to the person who judges, concludes, evaluates, etc. An example is (3), repeated as (8), in which Q expresses the speaker’s opinion that “having a pair of big eyes seems to be a necessary feature of a standard beauty”.

- (8) P [Wǒmen xíngróng měinǚ zǒng shuō dà yǎnjīng, xiǎo zuǐbā.] **Kějiàn**, Q
 [yōngyǒu yīshuāng dà yǎn sìhū shì gòuchéng biāozhǔn měinǚ de biyào yīnsù.]
 P ‘We always describe a beauty as having big eyes and a small mouth.’
Kějiàn, Q ‘having a pair of big eyes seems to be a necessary feature of a standard beauty.’

Q is annotated as *mental act* if the predicate represents an *intentional* activity indicating a change of mental state or an opinion/decision coming into being, which takes place in the mental world of the agent. For example, in (9), “he” came to the decision of “deliver[ing] the money order to the county of Nanyue by himself”.

- (9) P [Jiàoshìjié zàijǐ, xìnhuìkuǎn zài yínháng zhìshǎo
 Teacher: festival coming, mail: post: money in bank SUPL: little
 zhōuzhuǎn yī xīngqī.]
 turn: over one week
 P ‘Teacher’s Day is around the corner, and mail transfer via bank would take at least one week.’

7 For ease of reading, we dropped the interlinear glossing in the repeated examples.

Yúshì,
Thereupon,

Q [tā juéding bǎ huìpiào qīnzì sòngdào nányuè xiàn.]
3SG decide PREP post:ticket self deliver Nanyue:PRO county
Yúshì, Q ‘he decided to deliver the money order to the county of Nanyue by himself.’

Q is annotated as *physical act* if the predicate represents an *intentional* activity of the agent taking place and observable in the physical world. An example is (4), repeated below as (10): Here Q expresses the action of the villagers of “com[ing] to find *Běiyīng Wēn* one after another, asking for joining the chicken farm as chicken farmers”.

- (10) P [Fùjìn cūnmín kàndào yǎng jī yǒulikétú, shì zhèngjīngbābǎi de zhìfù lù.] **Yúshì** Q [Ø fēnfēn zhǎodào Wēn Běiyīng, yāoqiú zuò jīchǎng de yǎngjīhù.]
P ‘The villagers nearby saw that raising chickens was profitable, a serious way to get rich.’ **Yúshì** Q ‘(the villagers) came to find *Běiyīng Wēn* one after another, asking for joining the chicken farm as chicken farmers.’

Q expresses a *fact* if it describes an event or situation which is caused or takes place independent of any speech participant’s intentionality. In other words, it expresses an inevitable and/or unintended consequence, be it in the physical or the mental world. In Example (11), “in the past, there was a joke in Liuzhou, [...]” illustrates a fact in the world, i.e., the existence of a joke in that area.

- (11) P [Suāntān yěshì Liǔzhōu jiētóu zuì pǔbiàn de
Sour:stall also:COP Liuzhou:PRO street:head SUPL common ATTR
xiǎochītān zhīyī,]
small:eat:stall of:one
P ‘Sour food stall is also one of the most popular street food stalls in Liuzhou,’

yīncǐ,
for this reason

Q [guòqù liǔzhōu yǒu jù wánxiàohuà shuō, xiǎng kàn Liǔzhōu
past Liuzhou:PRO have CLF joke:saying say want see Liuzhou:PRO
měiniǚ, qù suāntānshàng zhǎo, kěndìng shì zuìduō de.]
beautiful:woman go sour:stall:on search sure COP SUPL:many PRT

yīncǐ, Q ‘in the past, there was a joke in Liuzhou, “If you want to see Liuzhou beauties, go to the sour-food stalls, where you find surely the most”.’

Note that the *illocutionary/perlocutionary* value of a speech act can be expressed as an interrogative (question), an imperative, or a declarative sentence. If Q has the form of a declarative sentence, it is sometimes unclear whether Q expresses a speech act or describes an act or situation. In such cases we checked whether Q expresses a performative utterance in the *here and now* and targets an addressee. If this is the case, as in Example (7), we coded it as a speech act. However, the speech-act-like utterance in (12) “I then suggested [...]” does not take place nor direct at the addressee in the current communicative *her and now*. Hence, it is not a speech act, but a description of what the speaker said at that moment. A proper paraphrase for the *yīncǐ*-segment is “I *there and then* made the suggestion that, I was not being logical, ‘we may as well have a look at the photos of real women’.”

- (12) P [*Háiyǒu yīzhǒng qīngtóngqì, ràng wǒ juéde xiàng*
 Still:have one:CLF Bronze let 1SG feel like
měirén yīyàng, shìba, yǒu zhèyàng yīzhǒng shěnměi.]
 beauty same PRT have this one:CLF aesthetic
 P ‘There is also a kind of Bronze that makes me feel it is a beautiful woman, such aesthetic, you know.’

Yīncǐ

For this reason

- Q [*wǒ jiù tíyì, wǒ jiù xiā liánxì, wǒmen*
 1SG EMP suggest 1SG EMP blind relate 1PL
hái kěyǐ kànkàn zhēnde nǚxìng de zhàopiàn.]
 still may look:look real female ATTR photo

Yīncǐ Q ‘I then suggested that, I was not being logical, “we may as well have a look at the photos of real women”.’

In some cases, the Q-segment consists of more than one finite clause. Consistently throughout the coding process, we focused on the propositional content expressed in the *nucleus clause* (Mann and Thompson 1988). When the clauses are in an additive relation, each being an individual nucleus, we took the first clause as the result segment for analysis; when they are in a hierarchical relation, we took the main clause (nucleus) for analysis. For example, in (13), the two segments following *suǒyǐ* are two parallel clauses connected by the coordinating connective *bìng* ‘and’. We annotated S1, which describes a fact (“[being] capable of labor work”), rather than S2, which is a mental act “took it as an honor”. In (14), S1 “*The Ball* [username of the blogger “I”] was very angry” and S2 “(I) then switched off your phone” form a causal relation: the fact that “*The Ball* was very angry” leads to

the physical act “[to switch] off your phone”. We coded it as a physical act, based on the nucleus, i.e., the Q-segment of this embedded causal construction.

- (13) P [Zhèxiē huà dāngshí dōu duì tóngxué yǐ shēnkè
 These remark that:time all to student AUX deep
 de qǐfā,]
 ATTR enlighten
 P ‘All these words were very enlightening to the students at that time,’
suǒyǐ
so
 S1 Q [kàngdà bìyè de tóngxué, dōu huì
 Counter:university graduate GEN student all can
 láodòngshēngchǎn,]
 labor:produce
 S2 bìng yǐcǐwéiróng.
 and take:this:as:honor
suǒyǐ S1 Q ‘the students who graduated from the Counter-Japanese
 Military and Political University were all capable of labor work,’ S2
 ‘and (they) took it as an honor.’
- (14) P [Nǐ shǒujīshàng de wēibó yèmiàn kāizhe, dàn
 2SG phone:on GEN microblog webpage open:TAM but
 Ø jūrán méiyǒu guānzhù běn Qiú,]
 (you) even neg:have follow 1SG Ball
 P ‘The microblog app on your phone was logged in, but (you) were
 not even following me, *The
 Ball.*’
Yúshì
As the result
 S1 běn Qiú hěn shēngqì, S2 Q [Ø jiù bǎ nǐ shǒujī guānle.]
 1SG Ball very angry (I) EMP PREP 2SG phone switch:TAM
Yúshì S1 ‘I, *The Ball*, was very angry,’ S2 Q ‘(I) then switched off
 your phone.’

3.2.3 Identity of the SoC

Our second step of annotation was to identify the SoC: whose perspective or intentionality is involved in the construction of the causal relation. In our study, the speaker, the writer, and the blogger are all first person SoCs, for which we used the term *speaker SoC*. The third person SoC, whose perspective/intentionality is not directly voiced but uttered by the speaker, is termed as *character SoC* and

considered as adding less subjectivity to the relation than *speaker SoC*. When the *PropAtt* expresses an unintentional act or fact, it takes place or exists independent of an *SoC*. We labeled it as *no SoC*, which encodes the least degree of subjectivity.

For example, in (7), it is the speaker “I” who is responsible for the performative utterance “I suggest you eat as usual”. In (9), it is the character “he” who comes to the idea of “deliver[ing] the money order [...] by himself”. In (11), “there was a joke in Liuzhou, [...]” states the existence of an entity, which does not rely on any subject for interpretation.

3.2.4 Linguistic realization of the *SoC*

Once the *SoC* was identified, we looked at the way it is linguistically realized. In line with Langacker (1990), we considered the implicit reference to *SoC* as indicating a higher degree of subjectivity than the explicit case. As the intentionality of the *SoC* is best reflected in the consequent, we analyzed only the Q-segment for this feature.

In addition to the implicit-explicit dichotomy, we have set *pro-drop* as a third category. It is especially necessary in this study because Chinese is a typical *pro-drop* language (Huang 1989), and *pro-drop* occurs frequently in spontaneous and informal contexts such as conversational exchanges and microblogs. To the best of our knowledge so far, no study has related this grammatical form to subjectivity. As the dropped *SoC* in the Q-segment is present and recoverable in the context, we assume its degree of subjectivity to be in between the implicit and the explicit categories. Cases without *SoC* are annotated as *absent* and considered as encoding the minimal degree of subjectivity in our model.

The *explicit SoC* in our annotation also includes the explicit reference to an *SoC* in the accusative or possessive case. For example, the speaker in (15) is present in the Q-segment in the accusative case “me”. It was, therefore, annotated as an explicit *SoC*.

- (15) P [*Wǒ* *bèn*, *shuōhuà* *zhīqián* *xūyào* *sīkǎo*.]
 1SG stupid say:word PREP:before need think
 P ‘Stupid as I am, I need to think before I speak.’

suǒyǐ

so

- Q [*liáotiān* *xiě* *yóujiàn* *zuì* *shìhé* **wǒ**.]
 chat write email SUPL suit **1SG:OBJ**

suǒyǐ Q ‘online-chat and writing email suit **me** best.’

Both Q-segments in Section 3.2.2, “I suggest [...]” in Example (7) and “he decided to [...]” in (9), are explicit cases of *LingReal-SoC*. Example (8) illustrates an implicit case: the speaker draws the conclusion “[...] seems to be [...]”, yet she is linguistically absent in the utterance. In Example (10) “∅ came to find *Běiyīng Wēn*

[...]” illustrates a *pro-drop* category. In (11), no SoC is responsible for the causal link, hence it was annotated as *absent*.

3.3 Inter-coder agreement and objectivity bias

The codebook for the present study has been tested and proved effective in the study of Mandarin REASON connectives (Xiao et al. 2021). Of the 600 instances analyzed in the present study, 180 (30%) were annotated separately by the first two authors, who are native speakers of Mandarin Chinese. The inter-coder agreement (Cohen’s kappa) was found to be substantial: *PropAtt* $\kappa = 0.80$; *IdSoC* $\kappa = 0.87$; *LingReal-SoC* $\kappa = 0.83$. All the discrepancies were discussed afterwards. In the end, complete agreement was reached upon *IdSoC* and *LingReal-SoC*; seven cases regarding *PropAtt* remained in dispute.

To settle the disputes, we referred to the context for justification. Four of the disagreements were related to the interpretation of Q as an epistemic conclusion or as a mental act. The difference between the two categories can be very subtle. By nature, a mental act is distinctive from a conclusion in that the former implies a change of mental state or expresses an activity taking place within the mental world of the SoC, whereas the latter expresses an opinion or the state of mind itself. For the subtleties, see (16), an example from the spoken corpus, for which the preceding context is provided, and Example (17), from the microblog corpus.

- (16) *Dànshì bùguǎn tā zěnme nǔlì, nàxiē yuángōng dōu jiào tā gōngzǐ, huòzhě shì, jiùshì bǎ, bǎ tā dāngchéng yīgè fùyù, fùyù jiāting de gōngzǐ.*
 ‘However, no matter how hard he worked, those employees still addressed him ‘young master’, or simply regarded him as a, as a wealthy, as the son from a wealthy family.’

P [*Zhújiàn de, tā jiù juéde zìjǐ*
 Gradually PRT 3SG EMP feel self
de nǔlì méiyǒu shénme jiéguǒ,
 GEN effort NEG:have what result
háowúyìyì.]

ADV:NEG:meaning

P ‘Gradually, he just felt that his efforts did not result in anything, (and were) downright meaningless.’

Yúshì

Thereupon

Q [*tā jiù xiǎng dāng yīmíng jǐngchá.*
 he EMP want become one:CLF policeman

Yúshì Q ‘he **then wanted** to become a police officer.’

- (17) P [Wōmen lái zì huí bù qù de dì fāng,
 1PL come:from return:NEG:to ATTR place
 zǒu xiàng huí bù lái de dì fāng,]
 walk:to return:NEG:back ATTR place
 P ‘We come from a place where we can’t go back and head for a place
 from where we can’t return,’

suǒ yǐ

so

- Q [wǒ xiǎng zǒu màn diǎn.]
 1 SG want walk slow:PRT
suǒ yǐ Q ‘I **want** to go a bit slowly.’

The ambiguity is caused by *xiǎng* ‘want’ in the predicate, which, depending on the context, can indicate both a mental act and an opinion. In (16), “he *then wanted to become* a police officer” indicates the change of his mind (from “working hard to gain the employees’ recognition” to “considering to take a different job”). This interpretation is also implied by the condition presented in the context: all his efforts failed to change other people’s opinion of him as a son from a wealthy family, and he felt frustrated with it. Moreover, this is also a typical instance of a narrative causal relation expressed by *yúshì*: an event takes place under some circumstances of another. The conjunction adverb *jiù* ‘then’ also implies an act of reasoning. Therefore, we annotated it as *mental act*. However, “I want to go slowly” in (17) indicates the SoC’s opinion. A natural paraphrase is: “we come from [...] and head for [...]”, so my opinion is “to go slowly”. Hence, we annotated it as *judgment*.

The remaining three disputes were over an *intentional* or *unintentional* reading of the causal event. If the context did not help, or if there was lack of context (especially in microblogs), we resolved the dispute by opting for the unintentional reading. Example (18) is a microblog message in our corpus. Due to lack of context, it remains unclear whether the blogger’s notice of “the change of numbers” means that he/she intentionally pays close attention to the number (in which case Q expresses an *act*), or that the change of number inevitably catches his/her attention (in which case Q expresses a *fact*). In such cases, we opted for the objective interpretation: *fact*.

- (18) P [Wǒ guānzhù hé fēnsī běnlái jiù shǎo,]
 1SG follow and fan ADV EMP few
 P ‘I just have a limited number of following accounts and fans,’

yīncǐ

for this reason

- Q [∅ duì shùzì biànhuà fēicháng mǐngǎn.]
 (I) to number change very:much sensitive
yīncǐ Q ‘(I) am very sensitive to the change of numbers.’

Note that in the complete dataset of the 600 instances, only 10 cases in total remain ambiguous. We consistently took the objectivity bias (i.e., chose an objective reading) for the coding of these cases.

4 Results

We first carried out general log-linear analyses to look into the interactions between *connective*, *subjectivity*, and *corpus*. The analyses enabled us to observe the co-occurrence pattern of each connective and subjectivity feature and whether or not it is robust across corpora (RQ1 and RQ2). Results regarding *PropAtt*, *IdSoC* and *LingReal-SoC* are presented in Section 4.1 to Section 4.3. We used an alpha level of 0.05 and chi-square test for these statistical tests; Fisher's exact test was used (see Section 4.3) when there is any cell with expected frequencies less than five (Field 2013: 335). In the cross-table analyses, we have used the Bonferroni correction test, which returns the adjusted residual values and an adjusted critical value of alpha for each individual test. Section 4.4 presents the results of the conditional tree analysis and random forest analysis that illustrate the predicative power of the three subjectivity features (RQ3).

4.1 Propositional attitude of the consequent

The frequencies and percentages of the five *PropAtt* categories observed in each corpus and in the relations signaled with each connective are presented in Figures 1 and 2 below, respectively.

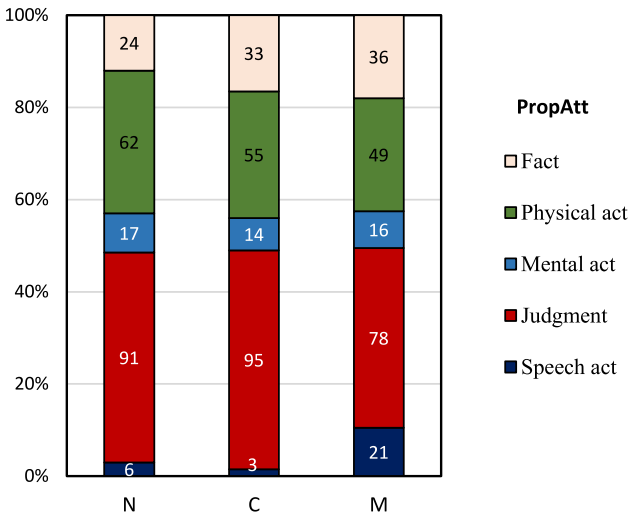


Figure 1: Frequencies and percentages of each *PropAtt* category observed in the corpus of Newspaper (N), Conversation (C) and Microblog (M).

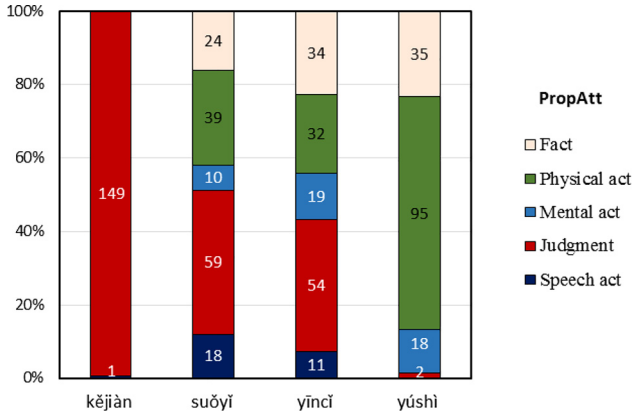


Figure 2: Frequencies and percentages of each *PropAtt* category co-occurring with *kējìàn*, *suǒyǐ*, *yīncǐ* and *yúshì*.

The log-linear analysis resulted in a model containing two significant two-way interactions. The fit of the model was substantial, $\chi^2(30, N = 600) = 13.65, p = 0.995$. The first interaction was between *Corpus* and *PropAtt*, $\chi^2(8, N = 600) = 24.07, p = 0.002$. The second was between *Connective* and *PropAtt*, $\chi^2(12, N = 600) = 451.95, p < 0.001$. The interactions were followed up by cross-table analyses summarized in Tables 4 and 5.

Table 4: Distribution of *PropAtt* categories in each corpus (frequencies and adjusted residuals).

		Speech act	Judgment	Mental act	Physical act	Fact	Total
Newspaper	Count	6	91	17	62	24	200
	Adj. Res.	-1.6	0.5	0.4	1.3	-1.7	
Conversation	Count	3	95	14	55	33	200
	Adj. Res.	-2.8	1.2	-0.5	-0.1	0.5	
Microblog	Count	21	78	16	49	36	200
	Adj. Res.	4.4	-1.7	0.1	-1.2	1.2	
Total	Count	30	264	47	166	93	600

The adjusted critical value of a.r. = 2.94.

The adjusted residuals in Table 4 show that the significant interaction between *Corpus* and *PropAtt* was exclusively due to the corpus-sensitive distribution of *Speech act*. A significantly high frequency was observed in *Microblog* (a.r. = 4.4). Example (19) illustrates a typical speech act instance from the microblog data.

- (19) P [Hūnyīn hé àiqíng bùtóng, tā yào
 Marriage and love NEG:same 3SG require
de shì chángcháng jiǔjiǔ de shēnghuó,
 PRT COP long:long ever:ever ATTR life

P ‘Unlike love, marriage aims for a long-term life,’

suǒyǐ

so

- Q [yīdìng yào nàizhùxìngzi, zhǎo yīgè shìhé
 sure MOD bear:temper find one:CLF suitable
zìjǐ de rén lái lái lái yìqǐ shēnghuó.
 self ATTR person INF together live

suǒyǐ Q ‘be sure to be patient and find a suitable person to spend the life together.’

Table 5: Relation between *PropAtt* and connective (frequencies and adjusted residuals).

		Speech act	Judgment	Mental act	Physical act	Fact	Total
<i>Kějiàn</i>	Count	1	149	0	0	0	150
	Adj. Res.	-2.8	15.8	-4.1	-8.7	-6.1	
<i>Suǒyǐ</i>	Count	18	59	10	39	24	150
	Adj. Res.	4.5	-1.3	-0.6	-0.5	0.2	
<i>Yīncǐ</i>	Count	11	54	19	32	34	150
	Adj. Res.	1.5	-2.3	2.5	-2.0	2.8	
<i>Yúshì</i>	Count	0	2	18	95	35	150
	Adj. Res.	-3.2	-12.2	2.2	11.3	3.1	
Total	Count	30	264	47	166	93	600
	Adj. Res.						

The adjusted critical value of a.r. = 3.02.

The adjusted residuals in Table 5 show that the co-occurrence pattern of *Connective* and *PropAtt* varies a lot from one connective to another. *Kějiàn* expresses predominantly *Judgments* (149 out of 150, a.r. = 15.8). *Yúshì* behaves just the opposite. It barely expresses any subjective category of *Speech act* (a.r. = -3.2) or *Judgment* (a.r. = -12.2), but mainly the objective ones: *Physical act* (a.r. = 11.3) and *Facts* (a.r. = 3.1). *Suǒyǐ* and *yīncǐ* appear less extreme. Generally, *suǒyǐ* is more on the subjective side, as witnessed by the high adjusted residuals for *Speech act* (a.r. = 4.5).

In Example (19) above, “be sure to be patient [...]” illustrates the typical speech act relation introduced by *suǒyǐ*. The Q-segment “there are often mice in this house” in Example (20) is a claim (*Judgment*) by the SoC, which is almost the only type of *PropAtt* found in relations signaled by *kějiàn*. *Yīncǐ* tends to express factual Q-segments. *Physical act* occurs most frequently in *yúshì* relations, as exemplified

by (10), repeated as (21): “came to find *Běiyīng Wēn* [...]”. Example (22) illustrates a factual Q-segment in *yúshì* cases: the appearance of “profiteers” is presented as a natural consequence due to the “rise in prices” on the tulip market.

- (20) P [*Qíángjiǎo, zhuōzi xià fàngzhe dú’ěr,*
 Wall:corner table under lay:PRT poison:bait
 P ‘There are poisonous baits in the corners and under the table,’

kějiàn,

it can be seen

- Q [*zhè wū cháng yǒu hàozi chūmò.*
 this room often have mouse appear
kějiàn, Q ‘there are often mice in this house.’

- (21) P [*Fùjìn cūnmín kàn dào yǎng jī yǒulì kě tú, shì zhèngjīngbābǎi de zhìfù lù.*
Yúshì Q [*Ø fēnfēn zhǎodào Wēn Běiyīng, yāoqiú zuò jī chǎng de yǎng jī hù.*]
 P ‘The villagers nearby saw that raising chickens was profitable, (which is) a serious way to get rich.’ **Yúshì** Q ‘(the villagers) came to find *Běiyīng Wēn* one after another, asking for joining the chicken farm as chicken farmers.’

- (22) P [*Suízhe jiàgé de shàngzhǎng, yǔ yùjīnxiāng*
 Follow:PRT price ATTR up:rise with tulip
zāipéi wú zhíjiē guānxì de rén
 plant NEG direct relation ATTR person
yě cānyùle jiāoyì,
 also join:TAM transaction
 P ‘With the rise of the price, people who have no direct relationship with the cultivation of tulips also participated in the transaction,’

yúshì

as a result

- Q [*chūxiànle yīpī “dǎoyé”, xǔduō rén yīyè*
 appear:TAM one:CLF profiteers many people one:night
zhījiān chéngwéi fùwēng.
 over become rich:man
yúshì Q ‘there has been a surge of “profiteers”, many of whom became wealthy overnight.’

4.2 Identity of the SoC

The frequencies and percentages of the three types of SoC identity observed in each corpus and in the relations marked with each connective are presented respectively in Figures 3 and 4 below.

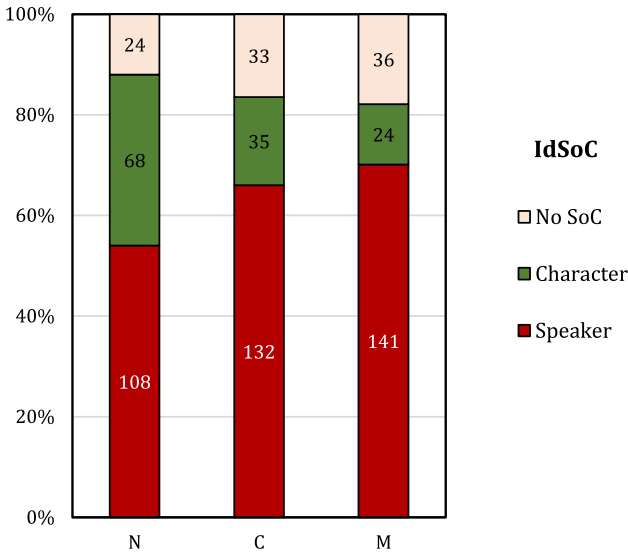


Figure 3: Frequencies and percentages of each *IdSoC* category observed in Newspaper (N), Conversation (C) and Microblog (M).

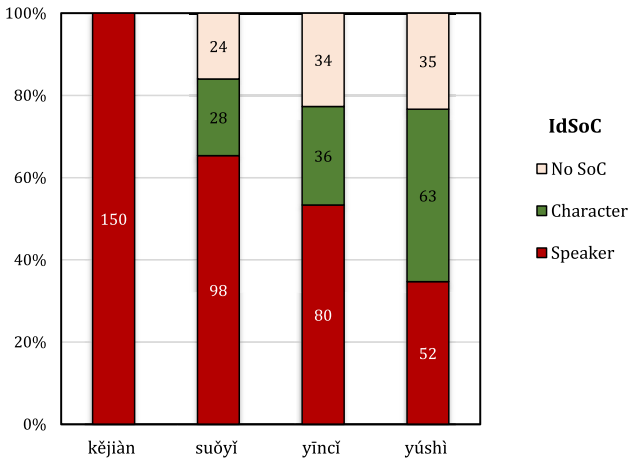


Figure 4: Frequencies and percentages of each *IdSoC* category co-occurring with *kějìàn*, *suǒyǐ*, *yīncǐ* and *yúshì*.

The log-linear analysis based on the identity of the *SoC* resulted again in a model containing two significant two-way interactions. One was between *Corpus* and *IdSoC*, $\chi^2(4, N = 600) = 30.98, p < 0.001$. The other was between *Connective* and *IdSoC*, $\chi^2(6, N = 600) = 197.30, p < 0.001$. The fit of the model was adequate, $\chi^2(18, N = 600) = 14.32, p = 0.708$. The following crosstab analyses provide us an insight into the two interactions.

Table 6: Distribution of *IdSoC* types in each corpus (frequencies and adjusted residuals).

		Speaker	Character	No <i>SoC</i>	Total
Newspaper	Count	108	68	24	200
	Adj. Res.	-3.4	5.4	-1.7	
Conversation	Count	132	35	33	200
	Adj. Res.	1.0	-1.6	0.5	
Microblog	Count	140	24	36	200
	Adj. Res.	2.4	-3.9	1.2	
Total	Count	380	127	93	600

The adjusted critical value of a.r. = 2.77.

The adjusted residuals in Table 6 show that the interaction between *Corpus* and *IdSoC* was largely due to the corpus-sensitive distribution of *Speaker SoC* and *Character SoC*. The former was relatively rare in *Newspaper* (a.r. = -3.4). The latter occurred relatively often in *Newspaper* (a.r. = 5.4), but rarely in *Microblog* (a.r. = -3.9). Example (21) from our written data exemplifies a *character SoC*, i.e. “the villagers”, undertaking the physical act “came to find *Běiyīng Wēn* [...]”.

Table 7: Relation between *IdSoC* and connective (frequencies and adjusted residuals).

		Speaker	Character	No <i>SoC</i>	Total
<i>Kějiàn</i>	Count	150	0	0	150
	Adj. Res.	10.8	-7.3	-6.1	
<i>Suǒyǐ</i>	Count	98	28	24	150
	Adj. Res.	0.6	-0.9	0.2	
<i>Yīncǐ</i>	Count	80	36	34	150
	Adj. Res.	-2.9	1.0	2.8	
<i>Yúshì</i>	Count	52	63	35	150
	Adj. Res.	-8.4	7.2	3.1	
Total	Count	380	127	93	600

The adjusted critical value of a.r. = 2.87.

The adjusted residuals in Table 7 cast light on the interaction between *Connective* and *IdSoC*. All instances with *kějiàn* have a speaker *SoC* ($n = 150$, a.r. = 10.8). By contrast, *yúshì* co-occurred relatively rarely with *Speaker SoC* (a.r. = -8.4) and mainly with *Character SoC* (a.r. = 7.2) and *No SoC* (a.r. = 3.1). As witnessed by the adjusted residuals, *suǒyǐ* relations appeared neutral, although the absolute frequencies of *Speaker SoC* were rather high (98 out of 150). In *yīncǐ* relations, the most subjective category *Speaker SoC* occurred less than expected (a.r. = -2.9), and the objective category *No SoC* almost reached significance (a.r. = 2.8 vs. 2.87).

Example (20) in Section 4.1 illustrates a typical *kějiàn* relation conceptualized by a speaker *SoC*: the speaker draws the conclusion that “there are often mice in this house”. Typical *yúshì* relations are exemplified via (21) with a character *SoC*, “the villagers” who “came to find *Běiyīng Wēn* [...]”, and (22) with a factual Q-segment describing a real-world event, i.e., “a surge of profiteers on the market of tulips”. A *yīncǐ* instance with no *SoC* is Example (11), repeated below as (23), which introduces the fact of an old joke exiting in the local area.

(23) P [Suāntān yěshì liǔzhōu jiētóu zuì pǔbiàn de xiǎochītān zhī yī,] **yīncǐ**, Q
[guòqù liǔzhōu yǒu jù wánxiàohuà shuō, xiǎng kàn liǔzhōu měinǚ, qù
suāntān shàng zhǎo, kěndìng shì zuìduō de.]

P ‘Sour food stalls used to be one of the most popular street food stalls in Liuzhou,’ **yīncǐ**, Q ‘in the past, there was a joke in Liuzhou, “If you want to see Liuzhou beauties, go to the sour food stalls. There you find surely the most”.’

4.3 Linguistic realization of *SoC*

Presented in Figure 5 are the frequencies and percentages of the four *LingReal-SoC* categories co-occurring with every connective in, respectively, the corpus of newspaper, conversation and microblog.

The log-linear analysis generated a model containing a significant three-way interaction: *Connective* LingReal-SoC* Corpus*, $\chi^2(18, N = 600) = 36.51, p = 0.006$, indicating that the relation between *Connective* LingReal-SoC* was moderated by *Corpus*. Table 8 below reveals the behaviors of the connectives, in *Newspaper* ($p < 0.001$, Fisher’s exact test), *Conversation* ($p < 0.001$, Fisher’s exact test), and *Microblog* ($\chi^2(9, N = 200) = 70.82, p < 0.001$), respectively.⁸

Across all the three corpora, *kějiàn* co-occurred predominantly with *Implicit*, rarely with *Pro-drop* and *Absent*; the pattern with *Explicit* varied from being rare in

⁸ We used Fisher’s exact test for the cross-table analyses in the newspaper and conversation corpora due to the observation that 4 cells had expected counts less than 5, which does not meet the chi-square assumptions. For the corpus of microblog, all assumptions were met; hence, chi-square was used.

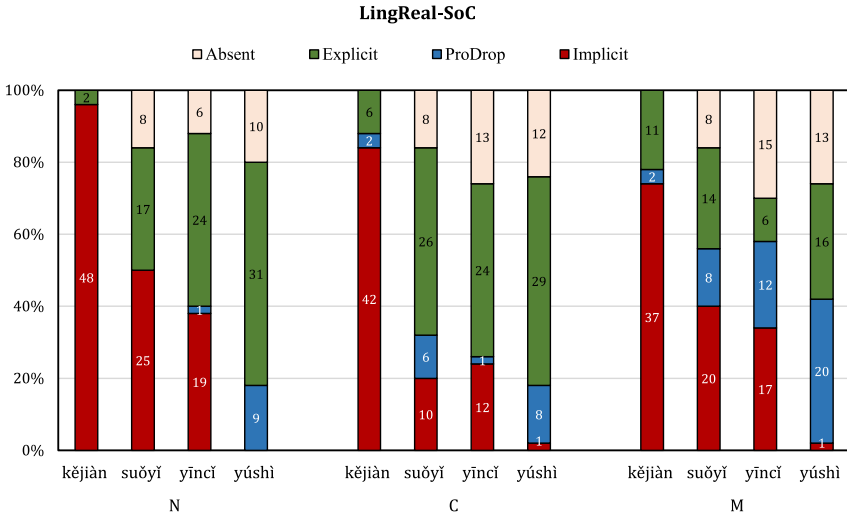


Figure 5: Frequencies and percentages of each *LingReal-SoC* category co-occurring with *kějiàn*, *suǒyǐ*, *yīncǐ* and *yúshì* in Newspaper (N), Conversation (C) and Microblog (M).

Table 8: Relation between *LingReal-SoC* and connective in each corpus (frequencies and adjusted residuals).

		LingReal-SoC							
		Implicit		Pro-drop		Explicit		Absent	
		Count	a.r.	Count	a.r.	Count	a.r.	Count	a.r.
Newspaper	Kějiàn	48	8.2	0	-1.9	2	-5.6	0	-3.0
	Suǒyǐ	25	0.7	0	-1.9	17	-0.5	8	1.0
	Yīncǐ	19	-1.3	1	-1.1	24	1.9	6	0.0
	Yúshì	0	-7.5	9	4.9	31	4.2	10	2.0
Total		92		10		74		24	200
Conversation	Kějiàn	42	9.0	2	-1.3	6	-5.0	0	-3.6
	Suǒyǐ	10	-2.2	6	1.0	26	1.6	8	-0.1
	Yīncǐ	12	-1.5	1	-1.9	24	0.9	13	2.1
	Yúshì	1	-5.3	8	2.2	29	2.6	12	1.6
Total		65		17		85		33	200
Microblog	Kějiàn	37	6.2	2	-3.4	11	-0.3	0	-3.8
	Suǒyǐ	20	0.4	8	-1.0	14	0.9	8	-0.4
	Yīncǐ	17	-0.6	12	0.6	6	-2.2	15	2.6
	Yúshì	1	-6.0	20	3.8	16	1.6	13	1.7
Total		75		42		47		36	200

The critical value of a.r. = 2.96.

Newspaper (count: 2; a.r. = -5.6) and *Conversation* (count: 6; a.r. = -5.0) to not so rare in *Microblog* (count: 11, a.r. = -0.3). In contrast to *kějiàn*, in all three corpora, *yúshì* rarely occurred with *Implicit*. The pattern regarding *Explicit* and *Pro-drop* varied across corpora: relatively many *Explicit* in *Newspaper* (count: 31; a.r. = 4.2), and relatively many *Pro-drop* in both *Newspaper* and *Microblog* (respectively, a.r. = 4.9 and 3.8). *Suōyǐ* and *yīncǐ* appeared relatively neutral in all corpora.

Exemplified in (24) is *kějiàn* co-occurring with an explicit SoC in microblog; Examples (25) and (26) demonstrate the use of *yúshì* with an explicit SoC in newspaper and *pro-drop* in microblog.

- (24) Preceding context: *Tóngyì de zhuǎn*.
‘Forward (the post if you) agree.’

P [Huíguó dàjiā dōu shuō wǒ shòule.]
Return:country everyone all say 1SG thin:TAM
P ‘Back to China, everyone said that I had lost weight.’

Kějiàn

It can be seen that

Q [wǒ zhīqián zhēnde yǒudiǎn pàng?]
1SG before real:PRT have:little fat?
Kějiàn Q ‘I was really a bit fat before?’

- (25) P [Hái zài běijīng shàngdàxué shí, xǔduō
Still in Beijing study:university time many
rén gàosù wǒ xīzàng zuì
people tell 1OBJ Tibet SUPL
quē jiàoshī.]
lack teacher
P ‘While still studying in university in Beijing, many people told me that Tibet was the worst with a shortage of teachers.’

Yúshì

Thereupon

Q [tíqián bànnián wǒ jiùzài bìyè
in:advance half:year 1SG EMION graduate
fēnpèi shēnqǐngshū shàng xiěxià “yuàn
assignment application:form on write:down “willing
dào xīzàng cóngshì jiàoyù gōngzuò”.]
go Tibet take education work”

Yúshì Q ‘I wrote “willing to go to Tibet to work in education field” on the application form of graduation assignment half year in advance.’

- (26) P [*Gāosān de shíhòu, bānzhūrèn rènwéi wǒ*
 High:three ATTR time class:director think 1SG
yǐjīng yǒu dúlì shēngcún de nénglì,
 already have independent survive ATTR ability,
 P ‘In the third year of high school, the teacher (charging the class)
 thought that I already had the ability to survive independently,’
- yúshì**
thereupon
- Q [*∅ ràng wǒ tuì xué.*]
 (the teacher) let 1SG quit:TAM school
yúshì Q ‘(the teacher) let me quit school.’

4.4 Relative importance of the three subjectivity features

To measure the strength of each of the three subjectivity features in distinguishing between the connectives under study, we performed the conditional inference trees and random forest analyses using the packages *party* and *Rling* in R (Levshina 2015: 291–300; R Core Team 2017). These are two non-parametric tree-structure models of regression and classification that are appropriate for our dataset, as it is relatively small in size and has multiple categorical predictors. The conditional inference trees analysis resulted in several splits (graphically represented as a tree) reflecting important predictors in the model (only significant splits are presented). The analysis also provides *p* values representing the confidence level of each split.

The conditional inference tree below (Figure 6) presents the decision rules for distinguishing *kějiàn*, *suǒyǐ*, *yīncǐ*, and *yúshì*.⁹ The first split involves *PropAtt*, which set apart *Judgment* from the rest. A further split within *Judgment* brought *LingReal-SoC* into play (Node 7). This node includes 52 cases of *Explicit Judgment* (Node 8) which contain mainly *suǒyǐ* (40%) and *kějiàn* (35%) and a moderate amount of *yīncǐ* (20%). The 212 cases of *Implicit/Pro-drop Judgment* (Node 9) contain predominantly *kějiàn* (60%); the rest are *suǒyǐ* and *yīncǐ* (each circa 20%).

⁹ To present only the relatively stronger predictors that make each split, we set 20 as the minimum number for each bin. This leads to the absence of the less powerful predictor *IdSoC* from the “tree”. The abbreviations in the tree graph: SpAct = speech act; MnAct = mental act; PhAct = physical act; KJ = *kějiàn*; SY = *suǒyǐ*; YC = *yīncǐ*; YS = *yúshì*.

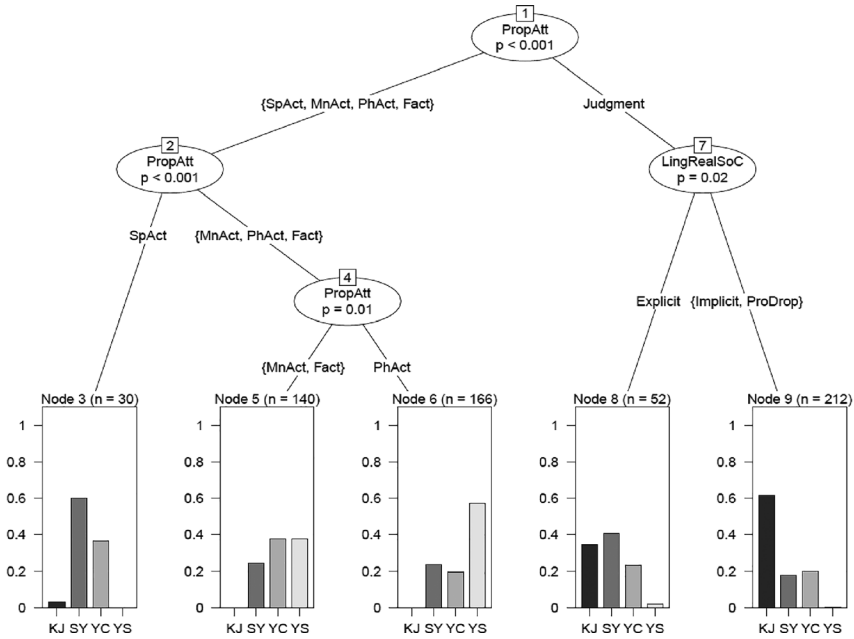


Figure 6: Conditional Inference Tree predicting the occurrence of four RESULT connectives on the basis of *PropAtt*, *LingReal-SoC*, *IdSoC*.

The second split (Node 2) again related to *PropAtt* and set apart *Speech act* from *Mental/Physical act* and *Fact*, resulting in a bin containing mainly *suōyǐ* and *yīncǐ* (respectively, 60 and 40% of the 30 instances, Node 4). The third split of *PropAtt* (Node 4) set apart *Physical act* from *Mental act/Fact*. The 140 cases of *Mental act/Fact* (Node 5) contain equal amounts of *yīncǐ* and *yúshì* (both nearly 40%) and circa 25% of *suōyǐ*. The bin with 166 instances of *Physical act* (Node 6) contains predominantly *yúshì* (60%) and roughly an equal amount of *suōyǐ* and *yīncǐ* (both circa 20%). The predictive accuracy of this model was moderate ($C = 0.53$). In our case, there are four choices to be made (the four connectives). A random model would predict that each connective occurs in 25% of the observations. It can be concluded that 0.53 is a considerable improvement over 0.25.

The analysis shows that *PropAtt* played a role in three of the four significant splits in the data, suggesting that it was the most important predictor in the model. This interpretation was confirmed in Figure 7, which was the result of the random forest analysis. It gave information about the importance status of the features in distinguishing between the connectives under study. *PropAtt* stood out as the most important feature in the model. *LingReal-SoC* ranked second, followed by *IdSoC*.

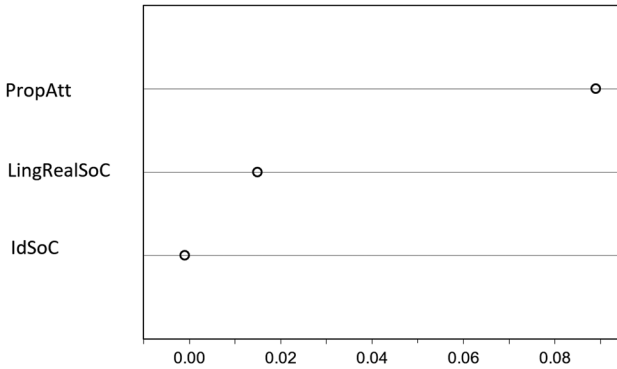


Figure 7: Evaluation of conditional importance of the variables *PropAtt*, *IdSoC* and *LingReal-SoC* in distinguishing between *kějiàn*, *suǒyǐ*, *yīncǐ* and *yúshì*.

5 Discussion

5.1 Summary of the results

This study was carried out to answer three research questions. The first asks whether the connectives differ systematically with respect to their prototypical behaviors in term of subjectivity, the notion of which has been operationalized in terms of three features “the propositional attitude of the result segment”, “the identity of the *SoC*” and “the linguistic realization of the *SoC*”, and whether the prototypical profile of each connective remains robust across three different discourse types: naturally occurring conversations, microblog messages and newspaper texts. The second asks whether (and to what extent) the three discourse types differ with respect to these subjectivity features, as far as the causal construction is concerned. The third research question asks which of the three features is the most important in determining the profile of a particular connective.

To answer the first two questions, a systematic log-linear analysis was carried out for the relationship between *connective*, *subjectivity* and *corpus*. We have found two significant interactions *connective*PropAtt* and *connective*IdSoC*, which were not moderated by the factor *corpus*. The results strongly support the conclusion that the connectives have distinctive and robust subjectivity profiles regarding these two features.

There was only one set of findings that ran counter to this claim of robustness: the interaction *connective*LingReal-SoC* was found to be moderated by *corpus* (Section 4.3). This moderation was mostly related to the connectives *kějiàn* and *yúshì* (for details, see Table 8). *Kějiàn* instances had relatively rare *explicit SoCs* in

newspaper and conversation data, while relatively many explicit SoCs in microblogs. This may well be a direct reflection of the property of the connective *kějiàn*, and the register of microblog discourse as well. It is part of the subjectivity profile of *kějiàn* to express opinions from the first-person perspective (as witnessed by the findings regarding *PropAtt* and *IdSoC* reported in the sections above). This profile fits well with the content of micro-blog posts, which is usually about the bloggers' own experiences, personal feelings or opinions. They themselves are often the participants of the speech event. This would naturally lead to relatively many cases of explicit reference to the SoC. By contrast, in writing or conversational communication, it can be expected that the speaker SoC is often drawing conclusions or exchanging opinions about the speech event concerning a third-person participant. As a consequence, there might be less cases of explicit reference to *the* SoC in the utterance.

For *yúshì*, the occurrences of *pro-drop* were more frequent than expected in newspaper and (especially) microblog data. A possible explanation is the various degrees of “reviewability” (Clark and Brennan 1991: 141) of the discourse in each corpus. Both microblogs and newspapers consist of written texts. When the reference to the SoC is dropped out in the utterance, it is still retrievable from the context and therefore can be easily recovered by the addressee. In oral communication, however, messages are exchanged in a dynamic continuum. Frequent use of *pro-drop* might lead to an ambiguous delivery of the message or hamper the communication flow. Another corpus sensitive finding concerned the case of *explicit SoC*, which occurred more than expected only in newspaper discourse. This we think reveals the *explicit* character of written language (Finegan 1987), especially with regard to formal writing like newspaper articles.

Only with respect to the linguistic realization of the SoC, *corpus* turned out to be a factor affecting the behaviors of the connectives, and that relevance was limited. Overall, the subjectivity profiles of the connectives remain stable across corpora. To be precise, *kějiàn* displayed a highly subjective profile as it co-occurred predominantly with *implicit SoC* in each corpus. *Yúshì*, by contrast, is the most objective connective, featuring rarely any *implicit SoC* and mostly *explicit* or *absent*. *Suǒyǐ* and *yīncǐ* had a generic character with respect to this feature.

In sum, the study provides clear answers to RQ1. *Kějiàn*, *suǒyǐ*, *yīncǐ*, and *yúshì* as causal connectives are distinct from each other in terms of subjectivity and their specific subjectivity profile is robust across corpora. *Kějiàn* proves to be the most subjective connective, expressing predominantly *judgments* drawn by an *implicit speaker SoC*. *Yúshì* is the most objective connective, introducing prototypically *factual* consequents that are independent of an SoC, or intentional *physical acts*. The SoC that undertakes the *physical act* can be both the *speaker* and (significantly more often) the *character* who are mostly *explicitly* referred to (and are dropped in microblogs). *Suǒyǐ* and *yīncǐ* are relatively neutral, both co-occurring with all

PropAtt categories and a majority of *speaker SoCs*, as well as a moderate amount of *character SoCs* and *no SoC*. Nevertheless, *suǒyǐ* appears to be slightly more subjective in that it expresses significantly more *speech acts*; *yīncǐ*, in comparison, has less cases of *speaker SoC* and more cases of *no SoC*. This leads to the following ordering of the connectives on a subjective-objective continuum, the order would be *kějiàn*>*suǒyǐ* >*yīncǐ*>*yúshì*.

The answer to RQ2 is that the variation between different discourse types with respect to subjectivity is also reflected at the level of the causal construction, as witnessed by the two significant interactions of *corpus*PropAtt* and *corpus*IdSoC*. The differences are relatively small. Microblog has relatively many *speech acts*; newspaper has relatively many *character SoCs* (which are rare in microblog). It seems that microblog discourse is the most subjective type, that newspaper discourse the most objective and that oral speech is in between.

The answer to RQ3 follows the findings from the conditional inference tree analysis and the random forest analysis (Section 4.4). The type of propositional attitudes expressed in the result segments proved to be the most important feature in characterizing the meaning and use of the connectives and making distinctions between them. It is involved in three of the four significant splits of our data. The linguistic realization of the *SoC* and the identity of the *SoC* follow in a decreasing order of importance.

5.2 Implications of the findings

The prototypical subjectivity profiles of the four RESULT connectives under analysis are to a large degree compatible with the conclusions drawn by previous studies (for details see Section 1). However, the robustness of the profiles of *suǒyǐ* and *yīncǐ* does not match the findings of a study using written data (Li et al. 2013). In that study, both connectives co-occur with more *physical facts* (counterpart of *physical acts* and *facts* in our model), *explicit* and *character SoC* in novels than in news reports and newspaper opinion pieces. As the present study did not include data from novels, there is no possibility for us to make direct comparisons. It could be argued that, compared to news reports and opinion pieces, the narratives in novels are more detached from the communicative *here and now*, and are more likely to be devoted to the description of the characters' activities.

Another point worth mentioning is our finding that *yīncǐ* bears an objective flavor. It occurs infrequently with *speaker SoC* and tends to favor the most objective category *no SoC* (Table 7), which is indicative of the objective non-volitional relation. Nevertheless, it should be noted that the significance is relatively small

based on our corpus data. It requires more empirical studies (for example, using experimental methods) for a clearer picture of this objective tendency.

By and large, our findings derived from the modified model are consistent across our dataset in each corpus. This consistency confirms the validity of the integrated subjectivity approach to the semantic-pragmatic properties that are prototypical of each connective. Moreover, through the systematic analysis of data from theoretically different discourse varieties, we have gained more insight into the use of Mandarin causal connectives, especially in the speech act domain, which has been substantially restricted in previous studies using written data. In the present study, a case in point is *suǒyǐ*, which is found to be preferred in the speech act domain (Table 5). This finding suggests that *suǒyǐ*, which prototypically has a relatively neutral profile, has a subjective flavor. In this sense it differs from *yīncǐ*, the other relatively neutral RESULT connective. The potential of a finer-grained distinction between these two connectives is promising.

Furthermore, the adapted model sets *pro-drop* apart as a separate category, as suggested in Xiao et al. (2021). This asks for further study concerning the subjectivity characteristics associated with this specific linguistic phenomenon. To the best of our knowledge, these two studies are the first to investigate the subjectivity characteristics of *pro-drop*. The model assumes that a dropped SoC is more subjective than an explicit SoC, but less subjective than an implicit SoC. Results show that *pro-drop* as a category rarely occurs in the relations with subjective *kějiàn* but mostly in its objective counterparts with *yúshì*. It may be that *pro-drop* fits best in a volitional content environment. This is obviously an issue for further study, not only on causal relations, but also other categories of coherence relations, and potentially not only restricted to Mandarin Chinese, but also considering other *pro-drop* languages.

6 Conclusion

In this article, we have studied the subjectivity profiles of four Mandarin RESULT connectives *kějiàn*, *suǒyǐ*, *yīncǐ*, and *yúshì*. It was based on an adapted model of analysis that integrated three subjectivity features (which function as three subjectivity predictors), namely, the propositional attitude of the consequent (domain), the identity of the SoC, and the linguistic realization of the SoC. The investigation was based on spontaneous oral speech, microblog messages and newspaper articles. The analyses show that these connectives differ systematically from each other with regard to the three subjectivity features, and that the differences have remained by and large stable across discourse types from

each corpus. The propositional attitude proved to be the most important feature, followed sequentially by the linguistic realization and the identity of the *SoC*.

In conclusion, this study demonstrates that causality and subjectivity are two cognitive notions that organize our knowledge of causal coherence relations. Notions like causality and subjectivity can help us explain the system underlying the categorization of causal relations and their linguistic expressions in everyday language use. In this study, we applied these notions to the analysis of Chinese RESULT connectives in three discourse types that are theoretically different in the overall contexts. In this way, it contributes to the subjectivity approach to Mandarin causal connectives and thereby provides a stepping stone for future studies.

Abbreviations

1	first person
2	third person
ADV	adverb(ial)
ATTR	attributive
AUX	auxiliary
CLF	classifier
COP	copula
EMP	emphatic
GEN	genitive case
INF	infinitive
INJ	interjection
MOD	mood
NEG	negation
OBJ	object
PL	plural
PREP	preposition
PRO	pronoun
PRT	particle
SG	singular
SUPL	superlative
TAM	tense

Acknowledgments: We would like to thank the editors and two anonymous reviewers for their comments on previous versions of this paper. We also thank the members in the Persuasive Communication group of the Centre for Language Studies, Radboud University Nijmegen, for their insightful feedback on the presentations of this study.

Data sources

<https://doi.org/10.5281/zenodo.4817295>.

BCC: Beijing Language and Culture University Corpus Center (BLCU BCC). <http://bcc.blcu.edu.cn/>.
Linguistic Data Consortium (LDC), University of Pennsylvania. <https://www ldc.upenn.edu/>.

Media Language Corpus (MLC). 2005. Monitoring and Research Center of National Broadcast Media Language Resources, Communication University of China. <http://ling.cuc.edu.cn/RawPub/>.

References

- Anderson, Marta. 2016. *The Architecture of RESULT relations: Corpus and experimental approaches to RESULT coherence relations in English*. Stockholm: Stockholm University dissertation.
- Austin, John Langshaw. 1962. *How to do things with words*. Oxford: Oxford University Press.
- Canavan, Alexandra & Zipperlen George. 1996. *CALLHOME Mandarin Chinese speech LDC96S34*. Web download. Philadelphia, PA: Linguistic Data Consortium.
- Chafe, Wallace. 1982. Integration and involvement in speaking, writing, and oral literature. In Deborah Tannen (ed.), *Spoken and written language: Exploring orality and literacy*, 35–53. Norwood, NJ: Ablex.
- Chafe, Wallace. 1984. Integration and involvement in spoken and written language. In Tasso Borbé (ed.), *Semiotics unfolding*, 1095–1102. Berlin & New York: Mouton de Gruyter.
- Clark, Herbert H. 1996. *Using language*. Cambridge: Cambridge University Press.
- Clark, Herbert H. & Susan E. Brennan. 1991. Grounding in communication. In Lauren B. Resnick, John M. Levin & Stephanie D. Teasley (eds.), *Perspectives on socially shared cognition*, 127–149. Washington, DC: American Psychological Association.
- De Smet, Hendrik & Jean-Christophe Verstraete. 2006. Coming to terms with subjectivity. *Cognitive Linguistics* 17(3). 365–392.
- Degand, Liesbeth & Henk Pander Maat. 2003. A contrastive study of Dutch and French causal connectives on the Speaker Involvement Scale. In Arie Verhagen & Jeroen Maarten van de Weijer (eds.), *Usage-based approaches to Dutch*, 175–199. Utrecht: LOT.
- Field, Andy. 2013. *Discovering statistics using IBM SPSS statistics*, 4th edn. Los Angeles & London: SAGE.
- Finegan, Edward. 1987. On the linguistic forms of prestige: Snobs and slobs using English. In Phillip C. Boardman (ed.), *The legacy of language: A tribute to Charlton Laird*, 146–161. Reno, NV: University of Nevada Press.
- Fung, Pascale, Shudong Huang & David Graff. 2005. *HKUST Mandarin telephone speech/transcripts, part 1 LDC2005S15/T32*. Web download. Philadelphia, PA: Linguistic Data Consortium. <https://doi.org/10.1515/cog.2006.011>.
- Glenn, Meghan, Haejoong Lee, Stephanie Strassel & Kazuaki Maeda. 2013–2016. *GALE Phase 2–4 Chinese broadcast conversation transcripts LDC2013T08; 2014T28; 2015T09; 2016T12*. Web Download. Philadelphia, PA: Linguistic Data Consortium.
- Groupe Lambda-I. 1975. Car, parce que, puisque. *Revue Romane* 10. 248–280.

- Guo, Jimao. 2006. 'Yúshì' he 'suǒyì' de yitong [A contrastive analysis between *yúshì* and *suǒyì*]. *Hanyu xuebao [Chinese Linguistics]* 4. 27–34.
- Halliday, Michael A. K. & Hasan. Ruquaiya. 1976. *Cohesion in English*. London: Pearson Longman.
- Hobbs, Jerry R. 1979. Coherence and coreference. *Cognitive Science* 3(1). 67–90.
- Huang, C. T. James. 1989. *Pro-drop* in Chinese: A generalized control theory. In Osvaldo Jaeggli & Kenneth J. Safir (eds.), *The null subject parameter*, 185–214. Dordrecht: Springer.
- Huang, Yan. 2006. Speech acts. In Keith Brown (ed.), *The encyclopedia of language and linguistics*, 2nd edn., 656–665. Amsterdam: Elsevier.
- Knott, Alistair & Robert Dale. 1994. Using linguistic Phenomena to motivate a set of coherence relations. *Discourse Processes* 18. 35–62.
- Lambrecht, Knud, Julia Bordeaux & Robert Reichle. 2006. Cognitive constraints on assertion scope: The case of spoken French *parce que*. In Chiyo Nishida & Jean-Pierre Y. Montreuil (eds.), *New perspectives on Romance linguistics, vol. 1: Morphology, syntax, semantics, and pragmatics*, 143–154. Amsterdam & Philadelphia: John Benjamins. https://doi.org/10.1207/s15516709cog0301_4.
- Langacker, Ronald W. 1990. Subjectification. *Cognitive Linguistics* 1(1). 5–38.
- Levshina, Natalia. 2015. *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam & Philadelphia: John Benjamins.
- Li, Shaoqun. 2012. 'Kejian' de biaoji gongneng he yufahua guocheng [The marking function of 'kejian' and its grammaticalization]. *Journal of Northwest University (Philosophy and Social Sciences Edition)* 42(3). 177–179.
- Li, Fang, Jacqueline Evers-Vermeul & Ted Sanders. 2013. Subjectivity and RESULT marking in Mandarin: A corpus-based investigation. *Chinese Language and Discourse* 4(1). 74–119.
- Li, Fang, Ted Sanders & Jacqueline Evers-Vermeul. 2016. On the subjectivity of Mandarin reason connectives: Robust profiles or genre-sensitivity? In Ninke Stukker, Wilbert Spooren & Gerard Steen (eds.), *Genre in language, discourse and cognition*, 15–50. Berlin & Boston: De Gruyter Mouton. <https://doi.org/10.1515/cogl.1990.1.1.5>.
- Lü, Shuxiang (ed.). 1999. *Xiandai hanyu babai ci* [Eight hundred words of modern Chinese], Expanded edn. Beijing: The Commercial Press.
- Mann, William C. & Sandra A. Thompson. 1988. Rhetorical structure theory: Toward a functional theory of text organization. *Text-Interdisciplinary Journal for the Study of Discourse* 8(3). 243–281.
- Martin, James Robert. 1992. *English text: System and structure*. Amsterdam & Philadelphia: John Benjamins. <https://doi.org/10.1075/cld.4.1.03li>.
- Nuyts, Jan. 2001. Subjectivity as an evidential dimension in epistemic modal expressions. *Journal of Pragmatics* 33(3). 383–400.
- Pander Maat, Henk & Liesbeth Degand. 2001. Scaling causal relations and connectives in terms of speaker involvement. *Cognitive Linguistics* 12(3). 211–245.
- Pander Maat, Henk & Ted Sanders. 2000. Domains of use or subjectivity? The distribution of three Dutch causal connectives explained. In Elizabeth Couper-Kuhlen & Bernd Kortmann (eds.), *Cause, condition, concession, contrast: Cognitive and discourse perspectives*, 57–82. Berlin & New York: Mouton de Gruyter.
- Pander Maat, Henk & Ted Sanders. 2001. Subjectivity in causal connectives: An empirical study of language in use. *Cognitive Linguistics* 12(3). 247–274.
- R Core Team. 2017. *R: A language and environment for statistical computing*, version 3.4.1. Vienna: R Foundation for Statistical Computing. <https://www.R-project.org/>.

- Sanders, Ted. 1997. Semantic and pragmatic sources of coherence: On the categorization of coherence relations in context. *Discourse Processes* 24(1). 119–147.
- Sanders, Ted & Wilbert Spooren. 1999. Communicative intentions and coherence relations. In Wolfram Bublitz, Uta Lenk & Eija Ventola (eds.), *Coherence in spoken and discourse*, 235–250. Amsterdam & Philadelphia: John Benjamins.
- Sanders, Ted & Wilbert Spooren. 2001. Text representation: Linguistic and psycholinguistic aspects. In Ted Sanders, Joost Schilperoord & Wilbert Spooren (eds.), *Text representation: Linguistic and psycholinguistic aspects*, 1–24. Amsterdam & Philadelphia: John Benjamins.
- Sanders, Ted & Wilbert 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.
- Sanders, Ted, Wilbert Spooren & Noordman Leo. 1992. Toward a taxonomy of coherence relations. *Discourse Processes* 15(1). 1–35.
- Sanders, Ted, José Sanders & Eve Sweetser. 2009. Causality, cognition and communication: A mental space analysis of subjectivity in causal connectives. In Ted Sanders & Eve Sweetser (eds.), *Causal categories in discourse and cognition*, 19–59. Berlin & New York: Mouton de Gruyter.
- Spooren, Wilbert, Ted Sanders, Mike Huiskes & Liesbeth Degand. 2010. Subjectivity and causality: A corpus study of spoken language. In Sally Rice & John Newman (eds.), *Empirical and experimental methods in cognitive/functional research*, 241–255. Stanford, CA: CSLI Publications.
- Stukker, Ninke, Ted Sanders & Arie Verhagen. 2008. Causality in verbs and in discourse connectives: Converging evidence of cross-level parallels in Dutch linguistic categorization. *Journal of Pragmatics* 40(7). 1296–1322.
- Sweetser, Eve. 1990. *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure*. Cambridge: Cambridge University Press. <https://doi.org/10.1080/01638539209544800>.
- Traugott, Elizabeth Closs. 1989. On the rise of epistemic meanings in English: An example of subjectification in semantic change. *Language* 65(1). 31–55.
- Traugott, Elizabeth Closs. 1995. Subjectification in grammaticalisation. In Dieter Stein & Susan Wright (eds.), *Subjectivity and subjectivisation: Linguistic perspectives*, 31–54. Cambridge: Cambridge University Press.
- van Dijk, Teun A. 1979. Pragmatic connectives. *Journal of Pragmatics* 3(5). 447–456.
- Voiskounsky, Alexander E. 1997. Telelogue conversations. *Journal of Computer-Mediated Communication* 2(4). <https://doi.org/10.1111/j.1083-6101.1997.tb00194.x>.
- Walker, Kevin, Christopher Caruso, Kazuaki Maeda, Denise DiPersio & Stephanie Strassel. 2013–2016. *GALE Phase 2–4 Chinese broadcast conversation speech LDC 2013S04; 2014S09; 2015S06; 2016S03*. DVD/Web download. Philadelphia, PA: Linguistic Data Consortium. <https://doi.org/10.2307/414841>.
- Wheatley, Barbara. 1996. *CALLHOME Mandarin Chinese transcripts LDC96T16*. Web download. Philadelphia, PA: Linguistic Data Consortium.
- Wu, Zhibiao. 1995. *Mandarin Chinese News Text LDC95T13*. Web Download. Philadelphia, PA: Linguistic Data Consortium. [https://doi.org/10.1016/0378-2166\(79\)90019-5](https://doi.org/10.1016/0378-2166(79)90019-5).
- Xiao, Hongling, Fang Li, Ted Sanders & Wilbert Spooren. 2021. How subjective are Mandarin REASON connectives? A corpus study of spontaneous conversation, microblog and newspaper discourse. *Language and Linguistics* 22(1). 166–210.
- Xing, Fuyi. 2001. *Hanyu fujū yanjiu* [A study of Chinese complex sentences]. Beijing: The Commercial Press.

- Xun, Endong, Gaoqi Rao, Xiaoyue Xiao & Jiaojiao Zang. 2016. Da shuju beijingxia BCC yuliaoku de yanzhi [The construction of the BCC corpus in the age of big data]. *Yuliaoku yuyanxue [Corpus Linguistics]* 3(1). 93–109.
- Yus, Francisco. 2011. *Cyberpragmatics: Internet-mediated communication in context*. Amsterdam & New York: John Benjamins.
- Zhao, Xin. 2003. ‘Yīncǐ, yúshì, cong’er’ de duo jiaodu fenxi [A multiple-perspective analysis of yīncǐ, yúshì and cong’er]. *Yuyan yanjiu [Linguistic Research]* 1. 26–34.
- Zufferey, Sandrine. 2012. *Car, parce que, puisque* revisited: Three empirical studies on French causal connectives. *Journal of Pragmatics* 44(2). 138–153.