

# Subjectivity and result marking in Mandarin

## A corpus-based investigation\*

Fang Li, Jacqueline Evers-Vermeul and Ted J.M. Sanders  
Utrecht University

Recent corpus studies have shown that differences in subjectivity – the degree to which speakers express themselves in an utterance – can account for the usage of causal connectives (*because, so*) in major European languages. If the notion of subjectivity is a basic cognitive principle, it ought to play a role in the description of connectives in other languages. In this paper, we present a corpus analysis of five Mandarin result connectives, *kějiàn*, *suōyǐ*, *yīncǐ*, *yīn'ér*, and *yúshì*. We used four subjectivity indicators: modality, domain (following Sweetser 1990), and the presence and identity of a Subject of Consciousness – the person responsible for constructing the causal relation. Results show that *kějiàn*, *suōyǐ*, *yīncǐ*, *yīn'ér*, and *yúshì* display different degrees of subjectivity. To a large degree, our findings corroborate previous observations in the literature (e.g. the ones in Xing 2001). However, our analysis also shows that while *kějiàn*, *yúshì*, and *yīn'ér* have robust profiles across genres, the subjective meanings of *suōyǐ* and *yīncǐ*, two common connectors, are genre sensitive.

**Keywords:** subjectivity, causal connectives, result connectives, reason connectives, subject of consciousness, domain

**关键词：**主观性，因果关系词，果标，因标，意识主体，域

### 1. Introduction

With the rise of functional linguistics, pragmatics, and Cognitive Grammar, the concept of ‘subjectivity’ has gained increasing prominence (see De Smet & Verstraete 2006; Langacker 1985; Lyons 1977, 1982a, 1995; Shen 2001; Traugott 1989, 1995, 2010; Zhang 1994). According to Lyons (1977), the linguistic notion of subjectivity refers to the speaker’s self-expression in the utterance. Speakers often express their attitude and emotion towards the propositions in an utterance. For instance, speakers sometimes explicitly express their attitude by using evaluative

adjectives, such as *wonderful* in the utterance *It is a wonderful stamp*. Alternatively, speakers can judge a situation by means of modal adverbs such as *probably* in the utterance *She probably fell ill*. Speakers can also express themselves by using other subjectivity markers such as prepositions (Langacker 1985) and connectives (Sanders & Sweetser 2009; Traugott 1995). The theme of this study is to examine the role of subjectivity in the use of result connectives in Mandarin Chinese.

Causal connectives are explicit linguistic markers of causal relations between adjacent segments (Murray 1997; Stukker & Sanders 2012). They can be used to mark the reason segment or the result segment of the causally related clauses. Accordingly, they divide into reason connectives and result connectives. In English, the result connective *so* can be used to mark the cause-consequence relation in (1), and the argument-claim relations in (2) and (3).

- (1) It was a very hot day, *so* Bob went swimming.
- (2) John is ill, *so* he probably won't come to the meeting.
- (3) The lights are out in the house, *so* nobody is at home.

In (1), *so* marks the fact that *Bob went swimming* as the consequence of the situation that *it was a very hot day*. In (2) and (3), *so* is used to present a conclusion or an inference on the basis of a fact or an observation. English speakers also have other linguistic markers at their disposal that seem specialized in expressing specific types of causal relations. For instance, cause-consequence relations, such as (1), are often marked by the linguistic cue phrase *that's why*, and the connective *therefore* is typically used to mark causal relations at an epistemic or argumentative level, as in (2).

In Mandarin Chinese, there are at least five connectives that can be considered to be the counterparts of the English *so/therefore*: *kějiàn*, *suǒyǐ*, *yīncǐ*, *yīn'ér*, and *yúshì* (see Chinese Academy of Social Sciences 2002). Just as *so/therefore*, the semantics of the five Mandarin connectives denote that the segment containing them expresses the consequent of the causally related events or states. That is why they are called result connectives. In terms of their syntactic status, they are as highly grammaticalized as *so/therefore*: they are acknowledged conjunctions.<sup>1</sup> The usage of the five Mandarin connectives is illustrated in (1')–(3').

- (1') Tiānqì fēicháng de rè, *yúshì/suǒyǐ* Bob qù yóuyǒng le.  
weather very MOD hot, *so* Bob go swim ASP  
'It was a very hot day, *so* Bob went swimming.'
- (2') John bingle, *yīncǐ/yīn'ér* tā kěnéng bùhuì lái kāihuì le.  
John ill:ASP, *so* 3SGM probably NEG will come meeting PRT  
'John is ill, *so* he probably won't come to the meeting.'

- (3') Fángzi li de dēng hēizhe, *kějiàn* méirén zài jiā.  
 house in MOD light dark:ASP, so nobody at home  
 'The lights are out in the house, so nobody is at home.'

Although *kějiàn*, *suǒyǐ*, *yīncǐ*, *yīn'ér*, and *yúshì* share great similarity in terms of semantics and syntax, they are only partially interchangeable in these contexts. For example, *yīncǐ* can be substituted for *yúshì* in (1') to express the causal connection between observed situations; however, *yúshì* cannot be substituted for *yīncǐ* in (2') to link a conclusion with its argument. The fact that Mandarin has at least five conjunctions to mark the consequent of causal relations and that these conjunctions are only partially interchangeable suggests a division of labor between these closely related connectives. How can this division of labor be characterized?

In this study, we will characterize these five Mandarin connectives in terms of subjectivity. One way of looking at subjectivity is to examine the type of causality marked by each connective. In the literature about English and other European languages, dichotomous distinctions have been made between external/internal causality (Halliday & Hasan 1976; Martin 1992), semantic/pragmatic causality (Moeschler 1989; Sanders 1997; Sanders, Spooren & Noordman 1992; Van Dijk 1979), and subject matter/presentational causality (Mann & Thompson 1986, 1988). A more fine-grained categorization is proposed by Sweetser (1990), who distinguishes between three domains of causal relations: the content domain (which corresponds to the external/semantic/subject matter type of causality), the epistemic domain, and the speech-act domain. According to Sweetser (1990:77), the content domain concerns "real-world causality" holding between events in the described world; in the epistemic domain the speaker's knowledge is involved as the basis for a logical conclusion; and in the speech-act domain the reason clause gives the cause of the speech act embodied in the main clause.

Shen (2008) has applied the Domain Theory to the interpretation of complex sentences in Chinese discourse. He observes that the division between three domains helps to account for the semantic relations in causally related sentences such as (4)–(6). Sentences (4)–(6) can actually serve to exemplify the content, epistemic, and speech-act domain, respectively.

- (4) Zhāng Gāng huíláile, *yīnwèi* tā hái ài Xiǎoli.  
 NAME come:back:ASP, because 3SGM still love NAME  
 'Zhang Gang has come back because he still loves Xiaoli.'
- (5) Zhāng Gāng hái ài Xiǎoli, *yīnwèi* tā huílái le.  
 NAME still love NAME, because 3SGM come:back ASP  
 'Zhang Gang still loves Xiaoli, because he has come back.'

(6) Zhāng Gāng hái huílái ma? *Yīnwèi* Xiǎoli zài děng tā.

NAME still come:back PRT? *because* NAME ASP wait 3SGM

‘Will Zhang Gang still come back? *Because* Xiaoli is waiting for him.’

Apparently, *yīnwèi*, just like *because*, can be used across domains. Now, it is interesting to ask whether or not there are causal connectives in Mandarin that specialize in specific domains, and whether the theory of subjectivity can help categorize causal connectives in Mandarin discourse. In fact, a number of studies have shown that causal connectives in European languages such as French, German, and Dutch can be categorized in terms of subjectivity (see Sanders & Sweetser 2009 for an overview). It is shown, for example, that subjectivity determines the choice of Dutch result connectives *dus* ‘so’ versus *daarom* ‘that’s why’ (Pander Maat & Sanders 2001) and Dutch reason connectives *omdat* ‘because’ versus *want* ‘because’ in a variety of discourse modes (Degand & Pander Maat 2003; Pit 2003; Sanders & Spooren 2009, submitted; Spooren, Sanders, Huiskes & Degand 2010). Investigations have also been made into other European languages such as French, German, and Polish (see Dancygier 2009; Evers-Vermeul, Degand, Fagard & Mortier 2011; Keller 1995). The results show that causal connectives of these languages can also be distinguished in terms of the degree of subjectivity they encode. For instance, French *car* and *puisque*, and German *denn* mark higher degrees of subjectivity in the sense that they are typically used to express epistemic causal relations that are constructed with higher speaker-involvement (see (2)). In contrast, French *parce que*, and German *weil* mark lower degrees of subjectivity in the sense that they typically involve lower speaker-involvement in constructing content causal relations as manifested in (1) (see Stukker & Sanders 2010, 2012).<sup>2</sup>

The studies mentioned above show that subjectivity plays an important role in categorizing causal connectives in major European languages. In fact, the theory of subjectivity is derived mainly from Germanic data. If we take seriously the proposals that the linguistic categories apparent in people’s everyday language use provide us with insights into the working of the mind (Lakoff 1987; Lakoff & Johnson 1999) and that the notion of subjectivity is likely to be a general cognitive mechanism underlying the categorization of causal connectives across different languages (Sanders & Sweetser 2009), then it would be reasonable to assume that subjectivity as a general factor should manifest in other language families as well. Therefore, we hypothesize subjectivity to be relevant to the categorization of causal connectives in Mandarin, in spite of the fact that Mandarin is typologically different from most of the major European languages.

Indeed, in the literature on Mandarin causal connectives there have been theoretical as well as corpus-based studies that suggest the possibility of categorizing Mandarin causal connectives with respect to subjectivity (Li 2011; Li & Liu 2004;

Shen 2008; Xing 2001; see also Y. Zhang 2012 for the categorization of reason connectives in terms of intersubjectivity). Referring to examples (4)–(6), Shen (2008) suggests that causal relations in Mandarin discourse can be categorized into different domains of causality, and therefore reflect different levels of subjectivity. Furthermore, Xing (2001) demonstrates systematic differences among a number of reason connectives in Mandarin discourse, as exemplified in (7) and (8).

- (7) *Yīnwèi* lǎoshī yǒu shì, xiàwǔ wǒmen zìxí.  
*because* teacher have thing, afternoon 1PL self:study  
 ‘Because the teacher has some other business to do, this afternoon we study by ourselves.’
- (8) *Jìrán* lǎoshī yǒu shì, xiàwǔ wǒmen zìxí.  
*since* teacher have thing, afternoon 1PL self:study  
 ‘Since the teacher has some other business to do, this afternoon we study by ourselves.’

According to him, *yīnwèi* as used in (7) is to mark the ‘descriptive causality’ on the basis of facts, whereas *jìrán* ‘since’ as used in (8) typically expresses an ‘inferential causality’ on the basis of rationality. We believe Xing’s concepts of descriptive causality and inferential causality to mirror the distinction between the content and the epistemic domains. In (7), the content causal relation holds between two observable situations, and there is no speaker involvement. In the epistemic relation in (8), by contrast, the speaker is directly involved in linking an argument with a conclusion, and thus we consider (8) to be more subjective than (7).

The examples provided by Xing (2001) only concern reason connectives (i.e. causal connectives that are used to mark the reason segment). Although the author does not provide examples to illustrate the systematic difference between result connectives, he does remark that the result connectives *suǒyǐ*, *yīncǐ*, and *yīn’ér* typically occur in descriptive causality, whereas *kějiàn* typically marks inferential causality (Xing 2001: 40). On the basis of this claim, we should expect to find more content use of the first three result connectives, and more epistemic relations in sentences connected by *kějiàn*.

Xing (2001: 527) claims that in clauses connected with *yúshì*, the second clause often expresses a natural consequence of the state of affairs, judgment, or some kind of feeling expressed in the first clause. This seems to suggest that *yúshì* occurs more often in the content domain to describe cause-consequence relations existing in the physical world. This is an issue we will investigate in this study.

To sum up, previous studies touched upon the issue of subjectivity in causal marking, reason marking in particular, in Chinese language and discourse. Concerning result marking, however, no systematic corpus-based investigations have been undertaken so far to explore conceptual differences among result

connectives in terms of subjectivity. In this study, we conducted a corpus-based analysis of the five Mandarin result connectives that are apparent equivalents in terms of both syntactic and semantic properties — *kějiàn*, *suǒyǐ*, *yīncǐ*, *yīn'ér*, and *yúshì* — to study whether and how subjectivity categorizes causal connectives in a non-Indo-European language such as Mandarin Chinese. Our research questions are:

1. Are previous claims (e.g. the ones in Xing 2001) about the usage of Mandarin result connectives supported by the corpus study?
2. Do the five result connectives *kějiàn*, *suǒyǐ*, *yīncǐ*, *yīn'ér*, and *yúshì* show systematic variation in terms of subjectivity?
3. Are the observed subjectivity profiles sensitive to text genres?

Studies on French, German, and Dutch causal connectives have already shown that the distribution of connectives over causal categories seems to vary in relation to the context (Degand & Pander Maat 2003; Frohning 2007; Pit 2003; Zufferey 2012). Subjective causal connectives such as French *car*, German *denn*, and Dutch *want*, which are all roughly translated as ‘because’, display consistent usage patterns across text types such as newspapers, novels, and periodicals. By contrast, usage patterns of their objective counterparts, French *parce que*, German *weil*, and Dutch *omdat*, are less consistent across these text genres (Stukker & Sanders 2012). These results raise the question whether the subjectivity profiles are really part of the inherent semantic characteristics of the connectives themselves, or whether they are (partially) determined by the pragmatics of the context.

The paper is organized as follows. In Section 2, theories about the linguistic notion of subjectivity are discussed in detail as the basis for the operationalization in our corpus-based analysis. In Section 3, we introduce our research method. In Section 4, the results of the corpus analysis are reported, showing how exactly the five causal connectives display differences along the dimension of subjectivity. In Section 5, the results are discussed further and suggestions are made for future studies.

## 2. Subjectivity defined

In order to give a precise characterization of the degrees of subjectivity expressed by the five result connectives, we adopt an integrative approach to the issue of subjectivity. In spite of various differences in their formulations, most definitions refer to subjectivity as a specific property of the utterance: speaker-relatedness. For example, Traugott (2010:30) defines subjectivity as the “relationship to the speaker and the speaker’s beliefs and attitudes”. In the same vein, De Smet and Verstraete

(2006: 365) claim that subjectivity “covers the fact that a particular element or construction requires reference to the speaker in its interpretation”. While focusing on discourse, Sanders and Spooren (2009, submitted) argue that speaker involvement is an important property of subjectivity in causally related sentences as well as in isolated sentences. Examples (9) and (10) illustrate their idea.

- (9) Quánqiú de yěyāróng niánchǎnliàng bùzú 1000 gōngjīn, yīn’ér jiàgé fēicháng ángguì.  
 global MOD wild:duck:down yearly:output NEG:enough 1000 kilograms,  
 as:a:result price very expensive  
 ‘The global annual output of wild-duck down is less than 1000 kilograms; as a result, the price (of wild-duck down) is very expensive.’
- (10) Zhè zhǒng yào hányǒu wēiliàng de yǒudú wùzhì, yīncǐ chángqī fúyòng hěn kěnéng duì jiànkāng bùlì.  
 this kind medicine contain tiny:amount MOD poisonous substance, so  
 long:term use very probably for health NEG:favorable  
 ‘This kind of medicine contains a tiny amount of poisonous substance, so long-term use (of the medicine) is probably unfavorable to the health.’

In (9), the relation between the antecedent and the consequent is objective, because it concerns a relation in an objective reality, which does not depend on the speaker’s reasoning or intentionality. In contrast, (10) is subjective because it involves a subjective construction of the causal connection, which is not apparent in the physical world but exists in the speaker’s mind. For cases like this, we need to refer to the speaker’s attitude for its interpretation. Therefore, on the basis of the property of speaker-relatedness we can make a division between subjectively constructed and objectively described causal relations.

In order to track any differences in the degree of subjectivity of the Mandarin result connectives, we go beyond the absence versus presence of speaker-relatedness. On the basis of earlier work on causal connectives (e.g. Pander Maat & Sanders 2000), subjectivity has also been defined in terms of the presence of an active ‘subject of consciousness’ (henceforth SoC) (see Sanders & Spooren submitted; Sanders & Sweetser 2009). An SoC is “an animate subject, a person, 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 & Sanders 2001: 251). The SoC may be the actual speaker or the character(s) in the utterance. Accordingly, the notion of subjectivity has been extended beyond speaker-relatedness: it could be character-related as well. To give an illustration, (11) is considered subjective because it involves a conclusion drawn by a speaker SoC, while (12) can also be called subjective because it involves the reasoning of a character SoC.

- (11) The lights are out in the house, so I think nobody is at home.
- (12) The lights are out in the house, so John thinks nobody is at home.

Both (11) and (12) contain a subjectively constructed causal relation because they are both SoC-related. They involve the SoC's point of view towards the state of affairs in the world. The author/speaker can be considered the first voice in the discourse, who has constant access to his/her feelings and thoughts. S/he does not have access to the feelings and thoughts of a third person. As a result, *I think Utrecht is nice* can be a direct report of an inner feeling, whereas *He thinks Utrecht is nice* is a description of an evaluation. As a consequence, first person evaluations are more subjective than third person evaluations. What distinguishes the speaker/author from a character as SoC is that the speaker concerns a first voice, which is grounded in the Deictic Centre of Communication (Sanders, Sanders & Sweetser 2009). This reflects Traugott's (1989, 1995) view on subjectivity as closeness to the communicative "here and now": the speaker/author here and now asserts that a particular state of affairs holds. By contrast, the character type concerns a third person in the discourse, who is more distant from the Deictic Centre of Communication. The notion of SoC, and the distinction we make here between speaker/author and character SoC, is related to the notion of *perspective* in literary studies (Fludernik 1993) and *voice* in linguistics (Ducrot 1980): the speaker/author is the first voice in the discourse, and the character is another person whose inner thoughts and evaluations can be accessible in a narrative (Sanders 2010).

Closely related to the concept of SoC is the Domain Theory, which we have briefly mentioned in the Introduction. Many previous studies show that the division between causal domains is an effective way to operationalize subjectivity in causally related sentences (see Degand & Pander Maat 2003; Evers-Vermeul *et al.* 2011; Sanders & Spooren 2009). These and other studies rank causal domains in terms of subjectivity according to the degree to which they necessarily imply the subjective involvement of an SoC. In the content domain, the speaker describes a causal relation in the physical world, as in (9). This content domain can be subdivided into two subtypes (Stukker, Sanders & Verhagen 2008): the volitional content domain always involves an SoC who performs an intentional act (see (13)), whereas the non-volitional content domain does not have an SoC at all (see (14)).<sup>3</sup>

- (13) It was a very hot day; that's why Bob went swimming.
- (14) Bob fell off the bike. As a result, he got hurt in the left leg.

Due to the absence of an SoC, the non-volitional content domain is considered the least subjective domain: it concerns causal processes in which human intentions do not play a role. Volitional relations carry a higher degree of subjectivity than



non-volitional ones, because human volition or motivation is a prerequisite for performing acts. Epistemic relations are of still higher subjectivity because they directly involve speakers' or characters' opinions, beliefs or inferences, as in (10)–(12). Speech act causal relations are also highly subjective because performing a speech act (making a promise, issuing a command, or raising a question, as in (6)) is bound to a situation in which the speaker is present.

As a final part of our integrative approach to subjectivity, we adopt Langacker's (1990) interpretation of subjectivity in terms of the explicitness of a ground element, an element that refers to the speech event, its participants and its immediate circumstances. From Langacker's point of view, the defining standard of subjectivity is not only speaker-relatedness, but also whether or not the speaker remains implicit in the utterance, without any formal linguistic encoding. An utterance is objectified when the speaker is put "on stage" and becomes observable (see Langacker 1985, 1990). For example, despite the fact that the causal relation expressed in (10) (repeated here as (15) for convenience) is identical to the one in (15'), we consider (15') as less subjective than (15) because the speaker is on stage, as an explicit reference point.

(15) Zhè zhǒng yào hányǒu wēiliàng de yǒudú wùzhì, yīncǐ chángqī fúyòng hěn kěnéng duì jiànkāng bùlì.

this kind medicine contain tiny:amount MOD poisonous substance, so long:term use very probably for health NEG:beneficial

'This kind of medicine contains a tiny amount of poisonous substance, so long-term use (of the medicine) is probably detrimental to health.'

(15') Zhè zhǒng yào hányǒu wēiliàng de yǒudú wùzhì, yīncǐ wǒ rènwéi chángqī fúyòng hěn kěnéng duì jiànkāng bùlì.

this kind medicine contain tiny:amount MOD poisonous substance, so I think long:term use very probably for health NEG:beneficial

'This kind of medicine contains a tiny amount of poisonous substance, so I think long-term use (of the medicine) is probably detrimental to health.'

### 3. Corpus-based analysis

#### 3.1 Mandarin result connectives

In the present study, we aim to investigate the subjectivity profiles of five Mandarin result connectives in a quantitative corpus-based analysis. We focus on single words, highly grammaticalized *guānxìcí* 'connectives' that are used to mark the consequent clause: *kějiàn*, *suǒyǐ*, *yīncǐ*, *yīn'ér*, and *yúshì*.<sup>4</sup> The frequencies of these

**Table 1.** Frequency of the result connectives in *the Lancaster Corpus of Mandarin Chinese* (LCMC) (McEnery & Xiao 2004)<sup>5</sup>

Connective	Absolute frequency	Frequency per 10,000 words
<i>yīncǐ</i>	439	4.39
<i>suǒyǐ</i>	393	3.93
<i>yúshì</i>	274	2.74
<i>yīn'ér</i>	150	1.50
<i>kějiàn</i>	59	0.59

result connectives are different (see Table 1). *Yīncǐ* is the most frequent of the five, followed in turn by *suǒyǐ*, *yúshì* and *yīn'ér*. The least frequent one is *kějiàn*.

In the *Contemporary Chinese Dictionary* (Chinese Academy of Social Sciences 2002), *yúshì* is annotated as indicating that the latter event immediately follows the former, and that the latter is often caused by the former (see also Lü 1999). On the grounds that it describes causal connections between physical events, we may predict that *yúshì* expresses causality within the content domain. This prediction is also in line with Xing's interpretation of the word (as discussed in Section 1). It also conforms to *yúshì*'s closeness to the descriptive temporal usage (Xing 2001; Zhang 2008). Xing (2001) labels *yúshì* as a 'cross-relation marker', because it is used to mark a relation of temporal succession as well as causality. For instance, in (16) *yúshì* is strictly temporal; it merely expresses the successive order of two connected events (see Xing 2001:528). Still, *yúshì* can also express relations that are solely causal. For example, in (17), *yúshì* connects a conclusion with its argument, and temporal ordering is irrelevant. Lu (2000) claims that *yúshì* cannot express inferences or conclusions, but (17) already serves as a counterexample. We will testify this further with the corpus data.

- (16) Guòle nà lín, chuán biàn wānjìnle Línggǎng, *yúshì* Zhào zhuāng biàn zhēn zài yǎnqián le.  
 pass:ASP that woods, boat then bend:into:ASP Ling:harbor, then Zhao:village then really at eye:front ASP  
 'The boat passed the woods, sailed into the Ling Harbor, and then Zhao Village was indeed in front of us.'
- (17) Jì méiyǒu yǒuxiào de jīngjì zérèn shěnjì, yòu quēfá biyào de juécè shīwù zhuīchéng, *yúshì* chāobiāo jiànzhù, "lànwěigōngchéng" de cūxiàn, yě jiù bù língréncHàyì le.  
 both NEG:have effective MOD financial responsibility auditing, and lack necessary MOD decision mistake chase:penalty, so exceed:standard building, waste:real:estate:project MOD appearance, likewise then NEG surprising PRT

‘There is neither effective auditing of financial responsibility nor essential penalty for mistakes on a decision, so it is not surprising that super-luxurious buildings and wasted real estate projects appeared.’

The dictionary annotation of *kějiàn* presents us a totally different picture. *Kějiàn* is translated as ‘it is thus obvious that, it shows that, that proves, so’. Clearly, its English translations are all closely related to opinions, beliefs, or conclusions, on the basis of which we may predict that *kějiàn* expresses causality within the epistemic domain. This prediction conforms to the word’s lexical origin. The conjunction *kějiàn* was grammaticalized from a verbal phrase *kě jiàn* (Liu & Yao 2011; Q. Zhang 2012), which can be roughly glossed as ‘can see’. Given this root meaning, it seems logical to expect that *kějiàn* is used in the subjective epistemic domain, expressing the speaker’s conclusion. In particular, the word *jiàn* ‘see’ has been demonstrated as a subjectivity indicator, which encodes perspective, hence the increased involvement of the speaker/writer (see Tao 2007 for evidence of *jiàn* indicating subjectivity in the context of existential/presentative constructions). In fact, this prediction about the use of *kějiàn* is also consistent with Xing’s (2001) claim that *kějiàn* typically expresses inferential causality on the basis of rationality.

The dictionary annotation does not give us a clear clue on the profile of *yīncǐ*, *yīn’ér*, or *suǒyǐ*, because only the English translation *so/therefore* is given in the dictionary and no further explanation can be found there. However, Xing (2001) has observed that these three connectives typically express descriptive causality on the basis of facts. Accordingly, we should expect them to occur more often in the objective content domain.

### 3.2 Corpus and sampling method

In order to study the impact of text genre on the connective meaning and use, i.e. the third research question, we collect the samples in a balanced way from three major discourse genres: argumentative, informative, and narrative. This is designed also to minimize any contextual bias towards discourse interpretation as earlier studies have shown that in canonical cases contextual factors play a smaller role in determining the coherence relation; however, when the relation is ambiguous, context can have a major impact on the interpretation (Sanders 1997). Language users show a systematic preference to interpret ambiguous cases as semantic in descriptive contexts, and a tendency to interpret them as pragmatic in argumentative contexts.

The argumentative and informative genres of fragments are taken from PPD: *People’s Daily Online* ([http://search.people.com.cn/rmw/GB/rmwsearch/gj\\_search\\_pd.jsp](http://search.people.com.cn/rmw/GB/rmwsearch/gj_search_pd.jsp)), which is a comprehensive and influential website claiming the

largest daily amount of news releases in Mainland China. As it has a database of 20 billion characters, it is an excellent data source for analyzing newspaper texts. The informative fragments are taken from the category of *Technology*, and the argumentative fragments are taken from the category of *Opinion*.

Although *People's Daily Online* is a large database for linguistic investigations into Mandarin Chinese, it is limited to newspaper articles. Thus, we look to another corpus for narratives: the CCL Corpus. This corpus, created by the Center for Chinese Linguistics of Peking University, consists of 477 million characters. The contemporary Chinese part of the corpus is divided into various subject areas, such as *Novel*, *Prose*, and *Biography*. The narrative fragments are taken from the category of *Novel*.

For each genre, seventy-five occurrences of each connective were randomly selected. See Table 2.

Table 2. Number and nature of connective fragments in the sample

Connective	Argumentative (PPD)	Informative (PPD)	Narrative (CCL)	Total
<i>kějiàn</i>	75	75	75	225
<i>suǒyǐ</i>	75	75	75	225
<i>yīncǐ</i>	75	75	75	225
<i>yīn'ér</i>	75	75	75	225
<i>yúshì</i>	75	75	75	225
Total	375	375	375	1125

For the 1125 fragments, special care has been taken to make sure that the connectives are used as causal connectives in every instance. For example, the temporal usage of *yúshì* in (16) is not included in our collection of fragments. In our final sample, four cases of *yúshì* are discarded due to its temporal usage. In addition, cases such as (18) are also excluded, because *kě jiàn* is used as a predicate 'may be seen' rather than as a connective. In total, six occurrences of verbal *kě jiàn* are replaced with causal uses.

- (18) Báifā yǐ qīngxī kě jiàn.  
 white:hair already clearly *can see*  
 'Grey hair *can* already *be seen* clearly.'

### 3.3 Analysis

In order to systematically measure the degree of subjectivity each causal connective expresses, we use an analytical model to operationalize the notion of subjectivity. The analytical model is created based on the integrative approach to subjectivity

described in Section 2. It contains all the important components of subjectivity: domain, the presence of an SoC, and the identity of the SoC. Apart from these three factors, we take a fourth variable into consideration: modality of Q (i.e. the consequent). We decide to add this variable because modality has been shown to be an important indicator of subjectivity in previous analyses (Sanders & Spooren 2009). Empirically, it also proved to be effective for several Germanic languages (Degand & Pander Maat 2003; Pander Maat & Degand 2001; Spooren *et al.* 2010).

Our analytical model is presented in Table 3. It shows the above-mentioned four variables together with their subjectivity values, to be elaborated on in the following subsections.

**Table 3.** Model for subjectivity analysis with variables and their values

Variables	+ .....subjectivity values..... -
Domain	Speech-act / Epistemic ... Volitional content .. Non-volitional content
Modality (Q)	Speech-act / Judgment ..... Mental fact ..... Physical fact
The presence of SoC	Implicit ..... Explicit ..... Absent
The identity of SoC	Author ..... Current speaker.... Character

In the literature, the distinction between subjectivity and objectivity is considered to be gradual rather than absolute (Lyons 1982b: 105). Therefore, we do not calculate an exact value of subjectivity for the connectives. Instead, we compare the relative degrees of subjectivity that are encoded in different connectives. For each connective, we use the four variables to examine the degree to which the speaker or SoC is responsible for the construction of the causal coherence relation. A connective is considered to signal a higher degree of subjectivity than another if there is at least one variable that discriminates between the connectives in the more subjective direction (e.g. more epistemic cases) and if none of the other variables shows a preference in the more objective direction (e.g. more physical facts). The fragments were coded independently by one of the authors and two other native speakers of Mandarin, who were working in the field of discourse studies at the Utrecht Institute of Linguistics OTS. The level of inter-rater agreement is quite high (for domain: Kappa = 0.93; for modality of Q: Kappa = 0.94; for presence of SoC: Kappa = 0.90; for identity of SoC: Kappa = 0.88). The team discussed discrepancies among their analyses until they reached an agreement.

### 3.3.1 *Domain*

As discussed in Section 2, we distinguish between four causal domains in the present study: the non-volitional content, volitional content, epistemic, and speech-act domain. In order to interpret the domains accurately, we use a paraphrase test.

The paraphrase test is presented in Table 4, in which P and Q correspond to the antecedent and the consequent of the causal relation, respectively.

**Table 4.** The paraphrase test used in the domain analysis

Domain	Paraphrase
Non-volitional content	P leads to the physical fact/mental fact that Q, and no intention is involved in Q
Volitional content	P leads to intentional physical act/mental act that Q
Epistemic	P leads to claim/decision/inference/conclusion that Q
Speech-act	P leads to question/advice/command/promise that Q

However, in natural discourse the division between domains is not always clear-cut. In cases where ambiguity emerges, we opt for the more objective interpretation. For example,

- (19) Yóuyú tāmen de zhōngzi cúnzàizhe xiūmiánqī, luò dì hòu bù líjì méngfā,  
 érshì yào děngdào cinián chūntiān láilín cái fāyá, yīn'ér nénggòu duǒ guò  
 dōngjì yánhán.  
 because 3PL POSS seed exist:ASP dormancy, fall ground after NEG  
 immediately sprout, but have:to wait:until next:year spring approach CAI  
 sprout, *as:a:result/therefore* can hide ASP winter sever:cold  
 ‘Their seeds have a period of dormancy. They don’t sprout immediately after  
 falling to the ground, but will wait till the next spring. *As a result/Therefore*  
 they can survive the severe cold during the winter.’

The ambiguity of the causal relation expressed in (19) springs from the lexical ambiguity of the word *nénggòu* ‘can’, which may denote capability as well as possibility. Consequently, we can plausibly arrive at two interpretations of (19). One way to interpret it is to treat the word *nénggòu* ‘can’ as expressing a kind of possibility: the fact that the seeds have a period of dormancy leads to a conclusion that it is possible for them to survive the severe cold during the winter. This can be regarded as a subjective epistemic relation. The other interpretation is to construe the *yīn'ér* segment as expressing an observable fact: the fact that the seeds have a period of dormancy leads to the fact that they are able to survive the severe cold in winter. This can be seen as denoting an objective content relation. In the analysis, we interpret the fragment in the latter way, that is, we have adopted an objectivity bias, for the sake of consistency. In fact, an ambiguity like this is often disambiguated by the context. For example, if we precede the sentence with the phrase “In his view,” (see (20)), the second interpretation is immediately ruled out.

- (20) Zài tā kànlái, yóuyú tāmen de zhōngzi cúnzàizhe xiūmiánqī, luò dì hòu bù lìjí méngfā, érshì yào děngdào cìnián chūntiān láilín cái fāyá, yīn'ér nénggòu duǒ guò dōngjì yánhán.  
 at 3SGM see, because 3PL POSS seed exist:ASP dormancy, fall ground after NEG immediately sprout, but have:to wait:until next:year spring approach CAI sprout, *therefore* can hide ASP winter sever:cold  
 'In his view, their seeds have a period of dormancy. They don't sprout immediately after falling to the ground, but will wait till the next spring. *Therefore* they can survive the severe cold during the winter.'

We have encountered a special type of causal relation during the analysis, which we label as 'chain causality'. In the so-called chain causality, readers need to add an inferential link to the causal relation to comprehend the sentence. For example,

- (21) (Bǎishìkèlè Gōngsī) zài qìshuǐ pínggài shàng yìn yǒu shùzì hàomǎ, dàjiǎng yǒu yībǎi wàn bǐsuǒ. *Yúshì* Bǎishìkèlè dàwéichàngxiāo.  
 (Pepsi-Cola Company) at soft:drink bottle:cap on print:have digit number, big:bonus have one:hundred ten:thousand peso. *So* Pepsi-Cola greatly:BEI:sell:well  
 '(The Pepsi-Cola Company) printed numbers on the bottle-caps; the biggest bonus was one million peso. *So* Pepsi was sold very well.'

In this case readers need to add the following inferential link: many customers intentionally went to buy Pepsi-Cola in order to win the big bonus. The next question to consider is: what is the causal domain of (21)? On the one hand, there is an implied SoC in it: the customers. On the other hand, the best paraphrase is "the fact that the Pepsi-Cola Company offered a big bonus leads to the fact that it was sold very well", which suggests that the domain is non-volitional content. During the analysis, we always choose the non-volitional content interpretation, to be consistent with the way in which we deal with ambiguous relations. That is also the case here.

### 3.3.2 Modality

We distinguish four values of modality. We code the result segment as expressing a speech act, a judgment, an observable physical fact, or a mental fact. A segment is coded as a speech act if it is in the form of a general question, a rhetorical question, or an imperative. A segment is considered to express a judgment if it presents opinions, decisions, conclusions, or inferences. A segment is coded as a physical fact if it describes events or states that take place in the observed world, and as a mental fact if it depicts mental states such as personal feelings, mental processes, or psychological activities. For example, we code the consequent in (21) as a physical fact, and the consequent in (20) (repeated here as (22) for convenience) as a

judgment. (23) is considered to have a speech-act consequent, and (24) a consequent representing a mental fact.

- (22) Zhè zhǒng yào hányǒu wēiliàng de yǒudú wùzhì, yīncǐ chángqī fúyòng hěn kěnéng duì jiànkāng bùlì.  
 this kind medicine contain tiny:amount MOD poisonous substance, so long-term use very probably for health NEG:beneficial  
 ‘This kind of medicine contains a tiny amount of poisonous substance, so long-term use (of the medicine) is probably detrimental to health.’
- (23) Tā de tǐzhì hěn bùhǎo, zhè shì yīdìng de. Suǒyǐ yào kuài, wǒmen zǒu ba.  
 3SGF POSS physique very NEG:good, this COP certain PRT. so have:to quick, 1PL go PRT  
 ‘Her physique is very bad, this is true for sure. So be quick, let’s go.’
- (24) Tā zhèng nǜlì huíxiǎngzhe dāngchū língchén yī jiǎo cǎi jìn éqún shí de qíngjǐng, yúshì tā fǎngfú yòu tīngdào le éqún yīnwèi jīngguāng fāchū de jiàoshēng.  
 3SGM ASP try recall:ASP at:that:time early:morning one foot step into goose:flock time MOD situation, so 3SGM seem again hear:ASP goose:flock because:of panic produce MOD call  
 ‘He is trying to recall the situation in which he stamped into a flock of geese in the early morning, so he seems to hear the calls made by the flock of geese out of panic.’

Mental facts and judgments are located in the inner world. They are not directly observable in the external world. By contrast, physical facts exist or take place in the external world. The distinction is clear-cut. However, the boundary between mental facts and judgments is less clear, since they both involve mental processes. To make the distinction, we examined whether an opinion is formed or a decision is made during these mental activities. For example, in (24) he felt as if he heard the calls of a flock of geese, but no opinion crossed his mind. It is a description of a mental illusion. By contrast, a typical judgment is formed in (22). Sometimes, a judgment is not as typical as the one in (22). For instance, we consider the second segment of the following sentence (25) as a judgment. Some analysts would probably argue that it is a description of a fact. In our view, one needs to make judgments before making decisions. Decisions are closely connected to judgments, and thus should be analyzed as judgments.

- (25) Xià yǔ le, yīncǐ tāmen juédìng qǔxiāo yěcān huódòng.  
 fall rain ASP, so 3PL decide cancel picnic activity  
 ‘It started to rain, so they decided to cancel the picnic.’



Physical facts are considered to be less subjective than mental facts, judgments and speech acts because physical facts do not involve mental processes. Mental facts are less subjective than judgments and speech acts since more cognitive efforts are involved on the part of the SoC to form an opinion, make a conclusion, or ask a question.

### 3.3.3 *The presence of an SoC*

We distinguish between utterances with no SoC, those with an implicit SoC, and those with an explicit SoC. There is no SoC in utterances where the causal relation exists independent of volition. For example,

- (26) Nàshí wǒmen dōu zhù zài zhè yīdài, suǒyǐ wǒmen jīngcháng pèngjiàn.  
 that:time 1PL all live at this area, so 1PL often encounter  
 ‘At that time we both lived in this area, so we often encountered each other.’

In (26), the fact of living in the same area automatically leads to the fact of encountering each other often. The causal connection has nothing to do with human volition: *we did not intentionally meet each other.* There is no SoC in it. When there is an SoC, we make a distinction between implicit and explicit SoCs. If the SoC is explicitly referred to by linguistic elements in the consequent, we will consider the SoC to be explicit (as in (27)). If the SoC is not linguistically realized, we will take it to be implicit (see (28)).

- (27) Tā tīngdào mén wài yǒu dòngjīng, yúshì tā tuī kāi mén qù kàn ge jiūjìng.  
 3SGM hear door outside there:be sound, so 3SGM push open door go see CL  
 what:actually:happen  
 ‘He heard some sounds outside the door, so he pushed open the door to see what actually happened.’
- (28) Xiǎodūnzi àn kāi diànshìjī, chūxiàn de huàmiàn yánsè tè dàn, kējìàn  
 xiǎnxiàngguǎn yǐrán lǎohuà.  
 NAME press open television, appear MOD picture color particularly light, so  
 kinescope already aging  
 ‘Xiaodunzi pressed on the television, the color of the picture was particularly light, so the kinescope has already been aging.’

Importantly, when *I* or *me* occur in a clause, we need to determine whether it functions as the SoC. Take (29), for example. The *me* in the consequent clause does not function as an SoC. We can easily replace *me* with *him* or *her*, and meanwhile the SoC will still be the speaker. Hence, we label the SoC implicit for (29). The explicit counterpart is provided in (30), in which the concluder is explicitly realized by the pronoun *I*. Again, we may use the substitution test to check whether *I* is indeed the SoC. If we replace *I* with *Xiǎoli* (see (31)), the SoC is changed to Xiǎoli

simultaneously. So, *I* is considered functioning as the SoC in (30), and therefore there is an explicit SoC.

- (29) tā shuōhuà shí bìng bù kàn wǒ, *suǒyǐ* tā bìng fēi tóng wǒ shuōhuà.  
 3SGM speak time EMP NEG look 1SG, so 3SGM EMP NEG with 1SG speak  
 ‘He did not look at me when he spoke, so he was not speaking to me.’
- (30) tā shuōhuà shí bìng bù kàn wǒ, *suǒyǐ* wǒ rènwéi tā bìng fēi tóng wǒ shuōhuà.  
 3SGM speak time EMP NEG look 1SG, so 1SG think 3SGM EMP NEG with 1SG speak  
 ‘He did not look at me when he spoke, so I think he was not speaking to me.’
- (31) tā shuōhuà shí bìng bù kàn wǒ, *suǒyǐ* Xiǎoli rènwéi tā bìng fēi tóng wǒ shuōhuà.  
 3SGM speak time EMP NEG look 1SG, so NAME think 3SGM EMP NEG with 1SG speak  
 ‘He did not look at me when he spoke, so Xiaoli thinks he was not speaking to me.’

We are aware of the fact that in Mandarin sometimes an SoC is implicit due to the pervasive phenomenon of subject drop (see (32)). This is not typically possible in European languages such as Dutch, French, and German. As a pro-drop language, Mandarin allows subjects to be omitted when they are in some sense recoverable. In (32), the SoC is *he*, who intentionally stayed to take care of his father. However, the SoC is omitted via subject drop in the consequent. We consider such cases as containing an implicit SoC.

- (32) Tā dānxīn fùqīn de shēntǐ, *yīn’ér* Ø liú xià lái zhàogù fùqīn.  
 3SGM worry father POSS body, therefore (3SGM) stay down to look:after father  
 ‘He worried about his father’s health, therefore he stayed to look after his father.’

Since the SoC is the ultimate source of the subjective construction of the causal relation, a relation with an SoC is more subjective than one with no SoC. In addition, following Langacker’s (1990) suggestion that an explicit reference to the SoC objectifies the SoC by making himself/herself observable, fragments with an implicit SoC are taken to be more subjective than fragments with an explicit SoC.

### 3.3.4 *The identity of SoC*

In utterances where SoCs are present, a distinction is often made between speaker SoCs, and character SoCs. There is a speaker SoC in (33), and a character SoC in (34).

- (33) Nóngmín zuì dānxīn de shì zhèngcè yǒu biàn, *suǒyǐ* dǎng zài nóngcūn de gè xiàng jībēn zhèngcè de wěndìngxìng hé liánxùxìng hé liángzhòngyào.  
farmer most worry MOD COP policy have change, so Party at countryside MOD each CL basic policy MOD stability and continuity of:critical:importance  
‘What farmers worry about the most is change of policy, so it is critical for the Party to maintain stability and continuity of policies in the rural areas.’
- (34) Suǐrán Zhōngguó mùqián hái bù shì qiángguó, dànshì wèilái Zhōngguó yǒu gèngduō de jīhuì, gèngdà de fāzhǎn, *suǒyǐ* tāmen yuànyì liú zài Zhōngguó.  
although China currently still NEG COP strong:country, but future China have more MOD opportunity, bigger MOD development, so 3PL willing stay at China  
‘Although China is not a great power yet, it will have more opportunities and bigger development in future. So they (Chinese people) are willing to stay in China.’

In (33) it is the speaker/author who draws the conclusion that it is critical to maintain the stability and continuity of policies in the rural areas, whereas in (34) it is the character *they* who volitionally wants to stay in China. Speaker SoC is considered more subjective than character SoC. After all, the linguistic term of subjectivity refers to the property of being speaker-related.

However, in natural discourse we may encounter sentences in which the author is quoting a causal relation constructed by another person. For example, the causally related sentence within the quotation marks in (35) is constructed by a third person *he* rather than the author. Nevertheless, *he* functions as a speaker, who is responsible for the content of the utterance — or the subjectively constructed causal relation.

- (35) Tā shuō: “tōnghuò-péngzhàng réngrán cúnzài, *yīncǐ* wǔyuè huì jiā xī, shènzhì liùyuè hái huì jiā xī.”  
3SGM say: “currency:inflation still exist, so May will add interest, even June also will add interest  
‘He said: “Inflation still exists, so the interest-rate will be raised in May and even in June.”’

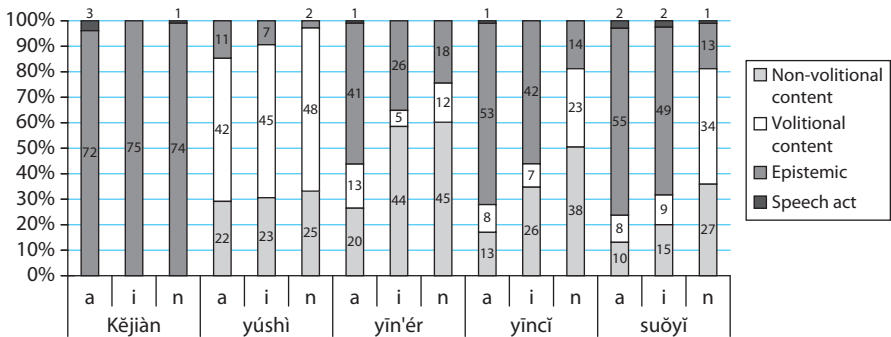
To deal with such cases, we have made a distinction within speaker SoCs: the author versus the current/quoted speaker. The author generally agrees with the quoted speaker while quoting his words, but the author is not directly responsible for the causal relation. Thus, we consider current speaker SoCs to be less subjective than author SoCs.

## 4. Results

Having spelled out the definitions of the key coding categories, we now analyze the use of each of the five connectives in terms of the four subjectivity factors: domain, modality of the consequent, and presence and identity of the SoC. For each subjectivity factor, we have conducted General Loglinear Analysis to find the model with the lowest amount of parameters (such as connective and genre) and the best fit to the observed data. Statistical details about the models for each subjectivity factor can be found in Appendices 1 to 4.

### 4.1 Domain

Figure 1 shows the distribution of the five connectives over domains. As can be seen in Figure 1, connective use in the speech-act domain is rare (only eleven cases in total). To avoid the loss of test power, we have collapsed the speech-act and the epistemic domain, the two most subjective categories.<sup>6</sup> And for this reason, we will not discuss in this paper the behavior of the five Mandarin result connectives in the speech-act domain.



**Figure 1.** The distribution of connectives over domains per genre (a = argumentative genre, i = informative genre, n = narrative genre)

As can be seen in Figure 1, the distribution over domains varies across connective tokens ( $\chi^2 [8] = 502.57, p < 0.05$ ). Moreover, the distribution of connectives over domains differs across genres as well ( $\chi^2 [24] = 66.37, p < 0.05$ ). Below we will translate these findings into domain profiles for each connective.

The data show that language users have a strong preference to use *kějiàn* in the epistemic domain across genres ( $z = 4.16, p < 0.001$ ). Fragment (36) is a typical example.

- (36) Cóng zhè xiàng wéiqī shí nián de yánjiū láikàn, zhèngmiàn qíngxù yǒuzhù yùfáng guànxīnbìng, kějiàn bǎohù xīnzàng xūyào zēngjiā zhèngmiàn qíngxù.  
 from this CL duration ten year MOD study see, positive mood have:help  
 prevent coronary:heart:disease, so protect heart need increase positive mood  
 ‘According to the ten-year long study, positive moods help to prevent the occurrence of coronary heart disease, so one needs to increase his positive moods to protect the heart.’

*Kějiàn* is typically used to connect clauses such as those in (36) to express an epistemic relation that is subjectively constructed by the speaker: the speaker is drawing a conclusion that one needs to increase his/her positive moods to protect the heart on the basis of the research finding that positive moods help to guard against coronary diseases.

The profile of *yīn’ér* is rather different from that of *kějiàn*. *Yīn’ér* exhibits a preference for the non-volitional content domain: it occurs less often in the volitional content domain ( $z = -5.31, p < 0.001$ ) and at the same it adheres to the overall main effect of having fewer occurrences in the epistemic domain ( $z = -2.72, p = 0.006$ ). This profile is rather stable across genres. Example (37) serves to illustrate that *yīn’ér* is typically used to express the non-volitional content domain in which one fact (i.e. the film-shaped solar battery uses only a small amount of silicon) leads to another (i.e. the price is low) in the observed world.

- (37) Bómóxíng tàiyángnéng diànchí zhǐ xūyào shǐyòng shǎoliàng de yuánliào guī, yīn’ér jiàgé dī.  
 film:type solar:battery only need use tiny:amount MOD material silicon,  
 as:a:result, price low  
 ‘The film-shaped solar battery uses only a small amount of silicon, as a result, its price is low.’

*Yīncǐ* and *suǒyǐ* seem to resemble each other. They exhibit a general preference for the epistemic domain (for *yīncǐ*:  $z = 3.32, p = 0.001$ ; for *suǒyǐ*:  $z = 4.60, p < 0.001$ ), and are less often used in the volitional content domain (for *yīncǐ*:  $z = -3.96, p < 0.001$ ; for *suǒyǐ*:  $z = -2.38, p = 0.02$ ). However, unlike the other three connectives, the profile of *yīncǐ* and *suǒyǐ* is not fully stable across genres: *yīncǐ* and *suǒyǐ* show an increase of their volitional use in the narrative genre (for *yīncǐ*:  $z = 2.31, p = 0.02$ ; for *suǒyǐ*:  $z = 2.92, p = 0.004$ ). The fragments (38) and (39) illustrate the typical causal domain that *yīncǐ* and *suǒyǐ* generally prefer to express.

- (38) Lìshǐ shì wánzhèng de, yīncǐ wénhuà yě yīnggāi shì wánzhèng de.  
 history COP continuative PRT, so culture also should COP continuous PRT  
 ‘History is continuative, so culture should also be continuous.’

- (39) Tā shǒu shàng dàile hǎo jǐ gè jièzhi, suǒyǐ wǒ yǐwéi tā jiéhūn le.  
 3SGM hand on wear:ASP quite several CL ring, so 1SG think 3SGM marry ASP  
 ‘He wore quite a few rings on the hand, so I thought he was married.’

Both (38) and (39) involve an SoC’s epistemic reasoning: the speaker makes a judgment (in (38): *culture should also be continuous*; in (39): *I thought he was married*) on the basis of the propositional fact expressed in the antecedent (in (38): *history is continuative*; in (39): *he wore quite a few rings on the hand*).

The final connective, *yúshì*, adheres to the overall main effects of domains: it has a clear preference for use in the volitional content domain ( $z = 2.60, p = 0.009$ ), and hardly occurs in the epistemic domain ( $z = -2.72, p = 0.006$ ). Just like *kějiàn* and *yīn’ér*, *yúshì* shows a stable pattern across genres. The typical use of *yúshì* is exemplified in (40): the fact that he heard some knocking sounds on the seabed causes his intentional act of swimming there to see what happened.

- (40) Tā tīngdào le hǎichuāng shàng qiāodǎ de shēngyīn, yúshì tā yóu guòqù xiāng kàn ge jiūjìng.  
 3SGM hear:ASP seabed on knock MOD sound, so 3SGM swim over want see CL what:actually:happen  
 ‘He heard some sounds produced by knocks on the seabed, so he swam over to see what happened.’

## 4.2 Modality

Figure 2 shows the distribution of the five connectives over the different modalities. Similar to what we have found for the speech-act domain, connective use in the speech-act modality is rare (only eleven cases in total). This is why we collapse the speech-act modality with another highly subjective modality: judgment.

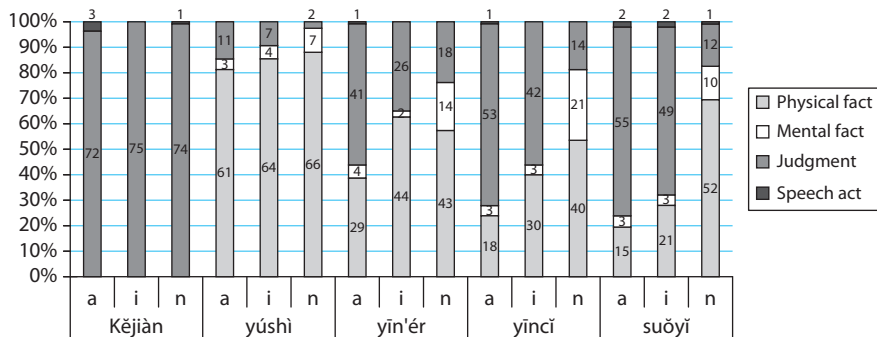


Figure 2. The distribution of connectives over modality per genre (a = argumentative genre, i = informative genre, n = narrative genre)

As can be seen in Figure 2, these connectives present different patterns of distribution over modalities ( $\chi^2 [8] = 429.61, p < 0.001$ ). Furthermore, the distribution patterns vary by genre ( $\chi^2 [24] = 61.09, p < 0.001$ ). Below we will translate these findings into modality profiles for each connective.

The modality profile of *kějiàn* appears to be salient and stable. Across genres *kějiàn* has a great chance to co-occur with judgments ( $z = 2.87, p = 0.004$ ). While *yúshì*, *yīncǐ*, and *suǒyǐ* show a decrease of judgments in narratives (cf. the interaction effect of genre and modality:  $z = -5.60, p < 0.001$ ), *kějiàn* maintains its strong preference for judgments even in the narrative genre ( $z = 3.87, p < 0.001$ ). Fragment (36), given in the previous section, also exemplifies the typical usage of *kějiàn* in this respect. The consequent clause *one needs to increase his positive moods to protect the heart* is clearly opinion-oriented, so it is considered a judgment.

*Yīncǐ* and *suǒyǐ* are very similar to each other with respect to modality. In the argumentative and the informative genres, they both adhere to the main effects of modality: they prefer to co-occur with judgments ( $z = 3.40, p = 0.001$ ), and appear less often with mental facts ( $z = -3.15, p = 0.002$ ). In the narrative genre, however, *yīncǐ* and *suǒyǐ* show an increase of physical facts and a decrease of judgments (cf. the interaction between genre and modality:  $z = -5.60, p < 0.001$ ).<sup>7</sup> To illustrate the typical modality that *yīncǐ* and *suǒyǐ* co-occur with, we may look at fragments (38) and (39) again. The consequent clause in (38) *culture should also be continuous* is conclusive in nature, and the consequent in (39) *I thought he was married* is a personal opinion. These are both typical examples of a judgment, the type of modality *yīncǐ* and *suǒyǐ* prefer to co-occur with.

*Yīn'ér* and *yúshì* are quite similar to one another in terms of modality. They exhibit a preference for physical facts: they appear less often with judgments (for *yīn'ér*:  $z = -4.13, p < 0.001$ ; for *yúshì*:  $z = -6.53, p < 0.001$ ) and mental facts (i.e. they adhere to the main effect of modality:  $z = -3.15, p = 0.002$ ). Their preference for physical facts is rather constant across genres (see Endnote 6). But *yúshì* is also sensitive to the genre: it has less chance to co-occur with judgments in narratives than in the other two genres (i.e. it adheres to the interaction between genre and modality:  $z = -5.60, p < 0.001$ ). The two connectives' preference for physical facts is illustrated in fragments (37) and (40). The consequent clause in (37) *the price is low* is an observable fact, rather than an SoC's conclusion. Similarly, the consequent in (40) *he swam over to see what happened* is a physical act that can be observed in the outside world.

#### 4.3 The presence of SoC

Figure 3 shows the distribution of the five connectives over the kinds of SoC they can co-occur with. The connectives show different distribution profiles ( $\chi^2$

[8] = 383.60,  $p < 0.001$ ). Moreover, the distribution of connectives over the kinds of SoC varies by genre ( $\chi^2 [24] = 44.73$ ,  $p = 0.006$ ). Below we will interpret these findings in detail.

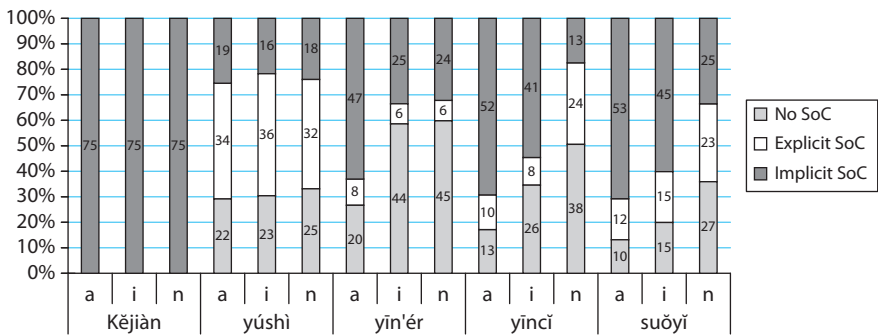


Figure 3. The distribution of connectives over presence of SoC per genre (a = argumentative genre, i = informative genre, n = narrative genre)

*Kějiàn* shows a strong preference for implicit SoCs ( $z = 3.69$ ,  $p < 0.001$ ). This profile appears to be stable across genres. The preference for an implicit SoC is illustrated clearly in fragment (36), where a conclusion is drawn by an SoC, *one needs to increase his positive moods to protect the heart*, and meanwhile the conclusion maker is not explicitly referred to by any linguistic element. *Suǒyǐ* also shows a preference for implicit SoCs across genres ( $z = 3.29$ ,  $p = 0.001$ ) (see (33)).

As to *yīncǐ*, in the argumentative and informative genres it exhibits a preference for implicit SoCs ( $z = 1.98$ ,  $p = 0.048$ ) and a dislike of explicit SoCs ( $z = -3.32$ ,  $p = 0.001$ ), whereas in the narrative genre it shows an increase of explicit SoCs ( $z = 2.52$ ,  $p = 0.01$ ) and a decrease of implicit SoCs ( $z = -2.68$ ,  $p = 0.007$ ). Fragment (38) exemplifies the kind of SoC *yīncǐ* generally co-occurs with. The person who makes the judgment *culture should also be continuous* is inferable from the context, but it is implicit in the consequent.

The connective *yīn'ér* is often used in causal relations with no SoC: speakers do not prefer to use *yīn'ér* together with explicit SoCs ( $z = -4.76$ ,  $p < 0.001$ ), and this lower number of explicit SoCs does not result in a higher number in the category of implicit SoC ( $z = -0.50$ ,  $p = 0.62$ ).<sup>8</sup> The profile of *yīn'ér* is rather constant across genres. Fragment (37) shows the typical usage of *yīn'ér*: the causal connection involved does not depend on intentionality. The result segment expresses the natural consequence of the economic law: when the cost of production is low, the price for the product is low.

Compared to the connectives *yīn'ér*, *yīncǐ* and *kějiàn*, the connective *yúshì* co-occurs more often with explicit SoCs (*yīn'ér*: fewer explicit SoCs,  $z = -4.76$ ,  $p < 0.001$ ; *yīncǐ*: fewer explicit SoCs,  $z = -3.32$ ,  $p = 0.001$ ; *kějiàn*: no instances of



explicit SoCs were observed). Fragment (40) exemplifies this prototypical use of *yúshì*, in which the source of the intentional act of *swimming there* is explicitly realized by the pronoun *he*.

In the analysis, we have observed 79 instances of implicit SoCs in volitional content relations. As introduced earlier (see (32) in Section 3.3.3), this type of implicit SoC occurs as a result of subject drop. It occupies the subject position of the result segment and functions as the agent of an intentional act. It is different from the type of implicit SoC commonly found in epistemic relations, which is outside the syntactic structure of the sentence. On the basis of their semantic roles, we name the former type of implicit SoC ‘implicit agent SoC’, and the latter type ‘implicit concluder SoC’. Table 5 shows the distribution of the connectives over the two types of implicit SoC.

**Table 5.** The distribution of the connectives over two types of implicit SoC

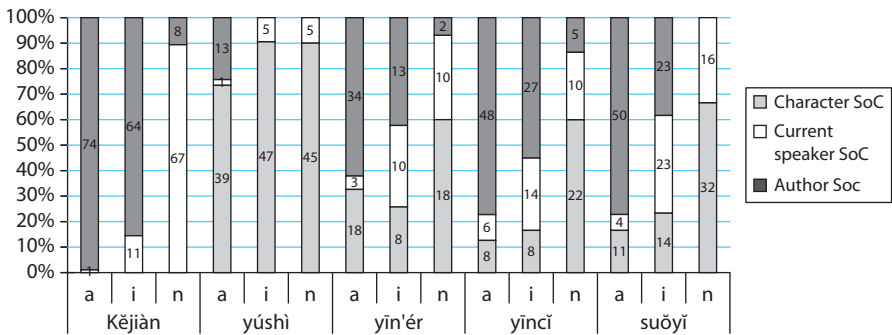
Connectives	Types of implicit SoC		Total
	Implicit agent SoC	Implicit concluder SoC	
<i>kějiàn</i>	0 (.0%)	225 (100.0%)	225 (100.0%)
<i>yúshì</i>	41 (77.4%)	12 (22.6%)	53 (100.0%)
<i>yīncǐ</i>	3 (2.8%)	103 (97.2%)	106 (100.0%)
<i>yīn'ér</i>	18 (18.8%)	78 (81.3%)	96 (100.0%)
<i>suǒyǐ</i>	17 (13.8%)	106 (86.2%)	123 (100.0%)
Total	79 (13.1%)	524 (86.9%)	603 (100.0%)

In contrast to the other four connectives, *yúshì* co-occurs more often with the implicit agent SoC resulting from subject drop. An example is given in (41) to illustrate the special usage of *yúshì*. We will discuss the two types of implicit SoC in depth in Section 5.

- (41) Tā gǎnjuédào nánbiān fùzá, *yúshì* Ø láidào Běipíng.  
 3SGM feel south complicated, so (3SGM) come:to Beiping  
 ‘He felt that the situation in the South was complicated, so he came to Beiping (the old name of Beijing).’

#### 4.4 The identity of the SoC

Figure 4 shows the distribution of the five connectives over the types of SoC they can co-occur with. The connectives’ distributional patterns differ ( $\chi^2 [8] = 374.80$ ,  $p < 0.001$ ), and vary with genre ( $\chi^2 [24] = 82.36$ ,  $p < 0.001$ ). Next, we will interpret these findings in more detail.



**Figure 4.** The distribution of connectives over identity of SoC per genre (a = argumentative genre, i = informative genre, n = narrative genre)

*Kějiàn* exhibits an overall preference for author SoCs ( $z=2.50$ ,  $p=0.01$ ), except that in the narrative genre there is an increase of current speaker SoCs ( $z=4.06$ ,  $p<0.001$ ). Fragment (36) illustrates the typical use of *kějiàn* with an author SoC: in (36) the author is the ultimate source of the judgment *one needs to increase his positive moods to protect the heart*. The typical use of *kějiàn* in the narrative genre is shown in the following sentence (42): it is used to express causal relations constructed by a first person character in the novel rather than the author. It is not the writer of the novel who draws the conclusion about the picture, but the first person character “I”, which we code as a case of current speaker SoC.

- (42) Wǒ zhīdào Zhāng Zéduān bùhuì fàn zhè zhǒng cuòwù, *kějiàn* zhè bùshì Zhāng Zéduān suǒ huà.  
 1SG KNOW NAME NEG:will make this kind error, so this (picture) NEG:COP NAME PRT paint  
 ‘I know that Zhang Zeduan won’t make this type of error, so the picture is not painted by Zhang Zeduan.’

In contrast with *kějiàn*, the connective *yúshi* prefers character SoCs: it co-occurs less often with current speaker ( $z=-4.31$ ,  $p<0.001$ ) and author SoCs ( $z=-3.88$ ,  $p<0.001$ ). Its preference for character SoCs is quite stable across genres. In general, there is an increase of character SoCs in the narrative genre (cf. the interaction between genre and identity of SoC: fewer cases of current speaker SoC,  $z=-2.27$ ,  $p=0.02$ ; fewer cases of author SoC,  $z=-4.19$ ,  $p<0.001$ ), but *yúshi* does not show such an increase of character SoCs ( $z=-2.24$ ,  $p=0.03$ ). The preference for character SoCs is illustrated in fragments (40) and (41). Although the character SoC *he* is implicit in (41), it can be recovered from the context: it is the character *he* who did the intentional act of coming to the city of Beijing.

The other three connectives (i.e. *suǒyǐ*, *yīncǐ*, and *yīn'ér*) also appear more often with author SoCs (cf. the main effect of the identity of SoC:  $z=2.99$ ,  $p=0.003$ ),

except in the narrative genre, where the three connectives show an increase of character SoCs (cf. the afore-mentioned interaction between genre and identity of SoC). Fragments (38) and (39) have exemplified the typical use of *yīncǐ* and *suǒyǐ*, respectively. Furthermore, we have already seen that there is often no SoC in the *yīn'ér* fragments. However, *yīn'ér* can also express causal relations that are subjectively constructed by an SoC. In such cases, the SoC is most often the author, as shown in the following fragment (43). Here it is the author who draws the conclusion that *changes must be made*.

- (43) Huángjīnzhōu de fùmiàn hòuguǒ yě rìyì xiǎnxiàn chūlái, *yīn'ér*  
 huángjīnzhōu de gǎibiàn yě jiù shìzài-bìxíng.  
 golden:week MOD negative consequence also gradually appear out, *therefore*  
 golden:week MOD change also then must:enforce  
 ‘The negative effects of Golden Week holidays have revealed themselves  
 gradually, *therefore* changes must be made.’

## 5. Discussion

The current study has produced several interesting findings and topics for discussion. Below we will reflect on the subjectivity profiles of the connectives (Section 5.1), as well as their genre-sensitivity (5.2). Furthermore, we will discuss the distributional patterns and frequencies of use in terms of speaker economy (5.3), and make some remarks on the analytical model we used in this paper (5.4).

### 5.1 Subjectivity profiles identified

Table 6 summarizes the subjectivity profiles of the five connectives in question. Three robust subjectivity profiles across genres can be identified for our data: epistemic *kějiàn*, volitional *yúshì*, and non-volitional *yīn'ér*. *Kějiàn*, a highly subjective causal connective, specializes in expressing epistemic relations and prefers to occur in highly subjective contexts involving judgments and implicit speaker SoCs. By contrast, *yīn'ér* has a rather objective profile. It prefers to occur within the non-volitional content domain to describe causal connections between physical facts in observable reality. *Yúshì* also exhibits a clear preference for the objective content domain and for the objective contexts involving physical facts. It should be noted that *yúshì* can occasionally express epistemic relations, which counters Lu's (2000) claim that *yúshì* does not introduce inferences and conclusions at all. Unlike *yīn'ér*, *yúshì* prefers to express content relations in which human intentions play a role: it is often used to express volitional content relations in which an SoC can be identified as the source of the causal process. Moreover, the SoC is often

**Table 6.** Overview of the subjectivity profiles per connective

Connective	Domain	Modality	Presence of SoC	Identity of SoC
<i>kějiàn</i>	+epistemic	+judgment	+implicit	+author * (current speaker)
<i>yīn'ér</i>	+non-vol. content	+physical fact	+no SoC	+author * (character)
<i>yúshì</i>	+vol. content	+physical fact	+explicit	+character
<i>yīncǐ</i>	+epistemic * (vol. content)	+judgment * (physical fact)	+implicit * (explicit)	+author * (character)
<i>suǒyǐ</i>	+epistemic * (vol. content)	+judgment * (physical fact)	+implicit	+author * (character)

Note: \* indicates that effects are genre-sensitive; within brackets we mention the category that increases in the narrative genre.

a character explicitly realized in segments containing *yúshì*. Thus, *yúshì* may be considered to be more subjective than *yīn'ér*, and less subjective than *kějiàn*.

These results are basically in line with observations in previous publications and our hypotheses formulated in Section 3.1. In particular, the objective profile of *yúshì* correlates with its closeness to a descriptive temporal use; the subjective profile of *kějiàn* conforms to the epistemic nature of its root meaning, ‘can see’; and *yīn'ér* expresses objective causal relations — descriptive causality — precisely as Xing (2001) predicts.

Generally speaking, *yīncǐ* and *suǒyǐ* are more subjective than *yīn'ér* and *yúshì*. Speakers prefer to use *yīncǐ* and *suǒyǐ* in the epistemic domain, linking arguments with conclusions, where the conclusion maker (i.e. the SoC) is often the author and is often not linguistically realized (i.e. implicit). *Yīncǐ* and *suǒyǐ* are also used in the objective content domain; it is on this basis we conclude that they are less subjective than *kějiàn*, which specializes in expressing epistemic relations and never occurs in the objective content domain.

Except for *yúshì*, each of the connectives under discussion has one or two instances of speech-act relations. The number of observed speech-act relations is rather low in our samples: only 11 cases in total. This is because the present study focuses on the written discourse. Previous studies have already shown that speech-act relations rarely occur in the written discourse. For example, Sanders & Spooren (submitted) find that speech-act relations occur relatively more often in chat (27.8%) and spoken language (10.1%) than in written language (5.0%). On the basis of the present study, we cannot draw a conclusion on the behavior of the five Mandarin result connectives in the speech-act domain. Spoken discourse should be investigated for that purpose.

## 5.2 Genre-sensitivity: Semantics vs. pragmatics

Given the fact that *kějiàn* and *yīn'ér* show stable usage patterns across genres in terms of domain, modality, and presence of SoC, why does it seem that their usage varies with genre when the fourth variable, identity of SoC, is concerned? As shown in Table 6, in the narrative genre there is an increase of current speaker SoCs in the *kějiàn* fragments and an increase of character SoCs in the *yīn'ér* fragments. Do these variations affect our conclusion about the subjectivity profiles of *kějiàn* and *yīn'ér*? The answer is negative. In fact, in spite of the observed variations in narratives, *kějiàn* and *yīn'ér* can still be argued to have stable subjectivity profiles across genres. For *kějiàn*, the variation from author SoC to current speaker SoC does not affect the nature of its subjectivity profile because both of these categories are highly subjective. For *yīn'ér*, the increase in character SoCs does not affect our conclusion that *yīn'ér* has a rather objective profile across genres, because the fragments with an SoC make up only 52% of the *yīn'ér* cases. This means that almost half of the fragments is of the most objective type: without any SoCs.

While *kějiàn*, *yúshì*, and *yīn'ér* have robust profiles across genres, the subjectivity profile of *yīncǐ* and *suǒyǐ* are genre-sensitive. In the argumentative and informative genres, they are most often used in the epistemic domain, but exhibit less subjective patterns in the narrative genre: in novels they co-occur more often with physical facts, volitional content relations, and character SoCs. The observed preference of *suǒyǐ* and *yīncǐ* for the epistemic domain seems to diverge from the findings in Xing (2001), who finds that these connectives typically mark descriptive causality (i.e. in the content domain). This apparent disagreement disappears when we take the genre-sensitivity of *suǒyǐ* and *yīncǐ* into account: in narratives, the only genre Xing (2001) focused on, these connectives are indeed more often used to express objective relations in the content domains.

The observed genre-sensitivity conforms to previous findings in the literature that the distribution of connectives over causal categories can vary across contexts. For example, the subjectively oriented French *car*, German *denn*, and Dutch *want* have a strong preference for subjective causal relations across text types, whereas the profiles of objectively oriented connectives such as French *parce que*, German *weil*, and Dutch *omdat* are inconsistent in newspaper corpora (for an overview, see Stukker & Sanders 2012).

The observed genre-sensitivity makes us wonder whether the subjectivity profiles are really part of the inherent semantic characteristics of the connectives themselves or whether they are (partially) determined by the pragmatics of the context. Neither option is sufficient to account for the entire set of data. Therefore, we propose a third option. If we assume that subjectivity is a semantic feature expressed by all causal connectives, we may propose that for *kějiàn*, *yúshì*, and

*yīn'ér*, the feature of subjectivity is specified in the lexicon, resulting in a subjectivity profile that is constant across genres. For *yīncǐ* and *suǒyǐ*, however, the subjectivity feature is semantically underspecified. Consequently, they depend on the pragmatics of the context for the exact value of subjectivity. For example, in novels authors prefer to describe observable situations and relate characters' physical activities to their motivations. Hence in novels *yīncǐ* and *suǒyǐ* show an increase of volitional use.

### 5.3 Speaker economy

The current study reveals that Mandarin has a specific result connective, *kějiàn*, which is restricted to expressing highly subjective epistemic relations. In comparison with the other four connectives, *kějiàn* is the least frequently used in the Mandarin written discourse (see Table 1). The other four connectives are quite general across causal domains: they can be used to mark different domains of causality, although they all have their own prototypical usage. Among them, the connectives *yīncǐ* and *suǒyǐ* are the most general ones because they are able to serve not only as markers for subjective epistemic relations (in the argumentative and informative genres) but also as markers for objective volitional content relations (in the narrative genre). These two connectives are also the most frequently used causal connectives in the Mandarin written discourse. Taken together, we can see that general causal connectives have a higher frequency of usage than specific ones do. The findings seem to suggest that speakers prefer to choose the most general connectives to express a relation, which is in line with a speaker economy account (Knott & Sanders 1998).

### 5.4 Remarks on the analytical model

The four variables in the analytical model have provided consistent information regarding the degrees of subjectivity that each result connective expresses (see Table 6). For example, whenever a connective is used in the non-volitional content domain, there is no SoC involved and the consequent always involves facts. This begs the questions as to why four variables instead of one would be needed and whether there is redundancy in positing multiple variables in characterizing subjectivity. We would like to stress that the integrative system in our analytic model is needed, as all four of the variables are needed to quantify degrees of subjectivity from different angles. Therefore, it is expected that the four variables can serve to complement each other while depicting a connective's degree of subjectivity. For example, in case the domain factor fails to distinguish one connective from another, other variables may provide complementary information to set them apart.

As reported in Section 4.3, two types of implicit SoC have been observed in Mandarin Chinese. This is a notable difference between Mandarin Chinese, a pro-drop language, and the Germanic languages, on whose basis studies on the subjectivity of connectives have been conducted. In Germanic languages such as Dutch and German in which subject drop is impossible, implicit SoCs are rarely found in the volitional content causal domain: the agent SoCs (in the volitional content domain) are always linguistically realized, except in passive structures such as *I/He did a good job, that's why I/he was promoted*, in which the agent of promotion is avoided. That is to say, in Germanic languages, implicit SoCs always occur in the epistemic domain, functioning as the source of a conclusion (see also Sanders, Sanders & Sweetser 2009, 2012). For this reason, previous studies based on non-pro-drop languages have not made a distinction between the implicit agent SoC (resulting from subject drop) and the implicit concluder SoC. The two types of implicit SoC can be illustrated with fragment (41) (repeated here as (44)) and fragment (43) (repeated here as (45)), respectively.

- (44) Tā gǎnjuédào nánbiān fùzá, yúshì Ø láidào Běipíng.  
 3SGM feel south complicated, so (3SGM) come:to Beijing  
 ‘He felt that the situation in the South was complicated, so he came to Beijing (the old name of Beijing).’
- (45) Huángjīnzhōu de fùmiàn hòuguǒ yě rìyì xiǎnxiàn chūlái, yīn’ér  
 huángjīnzhōu de gǎibiàn yě jiù shìzài-bìxíng.  
 golden:week MOD negative consequence also gradually appear out, therefore  
 golden:week MOD change also then must:enforce  
 ‘The negative effects of Golden Week holidays have revealed themselves gradually, therefore changes must be made.’

We think the two types of SoC can be considered as expressing different degrees of subjectivity. For example, the agent SoC in (44) *he* can be recovered at the syntactic level (e.g. by feature-checking), that is, *he* is actually present in the deep structure of the sentence. By comparison, the concluder SoC in (45), the speaker, cannot be recovered through syntactic mechanisms. It is more implicit in the sense that it is completely outside the structure of the sentence. Therefore, in order to get a clear picture of the subjectivity profiles of connectives in subject drop languages such as Mandarin, we should distinguish the two types of implicit SoC. A connective preferring implicit agent SoCs should be considered less subjective than a connective preferring implicit concluder SoCs.

## 6. Conclusion

In this paper, we have performed a corpus-based analysis on five Mandarin result connectives: *kějiàn*, *suǒyǐ*, *yīncǐ*, *yīn'ér*, and *yúshì*, in terms of four subjectivity indicators: modality, domain, and the presence and identity of an SoC. This analysis has uncovered distinct subjectivity profiles of the five result connectives. The findings allow us to arrange the degrees of subjectivity of the five connectives in a hierarchical order (from more to less subjective): *kějiàn* > *suǒyǐ*/*yīncǐ* > *yúshì* > *yīn'ér*. Genre-sensitivity has been observed for the connectives *yīncǐ* and *suǒyǐ*. Therefore, we propose that for connectives with a robust subjectivity profile (*kějiàn*, *yúshì*, and *yīn'ér*), the feature of subjectivity is specified in the lexicon, whereas for genre-sensitive connectives (*yīncǐ* and *suǒyǐ*), the subjectivity feature is semantically underspecified, leaving room for the pragmatics of the context to determine their exact value of subjectivity.

Furthermore, we have confirmed existing claims (e.g. the ones in Xing 2001) about the ways of result marking in Mandarin discourse with evidence from a systematic corpus-based study. Differences between Xing's and our findings could be ascribed to the genres that were studied: Xing only studied narrative texts, while we included argumentative and informative texts as well. The ostensible inconsistency between studies urges us to take the effect of genre into account as we discuss the meaning and usage of causal connectives and other discourse phenomena.

The present study has shown the relevance of the notion of subjectivity to the description of causal connectives in Mandarin Chinese — a language that is typologically very different from the Germanic and Romance languages that were so far the focus of studies on subjectivity and connectives. The results on Mandarin connectives suggest that the theory of subjectivity can be generalized across various languages. At the same time, we have seen differences that are — at least to a certain extent — related to the systematic differences between the languages. For instance, two types of implicit SoC (i.e. implicit agent SoC and implicit concluder SoC) come into our view in Mandarin Chinese, a pro-drop language, of which the first one is lacking in the major European languages. We hope to have shown how a theory-driven and corpus-based approach to the categorization of causal connectives produces interesting results for further cross-linguistic comparison.

## Notes

\* This research was enabled by the Dutch Ministry for Education, Culture, and Science, through Huygens Grant, awarded to Fang Li; and by The Netherlands Organization for Scientific research, through NWO-Vici-grant 277-70-003, awarded to Ted Sanders. We thank Hongyin Tao



and two anonymous reviewers for their valuable comments on earlier versions of this paper. We also thank Yueru Ni and Yuning Sun, for the tremendous efforts they made in analyzing the 1125 corpus fragments.

1. According to Chao (1968), *yīn'ér*, *yīncǐ*, and *suǒyǐ* are adverbial conjunctions because they can be put either in front of or immediately after the subject (which is the position for adverbs). However, many other authors disagree with this definition. For example, Lü (1979) and Zhang (1996) suggest that a lexeme is a conjunction if it can appear in both the pre- and post-subject position, and an adverb if it can only appear in the post-subject position.
2. But see Keller (1995) for recent developments in the German connective system.
3. In (14), there is an animate participant, Bob. However, Bob's intentionality is not the ultimate source of the causal event of getting hurt in the leg as Bob did not intentionally get himself hurt in the leg. The causal relation in (14) exists independent of Bob's intentionality. Neither does it depend on the speaker's intentionality. That is, there is no SoC at all in (14).
4. *Guānxi cí* or *liánjiē cí* are defined as words or expressions that have the function of connecting clauses and simultaneously marking the coherence relation between the connected clauses (Xing 1996:321; Xing 2001:26).
5. The LCMC Corpus is a balanced corpus representing modern Mandarin Chinese from Mainland China. It consists of five hundred 2,000-word samples taken from a wide range of text-types, including the informative, the argumentative, and the narrative genres.
6. The Loglinear analysis requires that there should be no expected counts less than 1 in the contingency table (Field 2009:710). However, in our study there are ten cells with expected counts of 0.4 due to the rare occurrence of the speech-act category. This is why we decide to collapse speech-act with another highly subjective category.
7. As clearly shown in Figure 2, *suǒyǐ* has (at least) twice as many physical facts in novels than in the informative and the argumentative genres. *Yīncǐ* does not significantly differ from *suǒyǐ* in this respect. The statistics in Appendix 2B only show this effect indirectly, because we take *suǒyǐ* as the reference category. *Suǒyǐ*'s (and *yīncǐ*'s) relatively high frequency in fragments with physical facts in narratives can be derived from the significantly lower frequency of this category in the use of *yīn'ér* and *yúshì* in narratives (for *yīn'ér*:  $z = -2.96$ ,  $p = 0.003$ ; for *yúshì*:  $z = -2.78$ ,  $p = 0.005$ ).
8. Statistically speaking, *yīn'ér*'s preference for "no SoC" can only be shown indirectly, because "no SoC" serves as the reference category in our statistical analyses.

## References

- Chao, Y. R. 1968. *A Grammar of Spoken Chinese*. Berkeley: University of California Press.
- Chinese Academy of Social Sciences (eds). 2002. *Contemporary Chinese Dictionary (Chinese-English Edition)*. Beijing: Foreign Language Teaching and Research Press.
- Dancygier, B. 2009. "Causes and Consequences: Evidence from Polish, English, and Dutch." In Sanders & Sweetser, 91–118.

- Degand, L., and H. Pander Maat. 2003. "A contrastive study of Dutch and French Causal Connectives on the Speaker Involvement Scale." In *Usage-Based Approaches to Dutch*, ed. by A. Verhagen, and J. van de Weijer, 175–199. Utrecht: LOT.
- De Smet, H., and J. C. Verstraete. 2006. "Coming to terms with subjectivity." *Cognitive Linguistics* 17(3): 365–392.
- Ducrot, O. 1980. "Essai D'application: MAIS — les allusions à l'énonciation — délocutifs, Performatifs, Discours Indirect." In *Le Langage en Context: Etudes Philosophiques et Linguistiques de Pragmatique*, ed. by P. Herman, 487–575. Amsterdam: John Benjamins.
- Evers-Vermeul, J., L. Degand, B. Fagard, and L. Mortier. 2011. "Historical and comparative perspectives on subjectification: A corpus-based analysis of Dutch and French causal connectives." *Linguistics* 49(2): 445–478.
- Field, A. 2009. *Discovering Statistics Using SPSS*, 3rd edition. London: Sage.
- Fludernik, M. 1993. *The Fictions of Language and the Languages of Fiction: The Linguistic Representation of Speech and Consciousness*. London: Routledge.
- Frohning, D. 2007. *Kausalmarker Zwischen Pragmatik und Kognition: Korpusbasierte Analysen zur Variation im Deutschen*. Tübingen: Niemeyer.
- Halliday, M. A. K., and R. Hasan. 1976. *Cohesion in English*. London: Longman.
- Keller, R. 1995. "The Epistemic 'Weil'" In *Subjectivity and Subjectivisation: Linguistic Perspectives*, ed. by Stein, D., and S. Wright, 16–30. Cambridge: Cambridge University Press.
- Knott, A., and T. Sanders. 1998. "The classification of coherence relations and their linguistic markers: An exploration of two languages." *Journal of Pragmatics* 30: 135–175.
- Lakoff, G. 1987. *Women, Fire and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press.
- Lakoff, G., and M. Johnson. 1999. *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*. New York: Basic Books.
- Langacker, R. 1985. "Observations and Speculations on Subjectivity." In *Iconicity in Syntax*, ed. by J. Haiman, 109–150. Amsterdam: Benjamins.
- Langacker, R. 1990. "Subjectification." *Cognitive Linguistics* 1: 5–38.
- Li, J. 2011. "Lun youyu yu yinwei de chayi [On the differences between 'youyu' and 'yinwei']." *Shijie Hanyu Jiaoxue [Chinese Teaching in the World]* 4: 490–496.
- Li, J., and Y. Liu. 2004. "Youyu yu jiran de zhuguanxing chayi [The differences of 'youyu' and 'jiran' in subjectivity]." *Zhongguo Yuwen* 299: 123–128.
- Liu, Y., and X. 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.
- Lu, Q. 2000. "Yushi yu shili chengjie ['Yushi' and logic connection]." *Journal of Yangzhou University (The Humanities and Social Science Edition)* 6: 40–44.
- Lü. S. 1979. *Hanyu Yufa Fenxi Wenti [An Analysis of Chinese Grammar]*. Beijing: Commercial Publishing House.
- Lü. S. 1999. *Xiandai Hanyu Babaici [Eight Hundred Words of Chinese]*. Beijing: Commercial Publishing House.
- Lyons, J. 1977. *Semantics*. Cambridge: Cambridge University Press.
- Lyons, J. 1982a. "Subjecthood and Subjectivity." In *Subjecthood and Subjectivity: The Status of the Subject in Linguistic Theory*, ed. by M. Yaguello, 9–17. Paris: Ophrys.
- Lyons, J. 1982b. "Deixis and Subjectivity: Loquor Ergo Sum." In *Speech, Place and Action: Studies in Deixis and Related Topics*, ed. by R. J. Jarvella, and W. Klein, 101–124. Chichester / New York: John Wiley & Sons.

- Lyons, J. 1995. *Linguistic Semantics: An Introduction*. Cambridge: Cambridge University Press.
- Mann, W. C., and S. A. Thompson. 1986. "Relational propositions in discourse." *Discourse Processes* 9: 57–90.
- Mann, W. C., and S. A. Thompson. 1988. "Rhetorical structure theory: Toward a functional theory of text organization." *Text* 8(3): 243–281.
- Martin, J. R. 1992. *English texts: System and Structure*. Amsterdam: John Benjamins.
- McEnery, T., and R. Xiao. 2004. *The Lancaster Corpus of Mandarin Chinese*. Lancaster: UCREL.
- Moeschler, J. 1989. "Pragmatic connectives, argumentative coherence and relevance." *Argumentation* 3: 321–339.
- Murray, J. D. 1997. "Connectives and narrative text: The role of continuity." *Memory & Cognition* 25(2): 227–236.
- Pander Maat, H., and L. Degand. 2001. "Scaling causal relations and connectives in terms of Speaker Involvement." *Cognitive Linguistics* 12: 211–245.
- Pander Maat, H., and T. Sanders. 2000. "Domains of Use or Subjectivity? The Distribution of Three Dutch Causal Connectives Explained." In *Cause, Condition, Concession and Contrast: Cognitive and Discourse Perspectives*, ed. by E. Couper-Kuhlen, and B. Kortman, 57–81. Berlin: Mouton de Gruyter.
- Pander Maat, H., and T. Sanders. 2001. "Subjectivity in causal connectives: An empirical study of language in use." *Cognitive Linguistics* 12: 247–273.
- Pit, M. 2003. *How to Express Yourself with a Causal Connective. Subjectivity and Causal Connectives in Dutch, German and French*. Ph.D. Dissertation, Utrecht University.
- Sanders, J. 2010. "Intertwined voices: Journalists' modes of representing source information in journalistic subgenres." *English Text Construction* 3(2): 226–249.
- Sanders, J., T. Sanders, and 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): 169–190.
- Sanders, T. 1997. "Semantic and pragmatic sources of coherence: On the categorization of coherence relations in context." *Discourse Processes* 24: 119–147.
- Sanders, T., J. Sanders, and E. Sweetser. 2009. "Causality, Cognition and Communication: A Mental Space Analysis of Subjectivity in Causal Connectives." In Sanders & Sweetser, 19–59.
- Sanders, T., and W. Spooren. 2009. "Causal Categories in Discourse — Converging Evidence from Language Use." In Sanders & Sweetser, 205–246.
- Sanders, T., and W. Spooren. Submitted. "Causality and subjectivity in discourse: The meaning and use of causal connectives in spontaneous conversation, chat interaction and written text".
- Sanders, T., W. Spooren, and L. Noordman. 1992. "Toward a taxonomy of coherence relations." *Discourse Processes* 15(1): 1–35.
- Sanders, T., and E. Sweetser (eds). 2009. *Causal Categories in Discourse and Cognition*. Berlin: Mouton de Gruyter.
- Shen, J. 2001. "Yuyan de zhuguanxing he zhuguanhua [A survey of studies on subjectivity and subjectivisation]." *Foreign Language Teaching and Research* 33(4): 268–275.
- Shen, J. 2008. "San ge shijie [Physical world, mental world, and linguistic world]." *Foreign Language Teaching and Research* 40(6): 403–408.
- Spooren, W., Sanders, T., Huiskes, M., and L. Degand. 2010. "Subjectivity and Causality: A Corpus Study of Spoken Language." In *Empirical and Experimental Methods in Cognitive/Functional Research*, ed. by S. Rice, and J. Newman, 241–255. Chicago: CSLI publications.

- Stukker, N., T. Sanders, and A. Verhagen. 2008. "Causality in verbs and in discourse connectives: Converging evidence of cross-level parallels in Dutch linguistic categorization." *Journal of Pragmatics* 40: 1296–1322.
- Stukker, N. M. & T.J.M. Sanders. 2010. "Diverging frequency effects in causal connectives across discourse contexts. A usage-based interpretation." *LAUD symposium Cognitive Sociolinguistics: Language variation in its structural, conceptual and cultural dimensions*. Essen: LAUD Linguistic Agency/Linse Linguistik-server.
- Stukker, N., and T. Sanders. 2012. "Subjectivity and prototype structure in causal connectives: A cross-linguistic perspective." *Journal of Pragmatics* 44(2): 169–190.
- Sweetser, E. 1990. *From Etymology to Pragmatics: Metaphorical and Cultural Aspects of Semantic Structure*. Cambridge: Cambridge University Press.
- Tao, H. 2007. "Subjectification and the development of special-verb existential/presentative constructions." *Language and Linguistics* 8(2): 575–602.
- Traugott, E. C. 1989. "On the rise of epistemic meanings in English: An example of subjectification in semantic change." *Language* 65: 31–55.
- Traugott, E. C. 1995. "Subjectification in Grammaticalization." In *Subjectivity and Subjectivisation: Linguistic Perspectives*, ed. by D. Stein, and S. Wright, 31–54. Cambridge: Cambridge University Press.
- Traugott, E. C. 2010. "(Inter)subjectivity and (inter)Subjectification: A Reassessment." In *Subjectification, Intersubjectification and Grammaticalization*, ed. by H. Cuyckens, K. Davidse, and L. Vandelanotte, 29–71. Berlin / New York: Mouton de Gruyter.
- Van Dijk, T. A. 1979. "Pragmatic connectives." *Journal of Pragmatics* 3: 447–456.
- Xing, F. 1996. *Hanyu yufaxue [Chinese Grammar]*. Changchun: Northeast Normal University Press.
- Xing, F. 2001. *Hanyu fujū yanjiu [A study of Chinese complex sentences]*. Beijing: Commercial Publishing House.
- Zhang, H. 1994. "The Grammaticalization of 'Bei' in Chinese." In *Chinese Languages and Linguistics II: Historical Linguistics*, ed. by P. J. K. Li, C. R. Huang, and C. C. J. Tang, 321–360. Taipei: Academia Sinica.
- Zhang, B. 1996. "Lianci de zaifenlei [Subdivision within Conjunctions]." In *Cilei wenti kaocha [An Examination of Syntactic Categories]*, ed. by M. Hu, 431–445. Beijing: Beijing Language and Culture University Press.
- Zhang, Q. 2012. "Kejian de yufahua [The grammaticalization of 'kejian']." *Journal of Hubei University of Education* 29(5): 28–32.
- Zhang, Yan. 2012. "Yinguo fujū guanlian biaoji jufa — yuyi yanjiu — jiyu 'jiaohu zhuguanxing' renzhiguan [A syntax-semantics interface study of causal connectives]." *Journal of Foreign Languages* 35(3): 42–50.
- Zhang, Ya-ru. 2008. "Yushiji de duojiaduo fenxi [A comprehensive analysis of 'yushi' sentences]." *Journal of Yunnan Normal University (Teaching and Research on Chinese as a Foreign Language Edition)* 1: 51–57.
- Zufferey, S. 2012. "'Car, parce que, puisque' revisited: Three empirical studies on French causal connectives." *Journal of Pragmatics* 44(2): 138–153.

## Appendix 1. The analysis of *domain*

### A. Results of the General Loglinear analysis of *domain*

Model	$\chi^2$ (model)	df	<i>p</i> (model)	$\chi^2$ (factor)	df	<i>p</i> (factor)
1. constant+connective	813.04	40	<.001			
+ 2. genre	813.04	38	<.001	0	2	> .05
+ 3. domain	643.37	36	<.001	169.67	2	< .001
+ 4. domain*connective	140.80	28	<.001	502.57	8	< .001
+ 5. domain*genre	66.37	24	<.001	74.43	4	< .001
+ 6. domain*connective*genre	.00	.00	1.000	66.37	24	< .001

### B. Parameter estimates *domain* for Model 6

Parameter	Estimate	Std. Error	<i>z</i>	<i>p</i>
constant	3.157	.206	15.304	< .001
[connective = kějiàn]	-3.850	1.429	-2.694	.007
[connective = suōyǐ]	-.416	.327	-1.272	.203
[connective = yīncǐ]	.120	.283	.424	.672
[connective = yīn'ér]	.638	.255	2.504	.012
[genre = argumentative]	-.043	.295	-.147	.883
[genre = narrative]	.082	.286	.286	.775
[Domain = volitional content]	.661	.254	2.601	.009
[Domain = epistemic/speech-act]	-1.142	.419	-2.723	.006
[connective = kějiàn] * [Domain = vol. content]	-.661	2.016	-.328	.743
[connective = kějiàn] * [Domain = epistemic]	6.159	1.480	4.163	< .001
[connective = suōyǐ] * [Domain = vol. content]	-1.150	.484	-2.376	.017
[connective = suōyǐ] * [Domain = epistemic]	2.343	.510	4.596	< .001
[connective = yīncǐ] * [Domain = vol. content]	-1.923	.485	-3.962	< .001
[connective = yīncǐ] * [Domain = epistemic]	1.614	.487	3.315	.001
[connective = yīn'ér] * [Domain = vol. content]	-2.751	.518	-5.307	< .001
[connective = yīn'ér] * [Domain = epistemic]	.624	.486	1.284	.199
[genre = argumentative] * [Domain = vol. content]	-.025	.364	-.068	.946
[genre = argumentative] * [Domain = epistemic]	.471	.554	.850	.396
[genre = narrative] * [Domain = vol. content]	-.018	.353	-.051	.960
[genre = narrative] * [Domain = epistemic]	-1.180	.784	-1.505	.132
[connective = kějiàn] * [genre = argumentative] * [Domain = vol.]	.068	2.011	.034	.973

Parameter	Estimate	Std. Error	z	p
[connective = kějiàn] * [genre = argumentative] * [Domain = epistemic]	-.427	.497	-.860	.390
[connective = kějiàn] * [genre = argumentative] * [Domain = non-vol.]	.043	2.022	.022	.983
[connective = kějiàn] * [genre = narrative] * [Domain = vol.]	-.064	2.011	-.032	.975
[connective = kějiàn] * [genre = narrative] * [Domain = epistemic]	1.099	.748	1.468	.142
[connective = kějiàn] * [genre = narrative] * [Domain = non-vol.]	-.082	2.020	-.040	.968
[connective = suǒyì] * [genre = argumentative] * [Domain = vol.]	-.043	.518	-.083	.934
[connective = suǒyì] * [genre = argumentative] * [Domain = epistemic]	-.317	.507	-.626	.532
[connective = suǒyì] * [genre = argumentative] * [Domain = non-vol.]	-.346	.497	-.696	.486
[connective = suǒyì] * [genre = narrative] * [Domain = vol.]	1.226	.421	2.915	.004
[connective = suǒyì] * [genre = narrative] * [Domain = epistemic]	-.169	.788	-.214	.830
[connective = suǒyì] * [genre = narrative] * [Domain = non-vol.]	.492	.427	1.150	.250
[connective = yīncǐ] * [genre = argumentative] * [Domain = vol.]	.193	.545	.355	.722
[connective = yīncǐ] * [genre = argumentative] * [Domain = epistemic]	-.179	.512	-.349	.727
[connective = yīncǐ] * [genre = argumentative] * [Domain = non-vol.]	-.631	.446	-1.415	.157
[connective = yīncǐ] * [genre = narrative] * [Domain = vol.]	1.078	.467	2.307	.021
[connective = yīncǐ] * [genre = narrative] * [Domain = epistemic]	.023	.791	.029	.977
[connective = yīncǐ] * [genre = narrative] * [Domain = non-vol.]	.292	.381	.765	.444
[connective = yīn'ér] * [genre = argumentative] * [Domain = vol.]	.966	.549	1.760	.078
[connective = yīn'ér] * [genre = argumentative] * [Domain = epistemic]	.045	.531	.085	.933

Parameter	Estimate	Std. Error	z	p
[connective = yīn'ér] * [genre = argumentative] * [Domain = non-vol.]	-.732	.398	-1.839	.066
[connective = yīn'ér] * [genre = narrative] * [Domain = vol.]	.757	.552	1.372	.170
[connective = yīn'ér] * [genre = narrative] * [Domain = epistemic]	.739	.791	.935	.350
[connective = yīn'ér] * [genre = narrative] * [Domain = non-vol.]	-.059	.355	-.167	.867

## Appendix 2. The analysis of *modality*

### A. Results of the General Loglinear analysis of *modality* (of the consequent, Q)

Model	$\chi^2$ (model)	df	p (model)	$\chi^2$ (factor)	df	p (factor)
1. constant+connective	1068.64	40	< .001			
+ 2. genre	1068.64	38	< .001	0	2	> .05
+ 3. modality	581.59	36	< .001	487.05	2	< 0.001
+ 4. modality*connective	151.98	28	< .001	429.61	8	< 0.001
+ 5. modality*genre	61.09	24	< .001	90.89	4	< 0.001
+ 6. modality*connective*genre	.00	.00	1.000	61.09	24	< 0.001

### B. Parameter estimates *modality* for Model 6

Parameter	Estimate	Std. Error	z	p
constant	3.068	.216	14.226	< .001
[connective = kějiàn]	-3.761	1.431	-2.629	.009
[connective = yúshì]	1.099	.249	4.412	< .001
[connective = yīncǐ]	.350	.282	1.242	.214
[connective = yīn'ér]	.793	.260	3.050	.002
[genre = argumentative]	-.327	.333	-.982	.326
[genre = narrative]	.893	.256	3.487	< .001
[modality = mental fact]	-1.815	.576	-3.149	.002
[modality = judgment]	.874	.257	3.402	.001
[connective = kějiàn] * [modality = mental fact]	1.815	2.081	.872	.383
[connective = kějiàn] * [modality = judgment]	4.144	1.442	2.874	.004
[connective = yúshì] * [modality = mental fact]	-.847	.755	-1.122	.262
[connective = yúshì] * [modality = judgment]	-3.025	.463	-6.528	< .001

Parameter	Estimate	Std. Error	z	p
[connective = yīncǐ] * [modality = mental fact]	-.350	.807	-.433	.665
[connective = yīncǐ] * [modality = judgment]	-.542	.350	-1.549	.121
[connective = yīn'ér] * [modality = mental fact]	-1.129	.868	-1.301	.193
[connective = yīn'ér] * [modality = judgment]	-1.457	.353	-4.126	< .001
[genre = argumentative] * [modality = mental fact]	.327	.826	.396	.692
[genre = argumentative] * [modality = judgment]	.437	.384	1.138	.255
[genre = narrative] * [modality = mental fact]	.206	.668	.308	.758
[genre = narrative] * [modality = judgment]	-2.232	.399	-5.596	< .001
[connective = kějiàn] * [genre = argumentative] * [modality = mental]	1.816E-16	2.138	.000	1.000
[connective = kějiàn] * [genre = argumentative] * [modality = judgment]	-.110	.252	-.438	.661
[connective = kějiàn] * [genre = argumentative] * [modality = physical]	.327	2.028	.161	.872
[connective = kějiàn] * [genre = narrative] * [modality = mental]	-1.099	2.093	-.525	.600
[connective = kějiàn] * [genre = narrative] * [modality = judgment]	1.339	.346	3.865	< .001
[connective = kějiàn] * [genre = narrative] * [modality = physical]	-.893	2.016	-.443	.658
[connective = yúshì] * [genre = argumentative] * [modality = mental]	-.251	1.039	-.242	.809
[connective = yúshì] * [genre = argumentative] * [modality = judgment]	.317	.507	.626	.532
[connective = yúshì] * [genre = argumentative] * [modality = physical]	.280	.378	.740	.459
[connective = yúshì] * [genre = narrative] * [modality = mental]	-.588	.858	-.685	.493
[connective = yúshì] * [genre = narrative] * [modality = judgment]	.240	.792	.303	.762
[connective = yúshì] * [genre = narrative] * [modality = physical]	-.862	.310	-2.781	.005
[connective = yīncǐ] * [genre = argumentative] * [modality = mental]	-4.225E-16	1.069	.000	1.000
[connective = yīncǐ] * [genre = argumentative] * [modality = judgment]	.138	.281	.494	.622
[connective = yīncǐ] * [genre = argumentative] * [modality = physical]	-.173	.445	-.388	.698



Parameter	Estimate	Std. Error	z	p
[connective = yīncǐ] * [genre = narrative] * [modality = mental]	.717	.844	.849	.396
[connective = yīncǐ] * [genre = narrative] * [modality = judgment]	.264	.431	.611	.541
[connective = yīncǐ] * [genre = narrative] * [modality = physical]	-.609	.351	-1.737	.082
[connective = yīn'ér] * [genre = argumentative] * [modality = mental]	.588	1.093	.538	.591
[connective = yīn'ér] * [genre = argumentative] * [modality = judgment]	.362	.313	1.156	.248
[connective = yīn'ér] * [genre = argumentative] * [modality = physical]	-.149	.407	-.366	.714
[connective = yīn'ér] * [genre = narrative] * [modality = mental]	.659	.922	.715	.475
[connective = yīn'ér] * [genre = narrative] * [modality = judgment]	.980	.430	2.276	.023
[connective = yīn'ér] * [genre = narrative] * [modality = physical]	-.981	.331	-2.962	.003

### Appendix 3. The analysis of *presence of SoC*

#### A. Results of the General Loglinear analysis of *presence of SoC*

Model	$\chi^2$ (model)	df	p (model)	$\chi^2$ (factor)	df	p (factor)
1. constant+connective	708.24	40	< .001			
+ 2. genre	708.24	38	< .001	0	2	> .05
+ 3. presence_SoC	475.05	36	<.001	233.19	2	< 0.001
+ 4. presence_SoC*connective	91.45	28	<.001	383.60	8	< 0.001
+ 5. presence_SoC*genre	44.73	24	.006	46.72	4	< 0.001
+ 6. presence_SoC*connective*genre	.00	.00	1.000	44.73	24	= 0.006

#### B. Parameter estimates *presence of SoC* for Model 6

Parameter	Estimate	Std. Error	z	p
constant	3.157	.206	15.304	< .001
[connective = kějiàn]	-3.850	1.429	-2.694	.007
[connective = suōyǐ]	-.416	.327	-1.272	.203

Parameter	Estimate	Std. Error	z	p
[connective = yīncǐ]	.120	.283	.424	.672
[connective = yīn'ér]	.638	.255	2.504	.012
[genre = argumentative]	-.043	.295	-.147	.883
[genre = narrative]	.082	.286	.286	.775
[presence_SoC = explicit]	.440	.264	1.665	.096
[presence_SoC = implicit]	-.354	.321	-1.101	.271
[connective = kějiàn] * [presence_SoC = explicit]	-.440	2.017	-.218	.827
[connective = kějiàn] * [presence_SoC = implicit]	5.371	1.455	3.692	< .001
[connective = suǒyǐ] * [presence_SoC = explicit]	-.440	.446	-.987	.324
[connective = suǒyǐ] * [presence_SoC = implicit]	1.431	.435	3.285	.001
[connective = yīncǐ] * [presence_SoC = explicit]	-1.577	.475	-3.323	.001
[connective = yīncǐ] * [presence_SoC = implicit]	.802	.406	1.975	.048
[connective = yīn'ér] * [presence_SoC = explicit]	-2.364	.496	-4.764	< .001
[connective = yīn'ér] * [presence_SoC = implicit]	-.203	.406	-.500	.617
[genre = argumentative] * [presence_SoC = explicit]	-.013	.379	-.034	.973
[genre = argumentative] * [presence_SoC = implicit]	.211	.446	.472	.637
[genre = narrative] * [presence_SoC = explicit]	-.198	.374	-.529	.597
[genre = narrative] * [presence_SoC = implicit]	.033	.443	.074	.941
[connective = kějiàn] * [genre = argumentative] * [presence_SoC = explicit]	.056	2.014	.028	.978
[connective = kějiàn] * [genre = argumentative] * [presence_SoC = implicit]	-.167	.372	-.449	.653
[connective = kějiàn] * [genre = argumentative] * [presence_SoC = no SoC]	.043	2.022	.022	.983
[connective = kějiàn] * [genre = narrative] * [presence_SoC = explicit]	.116	2.014	.058	.954
[connective = kějiàn] * [genre = narrative] * [presence_SoC = implicit]	-.114	.376	-.305	.761
[connective = kějiàn] * [genre = narrative] * [presence_SoC = no SoC]	-.082	2.020	-.040	.968
[connective = suǒyǐ] * [genre = argumentative] * [presence_SoC = explicit]	-.159	.448	-.354	.723
[connective = suǒyǐ] * [genre = argumentative] * [presence_SoC = implicit]	-.005	.391	-.013	.990
[connective = suǒyǐ] * [genre = argumentative] * [presence_SoC = no SoC]	-.346	.497	-.696	.486

Parameter	Estimate	Std. Error	z	p
[connective = suǒyì] * [genre = narrative] * [presence_SoC = explicit]	.532	.406	1.309	.190
[connective = suǒyì] * [genre = narrative] * [presence_SoC = implicit]	-.693	.419	-1.654	.098
[connective = suǒyì] * [genre = narrative] * [presence_SoC = no SoC]	.492	.427	1.150	.250
[connective = yīncǐ] * [genre = argumentative] * [presence_SoC = explicit]	.268	.519	.516	.606
[connective = yīncǐ] * [genre = argumentative] * [presence_SoC = implicit]	.068	.394	.173	.863
[connective = yīncǐ] * [genre = argumentative] * [presence_SoC = no SoC]	-.631	.446	-1.415	.157
[connective = yīncǐ] * [genre = narrative] * [presence_SoC = explicit]	1.175	.465	2.524	.012
[connective = yīncǐ] * [genre = narrative] * [presence_SoC = implicit]	-1.237	.461	-2.682	.007
[connective = yīncǐ] * [genre = narrative] * [presence_SoC = no SoC]	.292	.381	.765	.444
[connective = yīn'ér] * [genre = argumentative] * [presence_SoC = explicit]	.325	.573	.567	.571
[connective = yīn'ér] * [genre = argumentative] * [presence_SoC = implicit]	.455	.415	1.097	.273
[connective = yīn'ér] * [genre = argumentative] * [presence_SoC = no SoC]	-.732	.398	-1.839	.066
[connective = yīn'ér] * [genre = narrative] * [presence_SoC = explicit]	.116	.605	.192	.848
[connective = yīn'ér] * [genre = narrative] * [presence_SoC = implicit]	-.154	.441	-.350	.726
[connective = yīn'ér] * [genre = narrative] * [presence_SoC = no SoC]	-.059	.355	-.167	.867

## Appendix 4. The analysis of *identity of SoC*

### A. Results of the General Loglinear analysis of *identity of SoC*

Model	$\chi^2$ (model)	df	p (model)	$\chi^2$ (factor)	df	p (factor)
1. constant+connective	767.73	40	< .001			
+ 2. genre	767.73	38	< .001	0	2	> .05

Model	$\chi^2$ (model)	df	$p$ (model)	$\chi^2$ (factor)	df	$p$ (factor)
+ 2. identity_SoC	718.12	36	< .001	49.61	2	< 0.001
+ 3. identity_SoC*connective	343.32	28	< .001	374.80	8	< 0.001
+ 4. identity_SoC*genre	82.36	24	< .001	260.96	4	< 0.001
+ 5. identity_ SoC*connective*genre	.00	.00	1.000	82.36	24	< 0.001

B. Parameter estimates *identity of SoC* for Model 5

Parameter	Estimate	Std. Error	$z$	$p$
constant	2.140	.343	6.239	< .001
[connective = kějiàn]	-2.833	1.455	-1.947	.052
[connective = suōyǐ]	.534	.432	1.236	.216
[connective = yúshì]	1.721	.372	4.620	< .001
[connective = yīn'ér]	-9.725E-17	.485	.000	1.000
[genre = argumentative]	-3.353E-18	.485	.000	1.000
[genre = narrative]	.973	.403	2.418	.016
[identity_SoC = current speaker]	.534	.432	1.236	.216
[identity_SoC = author]	1.174	.392	2.992	.003
[connective = kějiàn] * [identity_SoC = current speaker]	2.601	1.508	1.725	.084
[connective = kějiàn] * [identity_SoC = author]	3.686	1.473	2.502	.012
[connective = suōyǐ] * [identity_SoC = current speaker]	-.051	.546	-.094	.925
[connective = suōyǐ] * [identity_SoC = author]	-.691	.515	-1.341	.180
[connective = yúshì] * [identity_SoC = current speaker]	-2.690	.624	-4.310	< .001
[connective = yúshì] * [identity_SoC = author]	-5.728	1.475	-3.884	< .001
[connective = yīn'ér] * [identity_SoC = current speaker]	-.323	.632	-.511	.610
[connective = yīn'ér] * [identity_SoC = author]	-.711	.588	-1.210	.226
[genre = argumentative] * [identity_SoC = current speaker]	-.802	.677	-1.185	.236
[genre = argumentative] * [identity_SoC = author]	.567	.541	1.049	.294
[genre = narrative] * [identity_SoC = current speaker]	-1.296	.571	-2.269	.023
[genre = narrative] * [identity_SoC = author]	-2.583	.617	-4.188	< .001

Parameter	Estimate	Std. Error	z	p
[connective = kějiàn] * [genre = argumentative] * [identity_SoC = current]	-1.235	.988	-1.249	.212
[connective = kějiàn] * [genre = argumentative] * [identity_SoC = author]	-.423	.293	-1.444	.149
[connective = kějiàn] * [genre = argumentative] * [identity_SoC = character]	9.924E-17	2.058	.000	1.000
[connective = kějiàn] * [genre = narrative] * [identity_SoC = current speaker]	2.093	.516	4.057	< .001
[connective = kějiàn] * [genre = narrative] * [identity_SoC = author]	-.417	.593	-.704	.482
[connective = kějiàn] * [genre = narrative] * [identity_SoC = character]	-.973	2.040	-.477	.633
[connective = suōyǐ] * [genre = argumentative] * [identity_SoC = current]	-.851	.698	-1.218	.223
[connective = suōyǐ] * [genre = argumentative] * [identity_SoC = author]	.198	.345	.572	.567
[connective = suōyǐ] * [genre = argumentative] * [identity_SoC = character]	-.232	.625	-.371	.711
[connective = suōyǐ] * [genre = narrative] * [identity_SoC = current speaker]	-.031	.517	-.060	.952
[connective = suōyǐ] * [genre = narrative] * [identity_SoC = author]	-2.241	1.504	-1.490	.136
[connective = suōyǐ] * [genre = narrative] * [identity_SoC = character]	-.166	.512	-.325	.745
[connective = yúshì] * [genre = argumentative] * [identity_SoC = current]	-.497	1.035	-.480	.631
[connective = yúshì] * [genre = argumentative] * [identity_SoC = author]	2.728	1.460	1.869	.062
[connective = yúshì] * [genre = argumentative] * [identity_SoC = character]	-.184	.531	-.348	.728
[connective = yúshì] * [genre = narrative] * [identity_SoC = current speaker]	.323	.727	.444	.657
[connective = yúshì] * [genre = narrative] * [identity_SoC = author]	1.609	2.054	.784	.433
[connective = yúshì] * [genre = narrative] * [identity_SoC = character]	-1.016	.453	-2.244	.025
[connective = yīn'ér] * [genre = argumentative] * [identity_SoC = current]	-.296	.777	-.381	.703

Parameter	Estimate	Std. Error	z	p
[connective = yīn'ér] * [genre = argumentative] * [identity_SoC = author]	.371	.400	.927	.354
[connective = yīn'ér] * [genre = argumentative] * [identity_SoC = character]	.778	.638	1.219	.223
[connective = yīn'ér] * [genre = narrative] * [identity_SoC = current speaker]	.323	.596	.542	.588
[connective = yīn'ér] * [genre = narrative] * [identity_SoC = author]	-.077	.832	-.092	.926
[connective = yīn'ér] * [genre = narrative] * [identity_SoC = character]	-.196	.578	-.339	.735

### *Authors' addresses*

Fang Li  
 Utrecht Institute of Linguistics OTS  
 Utrecht University  
 Trans 10  
 NL-3512 JK Utrecht, The Netherlands  
 F.Li@uu.nl

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

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