How subjective are Mandarin REASON connectives?

A corpus study of spontaneous conversation, microblog and newspaper discourse

Hongling Xiao,[1,4] Fang Li,[2] Ted J. M. Sanders,[3] and Wilbert P. M. S. Spooren[4]
[1] Xiamen University | [2] Beijing International Studies University |
[3] Utrecht University | [4] Radboud University Nijmegen

Studies in several languages find that causal connectives differ from one another in their prototypical *meaning and use*, which provides insight into language users' cognitive categorization of causal relations in discourse. Subjectivity plays a vital role in this process. Using an integrated subjectivity approach, this study aims to give a comprehensive picture of the semantic-pragmatic distinctions between Mandarin REASON connectives *jìrán* 'since', *yīnwèi* and *yóuyú* 'because'. The data come from spontaneous conversation, microblog, and newspaper discourse, while most previous studies have focused only on written data. The results show that, despite the contextual differences in discourse from each corpus, the connectives display distinctive and robust profiles. *Jìrán* is subjective. It prototypically expresses *speech act* and *epistemic* causalities featuring *speech act* and *judgment* in the consequent. *Speaker SoC* (subject of consciousness) is actively involved yet remains *implicit* in the utterances. *Yóuyú*, by contrast, is objective. It typically expresses *volitional* and *non-volitional content* causalities featuring the consequent of *physical act* and *fact,* which are usually independent of SoCs. *Yīnwèi* is neutral in general, with a slight preference to *volitional content* and *epistemic* relations, to the consequent of *fact*, and to *speaker SoC*. Only one interaction with discourse style is found: in relations introduced by *yīnwèi,* the linguistic realization of the SoC varies across corpora: significantly more *implicit* yet few *explicit* cases in microblogs, yet the opposite is true in conversations. The specific profile of *yīnwèi*, depending on the ordering of the antecedent and the consequent, is robust across corpora. Furthermore, the relative importance of the associated subjectivity features is determined. In conclusion, the study contributes to our understanding of causal coherence and extends the

https://doi.org/10.1075/lali.00080.xia
Language and Linguistics 22:1 (2021), pp. 166–210. issn 1606-822x | e-issn 2309-5067 © ILAS

empirical database that supports the claims of a cognitive account of causal coherence relations.

Keywords: subjectivity, Mandarin REASON connective, multi-style discourse

1. Introduction

1.1 Connective and subjectivity

Connectives like *because, but*, and *and* function as important linguistic devices to explicitly signal coherence relations in discourse. In most languages, there is more than one connective to mark a single type of coherence relation, be it causal, conditional or otherwise. Previous studies on causal connectives in particular, which include REASON connectives like *because* and RESULT connectives like *so*, have found that language users systematically prefer one connective over another to express a specific type of discourse level causality. This holds, for example, for the REASON connectives *omdat vs want* in Dutch, *parce que vs puisque* in French, and *yīnwèi vs yóuyú* in Mandarin, which are all roughly equal to their English counterpart *because* (Zufferey 2012; Sanders & Spooren 2015; Li et al. 2016). These differences are especially interesting because they provide insight into the specific function underneath the individual lexical form that each connective takes.

Linguists studying connectives agree that, in addition to its syntactic/grammatical role as a coherence marker, each connective is also inherently meaningful with an expressive function in both the semantic and the pragmatic senses (Van Dijk 1979; Sanders et al. 1992; Knott & Sanders 1998). In this paper, we focus on the *meaning and use* of Mandarin REASON connectives *jìrán* 'since', *yīnwèi* and *yóuyú* 'because'. As illustrated in (1–3),[1] they are not equally suitable to express the causal relation in each of the three fragments and their respective contexts.

(1) ***Jìrán/*Yīnwèi/*Yóuyú*** *P Ø shuō yě shì qiǎngbǎo zhōng de jùyīng, Q Ø shìbùshì qiàn wǒ yīshǒu cuīmiánqǔ a?*
Jìrán/*Yīnwèi/*Yóuyú *P* (you) say I am a big baby in the cradle, *Q* isn't it that (you) owe me a lullaby?

1. All illustrating causal constructions except (5), (6), and (7) are taken from our corpora. The unmarked connective is the one originally used in the corpus text; the question mark <?> indicates a possible alternative; the asterisk <*> indicates an improper option. The Mandarin text is put in italics, the connectives under study in bold and italics in both the original and the translated texts. The <Ø> symbol in Mandarin text indicates a null subject (*pro*-drop) which is made explicit and put in parentheses in the translated text. <P> and <Q> stand for the cause segment (the antecedent) and the result segment (the consequent), respectively.

(2) ***Yīnwèi /?Jìrán/*Yóuyú*** *P liàng dà fú dà, Q zài rénshēng de lǚtú shàng, zhǐyǒu kāikuò zìjǐ de xīnxiōng, tiāndì cái huì kāikuò!*
Yīnwèi/?Jìrán/*Yóuyú *P* generosity brings happiness, *Q* during the life journey, only by maintaining generous oneself, can his world be wide open!

(3) ***Yóuyú/?Yīnwèi/*Jìrán*** *P liánxì fāngshì hěnzǎo jiù shānchú le, Q wǒ xiànzài zhǐyǒu tā péngyǒu de.*
Yóuyú/?Yīnwèi/*Jìrán *P* (her) contact information has long been deleted, *Q* now I only have her friend's (contact information).

All three relations can be classified in terms of the tripartite system defined in Sweetser (1990). The causal relation holding in (1) corresponds to the speech act domain: "I" raised the question "isn't it that (you) owe me a lullaby" because "(you) say I am a big baby in the cradle". *Jìrán* is the sole proper connective that naturally links the two segments here. The relation in (2) corresponds to the epistemic domain: the speaker[2] concludes that "[…], only by […] can the world be wide open" based on the argument "generosity brings happiness". This argument-claim relation between the two segments needs to refer to the speaker for interpretation. *Yīnwèi* is the connective that was originally used by the blogger, which is certainly adequately used, but, intuitively, *jìrán* is also acceptable. The relation in (3) corresponds to the content domain: the fact that "(his) contact information has long been deleted" leads to the fact that "now I only have his friend's ([…])". The cause-consequence relationship exists *out there* in the physical world, independent of anyone's personal attitude or perspective. *Yóuyú* is originally from the blog, but *yīnwèi* may also fit. The apparent division of labor among *jìrán, yīnwèi*, and *yóuyú* suggests that Chinese speakers categorize causally related events by choosing different causal connectives, but at the same time, the overlap in some cases also shows that these connectives cannot be adequately distinguished only by the type of causal relations they typically signal.

The differences and similarities between *jìrán, yīnwèi,* and *yóuyú* have been widely discussed in literature. Some studies have taken into consideration the order and the information status of P (the antecedent) and Q (the consequent) of the causal event. It is found that *jìrán* normally signals forward causal constructions[3] (*jìrán P, Q*), wherein P "obligatorily" expresses given information (Eifring 1995: 49); *yóuyú* resembles *jìrán* in that it most often expresses given

2. The "speaker" refers to the first-person participant, which may be the speaker in oral speech, the writer in written texts, and the blogger in microblog messages.

3. Forward/backward causal construction has been termed, among others, as basic/non-basic order in Sander et al. (1992); forward/backward linking in Li & Thompson (1989) and Biq (1995); *zìrán/tèyì yǔxù* 'natural/special order'(Li & Liu 2004); initial/final causal clause sequence in Song & Tao (2009).

information in P and links forward constructions (*yóuyú P, Q*); *yīnwèi*, by contrast, is more flexible, yet it favors to some extent the backward construction (*Q, yīnwèi P*) and tends to provide new information in P (Liu et al. 2001; Guo 2008; Zhong & Zhang 2013; Zhao & Yao 2016).

The above findings are very inspiring and provide explanations for the focalization characteristics and the correlative usage of the connectives. With respect to focalization, *jìrán P,* for never providing pragmatically new information, normally cannot be focalized; *yóuyú P*, only when introducing new information, can be focalized but need to be explicitly marked with the COP *shì* 'be'(*Q, shì yóuyú P*); *yīnwèi*, given its flexibility, though tends to be focalized, a marker of focalization is not necessarily required (*Q, (shì) yīnwèi P*) (Liu 2002; Xing 2002; Li & Liu 2004; Li 2011; Zhong & Zhang 2013). Regarding the preferred correlative markers in the matrix clause (Q), *jìrán* usually correlates with the connective *nà/nàme* 'then' and the adverb *jiù* 'then', while *yīnwèi/yóuyú* most often with the connective *suǒyǐ* 'so' and both *jiù* and *cái* 'only then'. First, it has been argued that causality and inferentiality are inherent in the semantics of *nà/name*, in terms of the logical relationship between the two connected segments (e.g. Lǚ 1999; Wang 2015). Using *nà/nàme* to introduce Q makes it more salient as inferred from the premise (the given information) provided by *jìrán P*. Moreover, both *nà/nàme* and *suǒyǐ* may refer back to factual P, while only *suǒyǐ* may refer back to P capable of expressing new information (Eifring 1995: 384–385). That is why *suǒyǐ* is usually taken by *yīnwèi* (and sometimes by *yóuyú*), while *jìrán* usually takes *nà/nàme,* as the correlative marker in the matrix clause. Second, in the Q segment, the correlative adverb *jiù* [4] (for either *jìrán, yīnwèi*, or *yóuyú*) functions to assert an opinion that is already, or (as the speaker believes) should be, known and accepted by both the speaker and the interlocutor(s). The use of *cái*, on the other hand, is "to assert the speaker's own opinion (pragmatically new information) when there is a presupposition [...] of an opposite sort of opinion toward the subject of the discourse" (Biq 1988: 75–76).

Though it seems that the information status can explain major differences between *jìrán, yīnwèi*, and *yóuyú* in actual language use, one property seems to be more central to all these findings: subjectivity. It has long been established that *jìrán* is the prototypical connective expressing inferential causality while *yīnwèi* is typical in expressing expositive/descriptive causality (e.g. Huang 1998; Xing 2001; Guo 2008 and works cited there). It is also found that the difference between *jìrán* and *yīnwèi* lies in that *jìrán* constructions entail presupposition and inter-subjectivity, i.e. both P and Q are known and (at least as far as the

4. For a more extensive discussion of the differences and similarities between *jiù* and *cái*, please refer to Hole (2004) and Jin & Du (2014), and the vast literature mentioned there.

speaker believes) acceptable to both the speaker and the interlocutor(s), but this is not necessarily so with *yīnwèi* constructions (Eifring 1995: 384; Zhong & Zhang 2013). Furthermore, based on the domain analysis (Sweetser 1990; Shen 2003), *jìrán* is found to be used predominantly in the subjective domain (i.e. epistemic and speech act domains), yet *yóuyú*, by sharp contrast, is more frequently used in the objective domain (i.e. content domain); "*yīnwèi* is more subjective than *yóuyú*" (Li & Liu 2004; Li 2011: 492; Li et al. 2016).

Nevertheless, we still lack a comprehensive subjectivity framework (but see Li et al. 2016) and a solid empirical basis (e.g. a statistically evaluated corpus study using systematically selected data) for a fuller picture of these connectives' prototypical subjectivity profiles (and the causal relations they signal). In this paper, we shall perform a systematic analysis of the degree of subjectivity associated with them, based on the notions of *speaker involvement* and *subject of consciousness* (SoC) (Pander Maat & Sanders 2000, 2001; Pander Maat & Degand 2001). Here, the notion SoC refers to "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 2000: 64).

This is an integrated approach that considers subjectivity as a multi-dimensional concept. In addition to the domain of the relations (Sweetser 1990), subjectivity can be encoded in three other factors as well. That is, the propositional attitude of the consequent, e.g. *fact vs judgment*; the identity of SoC that conceptualizes this relationship, i.e. *first-person vs third-person* participant, and the linguistic realization of SoC, i.e. *explicit vs implicit*. This integrated subjectivity approach is elaborated on in § 2.

1.2 Connective and discourse styles

The above-mentioned studies on Mandarin REASON connectives are based on written, or mainly written with small amount of written-for-spoken discourse (e.g. play scripts). The analyses also lack in systematic statistic evaluations (but see Li et al. 2016). However, subjectivity is not only inherent in the positioning of the speaker with respect to a content by choosing certain linguistic items and structures, but also in the speaker-interlocutor interaction in the communication event (De Smet & Verstraete 2006). Will the findings in written discourse hold across discourse where messages are communicated via other distinct channels, say, spontaneous conversations and microblogging? In this section, we argue that, from both theoretical and empirical perspectives, it is highly necessary to add other than written data to the empirical foundation of the categorization of (Mandarin) REASON connectives.

In the first place, written language differs substantially from spoken language in linguistic and textual styles. Chafe (1982; 1984) has studied and explained the issue from the dimensions of time and space distance in relation to the cognitive and social processes typical of writing versus speaking. First, during the creation process, writers have the time to organize ideas and put them in well-planned linguistic structures, either simple or complex, and on solid argumentative foundation in order to stand the test of readers in different places and time periods. Therefore, written language is characteristic of being integrated. By contrast, the speaking process proceeds along real-time spontaneous interaction between speaker and addressee. Therefore, spoken language is fragmented, with less complex sentences and using words of more general meaning. Second, given the differing writer/speaker-interlocuter distance along the space dimension, written language, being decontextualized (Kay 1977), is more detached and the content is expressed in a more explicit way (Finegan 1987). Spoken language, on the contrary, is more involving and interactional. The first-person and second-person pronouns are more often used, as the interlocutors are situated in the Deictic Center of Communication (Sanders & Sweetser 2009; Sanders & Spooren 2013; Sanders & Spooren 2015).

To some extent, communication via social media (Twitter, Facebook, Microblog, etc.) is distinct from that via both traditional writing and speaking. It is true that the production process of texts in social media is controlled by the sender while the recipient(s) are unable to intervene (Voiskounsky 1997), which allows for the editing and planning of texts and ideas as in writing. The communication, nevertheless, also is characteristic of instantaneity and efficiency; consequently, the time pressure on both the sender and the recipient(s) may affect their strategy of choosing and using linguistic devices. Moreover, the sender often aims to exchange opinions and/or initiate interaction from the recipient(s) in a more direct and explicit way, for example, by using first and second person references as in fragment (4).

(4) *Q Wǒ bù jièyì nǐ de tǎoyàn,* ***yīnwèi P*** *Ø huózhe bìngfēi zhǐ wèi qǔyuè nǐ.*
Q I don't mind your dislike, ***yīnwèi P*** (I) am not living to please you.

Due to this speech-like profile, though technically typed and text-based, this type of internet-mediated discourse has recently been claimed to be *oralized written text* (Yus 2011) or *computer-mediated conversation*[5] (Herring 2011). In other words, the technical, cognitive, and social factors related to the networked environment have made this type of especially *interactive written discourse* as "an

5. For earlier terms used by other authors, please refer to Yus (2011: 175) and Herring (2011: 2).

emergent register" (Ferrara et al. 1991), with characteristics of both written and spoken language.

In the second place, studies based on other than only written data from some European languages have found either robust or dynamic pictures of the REASON connectives under analysis. French *parce que* has a dominant use in the content domain in written language, yet it is roughly equally distributed in spoken corpus over the content, the speech act and the epistemic domains; *puisque*, on the contrary, stays robust across spoken and written corpora: it is never used in the content domain and more often used in the epistemic than in the speech act domain (Zufferey 2012). German *weil* 'because' seems to be reserved for the content domain in written texts, but in spontaneous conversations it can also express epistemic relations (Günthner 1993, referred to in Sanders & Spooren 2015; Keller 1995). Spooren and colleagues (Spooren et al. 2010) conducted a comprehensive analysis of Dutch *omdat* and *want* by looking at the domain of use, the propositional attitude of the consequent, the identity of SoC and the way the SoC is linguistically realized. In both written and spoken data, *want* is more subjective than *omdat* with respect to all the four features (the identity of SoC differed between written and spoken discourse, yet without strong effect). This clear and stable difference was confirmed in the follow-up study based on more varied data, i.e. written texts, spontaneous conversation and online chat interactions (Sanders & Spooren 2015).

Given that the linguistic-textual features and the speaker-addressee interactions are typical of each style individually, will the *meaning and use* of these connectives remain robust or differ significantly across a range of distinctive discourse styles? This is an essential question we are going to explore in this study.

1.3 Research questions

This article aims at drawing a fuller picture of the subjectivity profiles of Mandarin REASON connectives *jìrán, yīnwèi*, and *yóuyú*. To that end, we made a systematic study using spontaneous conversation, microblog and newspaper discourse as the empirical data resources. The research questions we address are as follows:

RQ1. What is the prototypical subjectivity profile of each connective with regard to domain of the causal relation, the propositional attitude expressed in the consequent, the identity of SoC and the linguistic realization of the SoC in each corpus?

RQ2. Do the prototypical profiles remain robust or do they vary across corpora due to the contextual differences?

In naturally occurring language, *jìrán* and *yóuyú* mainly signal forward causal construction (*jìrán/yóuyú P, Q). Yīnwèi*, on the other hand, can signal both forward and backward constructions (*yīnwèi P, Q* and *Q, yīnwèi P*). In view of the functional nature of the two constructions, it is found that in both spoken and written data, the backward sequence *Q, yīnwèi P* typically expresses "interactionally motivated" causal relations; the "final" *yīnwèi* clause performs predominantly an "interactional" function with corresponding features of "negation, strong assertions, contrastiveness/comparison, and modality", as in (5); whereas the forward sequence *yīnwèi P, Q* expresses "ideationally (propositionally)" determined causal relations; the "initial" *yīnwèi* clause mainly performs an "information-sharing" function by providing more background information for an event or evidence for a statement, as in (6) (Biq 1995: 51; Song & Tao 2009: 87).

(5) *Q dāngshí wǒ céng xiǎng, búdàn rénrén yīnggāi xué*diǎn *xīnlǐxué, zhìguózhě sìhū yě yīng zuān- zuan zhè mén xuéwen.* ***Yīnwèi*** *P guīgēndàodǐ xīnlǐ jiù shì shìqì hé mínxīn.*
Q I was thinking then that not only **should** everybody study a little psychology, but the ruling party **should** especially do so, ***because*** *P* after all mentality has a lot to do with morale and popular sentiments.'

(6) *Jiù zài zhè nián 7 yuè,* ***yīnwèi*** *P pàntú chūmài, Q Huáng Jìng bèi mìmì dàibǔ.*
In July of the same year, ***because*** *P* a traitor sold him out (to the government), *Q* Huang Jing was secretly arrested.

(adapted from Song & Tao 2009: 89, 93)

Li et al. (2016) investigated the two positions of *yīnwèi* in written data from the perspective of subjectivity. The "sentence-initial *yīnwèi*" is found to typically express objective content relations existing between facts and independent of SoC, while the "inter-sentential *yīnwèi*" expresses subjective relations with judgments and implicit SoCs. The authors claim that the subjective profile of "inter-sentential *yīnwèi*" enables "final" causal clause to perform the interactional function; accordingly, the "information-sharing" function of "initial" causal clause relies on the objective profile of "sentence initial *yīnwèi*".

Inspired by the above interesting findings, in the present study, with microblog data, the *oralized written text,* added to the empirical basis next to news text and spontaneous conversation, we intend to find out answers to research question 3, and lastly, to research question 4 as follows:

RQ3. What are the prototypical subjectivity profiles of forward- and backward-linking *yīnwèi*? Is the respective profile robust or is it sensitive to the context characterized by discourse in each corpus?

RQ4. Do the factors under investigation, namely, the four subjectivity features and the ordering of the antecedent and the consequent, play equally important roles in characterizing each connective?

2. An integrated approach to subjectivity

Subjectivity is an innate nature of human language. "It is in and through language that man constitutes himself as a *subject*" (Benveniste 1971: 224). This has been clearly reflected in the cognitive categorization of causal relations in discourse. The involvement and viewpoint of the speaker and/or the SoC, which is realized through linguistic devices, plays a vital role in this process. Specifically in the present study, in order to systematically operationalize the complex notion of subjectivity, we employ an integrated approach by investigating (1) what domain does a causal relation fall into and whether its interpretation requires reference to an SoC; (2) who functions as the SoC; (3) how is the SoC realized linguistically; and (4) what are the propositional attitudes expressed in the consequent, given the assumption/argument in the antecedent.

This multi-dimensional concept of subjectivity (Spooren et al. 2010; Sanders & Spooren 2013, 2015; Li et al. 2013; Li et al. 2016) has integrated the influential views of Lyons (1977; 1982), Traugott (1989; 1995) and Langacker (1985; 1990). Lyons points out the importance of the expressive function of language for the "expression of attitude and personality" (1977: 50) and characterizes subjectivity as "the way in which natural languages, in their structure and in their normal manner of operation, provide for the locutionary agent's expression of himself and of his attitudes and beliefs" (1982: 102). Traugott (1982; 1989) takes a diachronic perspective studying the development of epistemic meanings in, for example, the auxiliary verb *must* and the discourse marker *while*. It is argued that the evolution of *epistemic* meaning of *must* in (7) from the *deontic* meaning of *must* in (8) and the change of discourse marker *while* from *temporal* meaning in (9) to *concessive* meaning in (10) indicate the tendency that meanings become increasingly situated in the speaker's subjective belief or attitude toward the proposition (1989: 35). This "expressiveness" of language that increases in such "a pragmatic-semantic process" is first defined as subjectivity (Traugott 1995: 47) and later refined as "speaker attitude or viewpoint" (Traugott 2010: 30).

(7) He must have worked very hard because he got a promotion.

(8) One must pass the exam for A1 before he/she can register the course for A2.

(9) He turned off the light while leaving.

(10) He didn't agree with her, while he didn't say it.

Following this notion of subjectivity, an utterance is considered subjective if its interpretation requires an SoC who evaluates and objective if not. The utterance (11) is presented as an opinion, the interpretation of which must refer to the speaker. Yet (12) expresses a physical fact, wherein no evaluation or personal viewpoint is present nor responsible for its truth value.

(11) Nijmegen is a lovely Dutch city.

(12) Nijmegen is located in the province of Gelderland.

When an SoC must be referred to for interpreting an utterance, the difference lies in the identity of the SoC. This is in line with the notion of subjectivity as closeness to the communicative *"here and now"* (Traugott 1989; 1995) or the *Deictic Center of Communication* (Sanders et al. 2009), i.e. the present time, location and participant role of the speech event. When the SoC coincides with the speaker, it is then part of the deictic center. The SoC's attitude/viewpoint are directly expressed in the first-person *voice* (Ducrot 1980). For example, by saying "I find Nijmegen a lovely city", the SoC "I" is actually making a personal judgment, whereas in "He thought Nijmegen is a lovely city", the SoC "he" and what "he thought" are outside the immediate communicative circumstances. What is expressed in the third person *voice* "he" takes its value as simply part of the discourse uttered by the speaker, which is not a judgment, but a description (Benveniste 1971). Consequently, it is less expressive (subjective). Hence, the distinction between speaker SoC and character SoC is made to evaluate the subjectivity of an utterance.

Focusing on the *perspective* of the speaker under the framework of cognitive grammar, Langacker (1985, 1990) studies the "gradience" of subjectivity in the construal of a particular entity or situation. The construal is considered as maximally objective when the participant is put *on-stage*, observable as an object or explicitly referred to as part of the *ground*, i.e. the speech event, its participants, and its immediate circumstances expressed via language (1990: 7–9). By contrast, the construal is maximally subjective when the participant is put *off-stage*, observing the scene as (a member of) the audience and remaining implicit in the utterance. For example, the construal of the situation is more subjective in (11), with the speaker responsible for the evaluation but remaining off-stage; whereas the construal will be relatively objectified in case the speaker "I" is put on-stage (e.g. ***I** think Nijmegen is a lovely city*. or *Nijmegen is a lovely city **to me**.*). Therefore, the implicit reference to the SoC is regarded as encoding a higher degree of subjectivity than the explicit reference.

Finally, since we are studying the causal coherence and connectives in discourse, the causal relation itself is where all the subjectivity features discussed above are anchored. Our concern naturally includes how subjective a causal event is in terms of the relation domain. Causality is a basic category in human cognition and natural language which can be subdivided further into different sets of relations (Sanders & Spooren 2009: 207–208). In this study, we follow the trichotomous classification by Sweetser (1990), as has been illustrated by Examples (1–3) in § 1.1: the content domain, the epistemic domain, and the speech-act domain. Content domain causal relations connect two events/situations that are causally related to each other in the content world. It is generally an objective category. Based on whether or not *intentionality* (of SoC) is involved in constructing the causal relation, content domain relations can be further divided into non-volitional content relations (e.g. *The ground is wet now because it has been raining this afternoon.*) and volitional content relations (e.g. *They want to postpone the game because the ground is still wet* .) (Degand & Pander Maat 2003; Evers-Vermeul 2005; Stukker et al. 2008). An epistemic relation (e.g. *It must have rained this afternoon since the ground is wet.*) is rooted in the mental world of the SoC who builds up the causal link through his/her personal reasoning. The same is true with speech act causal relations. For example, the interpretation of a causal relation in the fragment "May I have some salt, because I find the fish a bit light?" does not hold without referring to the SoC who makes the request because he/she personally "find[s] the fish a bit light". The degree of subjectivity encoded in causal relations increases from *non-volitional content* to *volitional content* and *epistemic/speech act.*

3. Methodology

3.1 Corpus and data collection

The empirical database for this study stems from a written corpus, a spoken corpus and a microblog corpus. The written corpus (a total of 163,393,972 Chinese characters) consists of newspaper texts from *Rénmín Rìbào* 'People's Daily', which is the sub-corpus of the news text corpus LDC95T13 from Linguistic Data Consortium (Wu 1995). From this corpus, 50 occurrences of *jìrán* were randomly selected. As *yīnwèi* and *yóuyú* both can be used as either a connective or a preposition, and due to the extremely high frequency (of *yīnwèi* in particular) in a large corpus like this, we randomly selected 100 fragments with *yīnwèi* and 100 with *yóuyú* and then manually counted for only the connective uses. This resulted in 81 fragments with connective *yīnwèi* and 68 fragments with connec-

tive *yóuyú*. Furthermore, we excluded 14 occurrences of *shì*(COP)*yīnwèi* and 3 of *shì*(COP)*yóuyú*, both literally meaning "be because" and usually occurring in the form of "*Zhī* (PRT) *suǒyǐ 'so' Q, shì*(COP) *yīnwèi/yóuyú 'because' P".* [6] Lastly, from the 67 *yīnwèi*- and 65 *yóuyú*-fragments, we randomly selected 50 of each for annotation and empirical analysis.

The microblog corpus, a sub-corpus of BCC (Xun et al. 2016), is made of posts in the year 2013 at *Sina Weibo*, the most popular microblogging platform in China, akin to a hybrid of *Facebook* and *Twitter*. The texts amount to 2,263,211,538 Chinese characters. Among all the occurrences, first we followed the same steps for collecting the written data, and then excluded those from institutional accounts to avoid the well-edited and planned linguistic features mirroring newspaper texts. In the end, 50 fragments with *jìrán, yīnwèi*, and *yóuyú* were randomly selected.

The spoken corpus consists of conversations on telephone and TV talk-show programs. The telephone corpus (Fung et al. 2005) contains 897 calls between Mandarin speakers in Mainland China, each lasting 10 minutes (150 hours in total). The transcripts amount to 2,030,730 Chinese characters. The corpus of TV talk-show conversations contains about 55 hours (92 texts) of *Luyu You Yue* (28 texts from Media Language Corpus [MLC 2005]), *Xinwen Huiketing* and *Shihua Shishuo* (translated respectively as *A Date with Luyu, People in the News* and *Tell It Like It Is)* (Walker et al. 2013–2015; Glenn et al. 2013–2015). The transcripts of this part amount to 911,816 Chinese characters. These programs were chosen for their characteristics of being interactive, casual, and entertaining, which provide a proper setting for spontaneous language use. The total size of the spoken corpus is 205 hours' audio speech with the transcripts of 2,942,546 Chinese characters. Table 1 presents the overall frequencies of *jìrán, yīnwèi*, and *yóuyú* in spoken corpus.

Table 1. Frequencies of *jìrán, yīnwèi*, and *yóuyú* in the spoken corpus

	Jìrán	*Yīnwèi*	*Yóuyú*
Telephone conversation (2,030,730)	151	3071	101
TV talk show conversation (911,816)	43	1757	44

All the fragments with *jìrán* and *yóuyú* were manually selected, disregarding repeated cases, false starts, incomplete sentences where Q segments cannot be

6. This structure has a specific function similar to the cleft sentence "Q, it is because P", wherein P is explicitly focalized and provides pragmatically new information. We abstract away from this special case as it differs from the rest on not only the syntactic level, but more importantly, on the pragmatic level.

specified, *shiyóuyú* and the prepositional uses of *yóuyú*. This resulted in 102 valid fragments with *jìrán* and 61 with *yóuyú*. 50 of each were randomly selected for final analysis. For *yīnwèi*, due to the extremely high frequencies and potentially large amount of invalid cases in spoken conversations, we first selected 200 occurrences in telephone speech and 200 in TV program conversations. In this process, we randomly jumped over the corpus to avoid getting many fragments in one text. Following the same procedure that was used for *yóuyú*, we ended up with 81 connective *yīnwèi,* from which 50 were randomly selected for the analysis.

In the end, the complete dataset under analysis are 450 fragments (50 per connective per corpus). See Table 2 for an overview of the dataset.

Table 2. Number of sampling fragments in each corpus

Connective	Newspaper (LDC)	Conversation (LDC&MLC)	Microblog (BCC)	Total
Jìrán	50	50	50	150
Yīnwèi	50	50	50	150
Yóuyú	50	50	50	150

3.2 Model of analysis

All the sampling fragments were annotated for the four subjectivity features discussed in § 2, each containing several categories of varying subjectivity degrees (Table 3). A similar model was employed in studies on Dutch and Mandarin REASON connectives (Spooren et al., 2010; Sanders & Spooren, 2015; Li et al. 2016). In the present study, we made modifications to some categories and explained in corresponding subsections.

Table 3. Model of integrated subjectivity

Subjectivity predictor	High.....Subjectivity degree.....Low
Domain	Speech-act/Epistemic…Volitional content...Non-volitional content
Propositional attitude (in Q)	Speech-act/Judgment…Mental act…Physical act…Fact
Identity of SoC	Speaker...Character...No SoC
Linguistic realization of SoC (in Q)	Implicit…Pro-drop…Explicit…Absent

3.2.1 *Causality domain*

Following Sweetser (1990), the causal relation expressed in each fragment was annotated as speech act domain, epistemic domain or content domain. The content domain was subdivided into volitional and non-volitional content domains based on whether or not the *intentionality* of SoC is involved in the conception of this causal relation. To systematically check the analyst's intuitions for domain of use, a framework of paraphrase test developed by Sanders (1997) was employed. See Table 4.

Table 4. Paraphrase test for determining *domain*

Domain	Paraphrase test
Speech act	The fact that P leads to the *SoC* (the speaker) *here and now* asking/warning/suggesting/offering/promising, etc. the addressee that Q.
Epistemic	The fact that P leads to the *SoC*'s *conclusion* here and now that Q.
Vol. content	The fact that P leads to the *SoC*'s *intentional* physical/mental *act* that Q.
Non-vol. content	The fact that P leads to the *fact* that Q.

Vol. content = volitional content; Non-vol. content = non-volitional content

In the speech act domain, explicit cases are utterances in the form of a (rhetorical) question and an imperative. This is the speech act causality considered in Li et al. (2016). In the current study, speech act is meant in the classical Austinian sense that an utterance can also be a social act, such as a declaration, a question, a promise, namely, a performative utterance (Huang 2006). That is, speech act utterances may also include (or even more often) declaratives carrying the *illocutionary* and/or the *perlocutionary* value of a speech act: suggesting, warning, commanding, offering, etc. targeting at the addressee (Austin 1962; Searle 1969). The decisive rule for determining the implicit cases lies in the communicative *here and now* and *the presence of addressee*: the speaker *here and now* performs or motivates the speech act targeting at the *addressee*. Example (13), from our spoken corpus, presents an instance of implicit speech act causality. The speaker *here and now* suggests the addressee (*you*) "take the initiative, uh, to contact him". At first sight, (14) seems to be similar, but this is not a case of speech act causality as it is not a suggestion located *here and now* but in the historical past "*dāngshí* (at that time)". Accordingly, Q is to express the speaker's personal opinion rather than to motivate a speech act.

(13) *Nǐ* ***jìrán*** *P bǎ tā dāng wèi hǎo péngyǒu dehuà, Q nǐ jiù yīnggāi qù zhǔdòng, è qù zhǎo tā a.*
Jìrán *P* you regard him as a good friend, *Q* then you should take the initiative, uh, to contact him.

(14) *"Nǐ **jìrán** P Ø xǐhuān fǎlǜ, Q dāngshí Ø jiù yīnggāi xuǎn fǎlǜ." Dàyé de! Wǒ nà shì kǎo bù shàng hǎo me?*
"***Jìrán** P* (you) were interested in law studies, *Q* then at that time (you) should apply for it". Damn it! (Didn't he know that) my scores were not good enough?

Examples (15–17) present an instance of the epistemic, the volitional and the non-volitional content causality, respectively.

(15) *Guóxué, hóngyáng zhōnghuá yōuxiù wénhuà. **Jìrán** P Ø shì shùn hū rénxīn, yīng hū cháoliú deshìyè, Q Ø bìrán huì fāzhǎn xiàqù de.*
Sinology carries forward the outstanding culture of China. ***Jìrán** P* (it) is a cause in harmony with people's wish and the popular trend, *Q* surely (it) will continue to develop.

(16) ***Yīnwèi** P wǒ shuō wǒ shì nǐ tóngshì, Q ránhòu tāmen jiù dǎ wǒ.*
***Yīnwèi** P* I said I was your colleague, *Q* then they beat me.

(17) *Dàn dào 80 niándài, **yóuyú** S1 [P yàzhōu yìxiē guójiā hé dìqū gǔlì chūkǒu, zhùzhòng chǎnpǐn zhiliàng,] S2 [Q zījīn liúrù zēngjiā,] S3 jīngjì chíxù fāzhǎn.*
But by the 1980s, ***yóuyú** S1* [*P* some Asian countries and regions promoted exports and paid attention to product quality,] *S2* [*Q* the capital inflows increased.] *S3* The economy continued to develop.

In (17), the three segments, S1 "some Asian countries and regions promoted exports and paid attention to product quality", S2 "the capital inflows increased", and S3 "the economy continued to develop", form a causal chain: S1 leads to the consequent S2, which in turn leads to the consequent S3. For cases like this, we simply annotated the first causal relation, leaving out the second one. However, when such a causal chain is embedded inside a *double-marked* fragment "REASON connective S1, S2, RESULT connective S3" like (18), we divided up P and Q in between the two correlative connectives. Therefore, (18) was annotated as a volitional content relation following the paraphrase test: the fact that "the activity venue was nearby the sea, it was very cold at night" leads to the intentional act "put on an extra coat while leaving (for it)".

(18) ***Yóuyú** P [S1huódòng chǎngdì kàojìn hǎibiān, S2 wǎnshàng tèbié lěng,] suǒyǐ Q [S3 chūmén de shíhòu Ø duō jiāle jiàn yīshang.]*
***Yóuyú** P [S1* the activity venue was nearby the sea, *S2* (it would be) very cold at night,] *suǒyǐ 'so' Q* [*S3* (I) put on an extra coat while leaving (for the activity).]

3.2.2 *Propositional attitude of the consequent*

In a causal coherence relation linking a consequent (Q) to an antecedent (P), P is given as the premise or an assumption, which usually is not subject to evaluations. The correlative consequence or the claim given the premise in P, how-

ever, is where the SoC's viewpoints/evaluations are accommodated. Therefore, the propositional attitude of Q segment (or *PropAtt* hereafter) is taken as a feature predicting the subjectivity degree of a given relation. We made a classification into five types:[7] Q expresses (1) a *speech act* if the SoC (the speaker) *here and now* makes a performative utterance targeting at the addressee (e.g. asking, suggesting, offering, commanding); (2) a *judgment* if it concerns the SoC's personal feelings, opinions, attitudes, etc. in response to the assumption in P; (3) a *mental act* if the predicate represents an intentional act involving the change of mind or the forming of an opinion, etc., such as decision-making, that takes place in the mental world of the SoC; (4) a *physical act* if the predicate represents an intentional act of the SoC taking place and observable in the physical world; (5) a *fact* if it describes an event or situation independent of the SoC's intention, be it in the mental or the physical world. Accordingly, the Q segment "you should take the initiative, uh, to contact him" in (13) was annotated as a *speech act*, "surely (it) will continue to develop" in (15) as a *judgment*, the intentional act "they beat me" in (16) as a *physical act*, "the capital inflows increased" in (17) as a *fact*, and "today (I) decide to take a break" in (19) as a *mental act*.

(19) *Q Tiàole liǎngtiān cāo, jīntiān Ø juédìng xiūxí yītiān,* ***yīnwèi*** *P Ø méiyǒu yīfú huànle.*
Q After exercising for two days (in a row), today (I) decide to take a break, ***yīnwèi*** *P* (I) have no outfits to change.

(20) *Q Ó, lánqiú Ø dào bùshì hěn xǐhuān,* ***yīnwèi*** *P Ø gèzi bùshì hěn gāo.*
Q Well, basketball, (I) don't like it very much, ***yīnwèi*** *P* (I) am not tall.

Note that in some cases, the line between *judgment* and *mental act* is not clear, since both develop in the mental world of the SoC. To specify the subtle differences and maintain consistency in our coding, we always refer to the context for the interpretation. If the interpretation of an utterance does not fundamentally change the meaning of the utterance in the context, then we would code it as the interpretation suggests. If the context denotes a change of state or the forming of an idea at a particular point in time, such as (19), it was coded as *mental act*. If it evokes the evaluative desire of the speaker, which is static and evaluative, as in (20), then it is coded as *judgment*. If the context does not allow us to distinguish between the two, then we would code it as "ambiguous" and later use the objectivity bias (see § 3.3).

7. For other classifications of the propositional attitude of Q, please refer to Li et al. (2016: 25) and Spooren et al. (2010: 244).

3.2.3 *Identity of SoC*

When an SoC is responsible for the construction of the causal relation, we made a distinction between the *speaker SoC* and the *character SoC*. Relations conceptualized by the speaker are more subjective than those by the character in the communication event; relations that hold independent of an SoC are the least subjective. In the above Example (20), the cause-effect relation between "*(I)* am not tall" and "well, basketball, *(I)* don't like it very much" is constructed from the speaker's personal point of view. In (21), that "*he* accepted the tough mission" leads to "*he* then devotes himself to it" is a relation conceptualized by the character (*he: Qiu Jinghua*) while the speaker is simply the reporter.

(21) *Rán'ér, qiūjìnghuá bùshì ruòzhě. Tā* ***jìrán*** *P Ø jiēle shuài bǎng, Q Ø biàn quán xīnshēn de tóurù.*
However, Qiu Jinghua was not a loser. ***Jìrán*** *P* he '*tā*' accepted the tough mission, *Q* he then devoted himself to it.

In the annotating stage, we encountered a few fragments with first-person plural "we" acting as the SoC. The speaker presents something on behalf of both herself and the other(s), all of whom are responsible for drawing the causal link between the two segments, Example (22). Such cases (eight in total) were first coded as having a *speaker+ SoC* and later merged into the category of *speaker SoC* in the stage of statistical analysis.

(22) *Q Wǒmen bù fǒurèn shāngyè diànyǐng de jiàzhí,* ***yīnwèi*** *P jìn jǐ nián shāngyè diànyǐng sìhū gèng yǒu kàn tóu, fǎn'ér wényìpiàn chéng le "chòu dàjiě".*
Q We do not deny the value of commercial films, ***yīnwèi*** *P* in recent years the commercial films seem more worth seeing, while the art pieces instead were poor in quality.

3.2.4 *Linguistic realization of SoC*

Following Langacker's (1985; 1990) view of subjectivity as a gradient phenomenon related to the implicit or explicit construal of a situation, we take the linguistic realization of the SoC (*LingReal-SoC*) as another subjectivity feature. The explicit reference to the SoC objectifies the construal of the causal relation to some extent. An implicit SoC, on the other hand, indicates a more subjective construal. As evaluations by SoC are manifest in the consequent, we look only at Q for this feature.

Given that Chinese is a typical *pro*-drop language, and *pro*-drop occurs especially often in spontaneous and informal discourse like conversations and microblogs, in this study, we set *pro*-drop as a separate category in addition to the implicit/explicit dichotomy used in previous studies. Fragment (23) illustrates

this phenomenon: the dropped-off speaker SoC "I" in Q can be easily identified from the previous utterance "Oh, uh, me too".

(23) *Ó, è, wǒ yěshì.* ***Yóuyú, yóuyú*** *P Ø zuìjìn gōngzuò bǐjiào máng, Q Ø yě méiyǒu tiāntiān kàn àoyùnhuì.*
Oh, uh, me too. ***Yóuyú, yóuyú*** *P* recently, (I) am busy, *Q* (I) do not watch the Olympic Games every day.

As the dropped SoC is actually present and recoverable in the context, we suppose *pro*-drop is a stronger marker of subjectivity than an explicit SoC, but not as strong as an implicit SoC. Cases without SoC (annotated as *Absent* in our model), are the least subjective.

Another possibility in Chinese is to drop the syntactic subject shared by both P and Q and position it instead in front of the sentence-initial connective. In such cases, when the *prepositioned subject* acts as SoC, Q was coded as having an explicit SoC. For example, (21) was coded as having an explicit SoC, i.e. *tā* 'he'.

3.3 Coding strategy for ambiguous cases

During the coding stage, we encountered some ambiguous cases with regard to the causality domain and the *PropAtt* (respectively, 11 and 7 cases out of 450), for which we aimed at consistency by choosing, slightly arbitrarily, the objective option (objectivity bias). By way of example, we here discuss two of these cases.

When both volitional and non-volitional content interpretations are acceptable, we opted for the non-volitional. For example, (24) was coded as non-volitional content domain in our study, although the consequent "(I) watch (NBA) wherever (I) can" might be an *intentional act* (volitional) or simply a *fact* due to the outside force "(I) do not have (my) own TV" (non-volitional).

When ambiguity occurred between interpreting Q as expressing *judgment* versus *mental act*, we opted for the interpretation of *mental act*. An example is (25), which can be interpreted as a *judgment* by the writer (the journalist) based on her observation of the situation that "the Community farmers" face, or as a *mental act* of the character in the speech event, i.e. "the Community farmers" themselves, that they (start to) "worry more about […]" under the circumstances provided in P.

(24) ***Yīnwèi*** *P zuòwéi xuéshēng Ø méiyǒu zìjide diànshì, suǒyǐ Q néng zànǎ kàn (NBA) wǒ jiù zàinǎ kàn.*
Yīnwèi *P* as a student, (I) don't have my own TV, *suǒyǐ* 'so' *Q* I then watch (NBA) wherever I can.

(25) *Q Gòngtóngtǐ nóngmín duì dìng yú běn yuè 15 rì zài rìnèiwǎ huífù de wūlāguī huíhé màoyì tánpàn gèng wèi yōulǜ,* ***yīnwèi*** *P jíshǐ àn gòngtóngtǐ tíchū de jiāng nóngyè bǔtiē xuējiǎn 30% de fāng'àn, tāmen de nián shōurù yě jiāng jiǎnshǎo 210 wàn ōuzhōu huòbì dānwèi.*
Q The Community farmers worry more about the Uruguay Round of trade negotiations scheduled to resume in Geneva on the 15th of this month, ***yīnwèi*** *P* even if the Community proposes a 30% reduction in agricultural subsidies, their annual income will still be reduced by 2.1 million in European currency.

3.4 Procedure

Of all the 450 fragments under analysis, 120 were first coded independently by the first two authors, who are native Mandarin speakers. Discrepancies were discussed in order to better define the coding criteria. Then another 90 fragments were coded independently by them and the Cohen's kappa was calculated. The kappa coefficients were: domain = 0.72; *PropAtt* = 0.74; *IdSoC* = 0.79; *LingReal-SoC* = 0.84. The kappa's indicate that the coding is sufficiently reliable. All discrepancies were discussed. Two fragments remained in discrepancy over domain (*volitional* vs *non-volitional content*) and two over *PropAtt* (*judgment vs mental act*). For these cases, we opted for the objective alternative (objectivity bias introduced in § 3.3). Final agreement was reached upon all data concerning *IdSoC* and *LingReal-SoC*.

For each of the four subjectivity features, a log-linear analysis was carried out to get the best fitting model of the relationship between *connective, subjectivity*, and *corpus* (to answer RQs 1 &2) and the relationship between *corpus, subjectivity*, and the forward/backward linking *yīnwèi's* (*Fwd/Bwd yīnwèi* hereafter) (to answer RQ3). To address RQ4, the relative importance of each factor involved in predicting the *meaning and use* of a particular connective, we carried out an analysis of conditional inference trees and random forests as introduced in Levshina (2015) and Tagliamonte & Baayen (2012).

4. Results

Results of the log linear analyses concerning RQs 1 & 2 are presented separately in terms of the four subjectivity features (§ 4.1–4.4); those concerning RQ3 are presented in § 4.5. We used an alpha level of .05 for all these statistical tests. The output of the conditional inference trees and the random forests analyses is presented in § 4.6.

4.1 Domain

Data with respect to *domain* are summarized in Figure 1 and Figure 2.

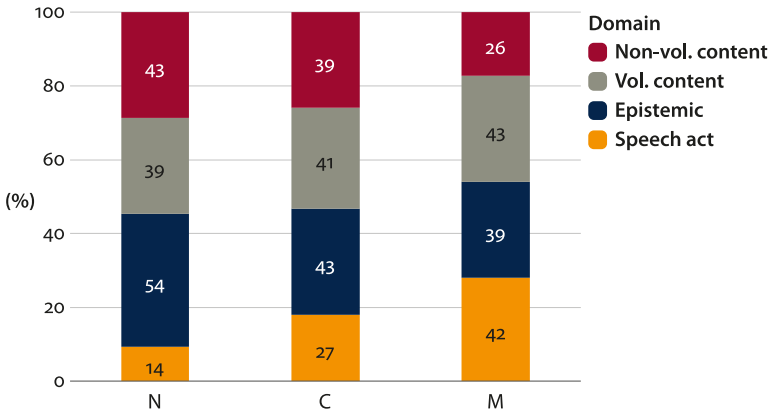

Figure 1. Frequencies and percentages of each domain category observed in the corpus of Newspaper (N), Conversation (C) and Microblog (M)

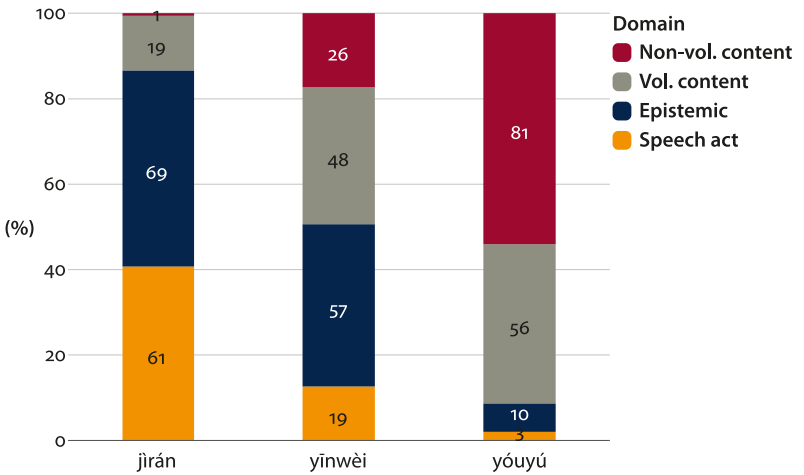

Figure 2. Frequencies and percentages of each domain category co-occurring with *jìrán, yīnwèi*, and *yóuyú*

The log-linear analysis resulted in a model containing two significant two-way interactions. The fit of the resulting model is moderate, $\chi 2$ (16, $N = 450$) = 24.93, $p = .07$. The first significant interaction is that between *Corpus* and *Domain*, $\chi 2$ (6, $N = 450$) = 22.08, $p < .01$. The other is between *Connective* and *Domain*, $\chi 2$ (6,

$N = 450) = 250.80$, $p < .001$. A cross-table analysis was made for a closer look at these interactions. See Table 5 and Table 6.

Table 5. Distribution of domain categories in each corpus (frequencies and standardized residuals)[8]

Corpus		Domain				
		Speech act	Epistemic	Vol. content	Non-vol. content	Total
Newspaper	Count	14	54	39	43	150
	Std. Res.	−2.6	1.3	−.3	1.2	
Conversation	Count	27	43	41	39	150
	Std. Res.	.1	−.3	.0	.5	
Microblog	Count	42	39	43	26	150
	Std. Res.	2.7	−.9	.3	−1.7	
Total		**83**	**136**	**123**	**108**	**450**

Seen from the standardized residuals, the interaction *Corpus*Domain* is mainly caused by a relatively low amount of *Speech act* in the newspaper corpus (s.r. = −2.6) and a high amount in microblogs (s.r. = 2.7); the slightly fewer occurrences of *Non-vol. content* in microblogs (s.r. = −1.7) also contribute to this interaction to some extent. Below, (26) exemplifies a typical speech act relation in the microblog corpus.

(26) *Q Ná qǐ diànhuà de shíhòu qǐng wéixiào,* ***yīnwèi*** *P duìfāng néng gǎnjuédào!*
Q Please smile while picking up the phone, ***yīnwèi*** *P* people on the other side can feel it!

Table 6. Co-occurring of domain categories and connective (frequencies and standardized residuals)

Connective		Domain				
		Speech act	Epistemic	Vol. content	Non-vol. content	Total
jìrán	Count	61	69	19	1	150
	Std. Res.	6.3	3.5	−3.4	−5.8	
yīnwèi	Count	19	57	48	26	150
	Std. Res.	−1.6	1.7	1.1	−1.7	
yóuyú	Count	3	10	56	81	150
	Std. Res.	−4.7	−5.2	2.3	7.5	
Total		**83**	**136**	**123**	**108**	**450**

8. Standardized residuals are delimited as "Std. Res." in tables and "s.r." in the text.

The interaction *Connective*Domain* is mainly caused by the extremely high amount of *jìrán* expressing *Speech act* (s.r. = 6.3) and *Epistemic* (s.r. = 3.5) rather than *Vol. content* (s.r. = −3.4) or *Non-vol. content*(s.r. = −5.8), and by the extremely low frequency of *yóuyú* in *Speech act* (s.r. = −4.7) and *Epistemic* (s.r. = −5.2) yet high frequency in *Vol. content* (s.r. = 2.3) and *Non-vol. content* (s.r. = 7.5). *Yīnwèi* seems to have a slight preference for *Epistemic* (s.r. = 1.7), whereas it is relatively rare in *Speech act* (s.r. = −1.6) and *Non- vol. content* (s.r. = −1.7).

Example (15), repeated here as (27), is the typical use of *jìrán* in the epistemic domain: the SoC comes to the conclusion "surely (it) will continue to develop" based on the premise that "(it) is a cause in harmony with […]"; (1), repeated here as (28), exemplifies *jìrán* in the speech act domain: the SoC asks "isn't it that (you) owe me a lullaby" because "(you) say I am a big baby in the cradle".

(27) *Guóxué, hóngyáng zhōnghuá yōuxiù wénhuà.* ***Jìrán*** *P Ø shì shùn hū rénxīn, yīng hū cháoliú deshìyè, Q Ø bìrán huì fāzhǎn xiàqù de.*
Sinology carries forward the outstanding culture of China. ***Jìrán*** *P* (it) is a cause in harmony with people's wish and the popular trend, *Q* surely (it) will continue to develop.

(28) ***Jìrán*** *P Ø shuō yě shì qiǎngbǎo zhōng de jùyīng, Q Ø shìbùshì qiàn* ***wǒ*** *yīshǒu cuīmiánqǔ a?*
Jìrán *P* (you) say I am a big baby in the cradle, *Q* isn't it that (you) owe **me** a lullaby?

Example (29) illustrates the typical use of *yóuyú* in a volitional content relation: the SoC "people engaged in corruption (*intentionally*) take various means to cover up" because "corruption violates the law by nature". (3), repeated here as (30), illustrates the typical use of *yóuyú* in a non-volitional content relation: the fact that "(her) contact information has long been deleted before" leads to the fact that "now I only have her friend's ([…])", wherein no evaluation nor intention of an SoC is involved.

(29) ***Yóuyú*** *P fǔbài xíngwéi shǔyú wéijì wéifǎ xìngzhì, Q gǎo fǔbài de rén yībān yào cǎiqǔ zhǒngzhǒng shǒuduàn jìnxíng yǎnshì.*
Yóuyú *P* corruption violates the law by nature, *Q* people engaged in corruption generally take various means to cover up.

(30) ***Yóuyú*** *P liánxì fāngshì hěn zǎo jiù shānchú le, Q wǒ xiànzài zhǐyǒu tā péngyǒu de.*
Yóuyú *P* (her) contact information has long been deleted before, *Q* now I only have her friend's (contact information).

Yīnwèi shows a relatively neutral profile, but with a slight preference to the epistemic domain, as in (31), and the volitional content domain, as in (32). The

two relations can be paraphrased respectively as follows: the fact that "you know everything (in your home country)" leads to the speaker's *conclusion* that "[...], you should be able to [...]"; the fact that "[...], I had too much food [...]" leads to *"my" intentional act* to "first have a banana for digestion".

(31) *Q Rúguǒ nǐ yǒu- yǒu zhè yībǎi wàn Ø yìng gāi néng zài zìjǐ de zǔguó shēnghuó dé gèng- bǐ dào wàiguó qù- qù gèng hǎo,* ***yīnwèi*** *P nǐ shénme dōngxī dū zhīdào.*
Q If you have, have a million (Rmb), (you) should be able to live a better, compared to living abroad, better life in your home country, ***yīnwèi*** *P* you know everything (in your home country).

(32) *Āiyā, nèitiān zǎoshang* ***yīnwèi*** *P tóu yītiān wǎnshàng wǒ chīde tàiduōle, chī bùxià, suǒyǐ 'so' Q Ø chīle xiāngjiāo qīng qīng chángzi.*
Oops! That morning, ***yīnwèi*** *P* the night before that morning, I had too much food and couldn't eat any more next morning, *suǒyǐ 'so' Q* (I) first had a banana for digestion.

4.2 Propositional attitude of the consequent

Data with respect to *PropAtt* are summarized in Figure 3 and Figure 4.

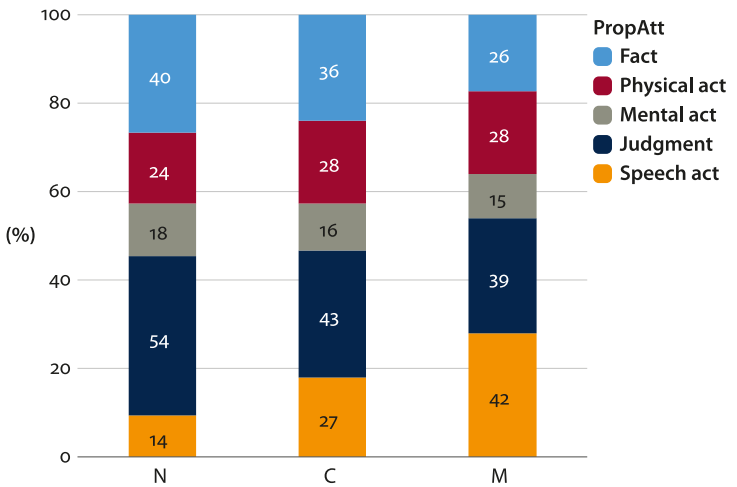

Figure 3. Frequencies and percentages of each PropAtt category observed in the corpus of Newspaper (N), Conversation (C) and Microblog (M)

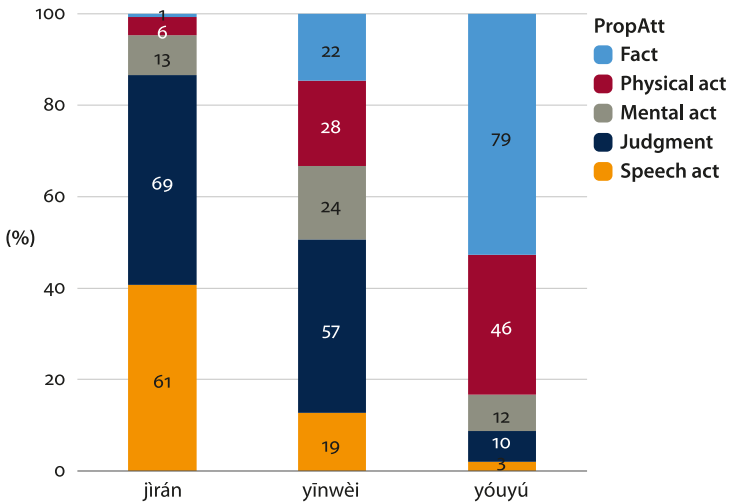

Figure 4. Frequencies and percentages of each PropAtt category co-occurring with *jìrán, yīnwèi*, and *yóuyú*

The log-linear analysis resulted in another model containing two significant two-way interactions. The fit of the model is adequate, $\chi 2\ (16, N=450)=26.81$, $p=0.14$. The first interaction is that between *Corpus* and *PropAtt*, $\chi 2\ (8, N=450)=21.15$, $p<.01$. The second is between *Connective* and *PropAtt*, $\chi 2\ (8, N=450)=269.83$, $p<.001$. The interactions are further elaborated on by a cross-table analysis summarized in Table 7 and Table 8.

Table 7. Distribution of *PropAtt* categories in each corpus (frequencies and standardized residuals)

Corpus		***PropAtt***					
		Speech act	**Judgment**	**Mental act**	**Physical act**	**Fact**	**Total**
Newspaper	Count	14	54	18	24	40	150
	Std. Res.	−2.6	1.3	.4	−.5	1.0	
Conversation	Count	27	43	16	28	36	150
	Std. Res.	.1	−.3	.1	.3	.3	
Microblog	Count	42	39	15	28	26	150
	Std. Res.	2.7	−.9	−.3	.3	−1.4	
Total		**83**	**136**	**49**	**80**	**102**	**450**

The standardized residuals demonstrate that the interaction *Corpus*PropAtt* is mainly caused by the variant distributions of *Speech act* over corpora: signifi-

cantly low in *Newspaper* (s.r. = −2.6) yet high in *Microblog* (s.r. = 2.7). The Q segment of fragment (26), "please smile while answering the phone", is a typical speech act utterance in the microblog corpus.

Table 8. Co-occurring of *PropAtt* categories and connective (frequencies and standardized residuals)

Connective		*PropAtt*					
		Speech act	**Judgment**	**Mental act**	**Physical act**	**Fact**	**Total**
jìrán	Count	61	69	13	6	1	150
	Std. Res.	6.3	3.5	−.8	−4.0	−5.7	
yīnwèi	Count	19	57	24	28	22	150
	Std. Res.	−1.6	1.7	1.9	.3	−2.1	
yóuyú	Count	3	10	12	46	79	150
	Std. Res.	−4.7	−5.2	−1.1	3.7	7.7	
Total		**83**	**136**	**49**	**80**	**102**	**450**

The interaction *Connective*PropAtt* is caused by high numbers of *jìrán* expressing *Speech act* (s.r. = 6.3) and *Judgment* (s.r. = 3.5) yet very low numbers of *Physical act* (s.r. = −4.0) and *Fact* (s.r. = −5.7), and by few instances of *yóuyú* expressing *Speech act* (s.r. = −4.7) and *Judgment* (s.r. = −5.2) while a great many *Physical act* (s.r. = 3.7) and *Fact* (s.r=7.7). The Q segments in *yīnwèi* fragments express relatively more *Judgment* (s.r. = 1.7) and *Mental act* (s.r. = 1.9) while less *Speech act* (s.r. = −1.6) and *Fact* (s.r. = −2.1).

The Q segments of (27), "surely (it) will continue to develop", and of (28), "isn't it that (you) owe me a lullaby", express respectively the typical propositional attitudes of *Judgment* and *Speech act* in *jìrán* fragments. Q in (29), "people engaged in corruption generally take various means to cover up", and in (30), "now I only have her friend's", express respectively *Physical act* and *Fact* in *yóuyú* fragments. The segments in (31) "[…], (you) should be able to […]" and in (22) "we do not deny the value of commercial films", illustrate respectively *Judgment* and *Mental act* in *yīnwèi* fragments.

4.3 Identity of SoC

Figure 5 and Figure 6 summed up data regarding the identity of SoC.

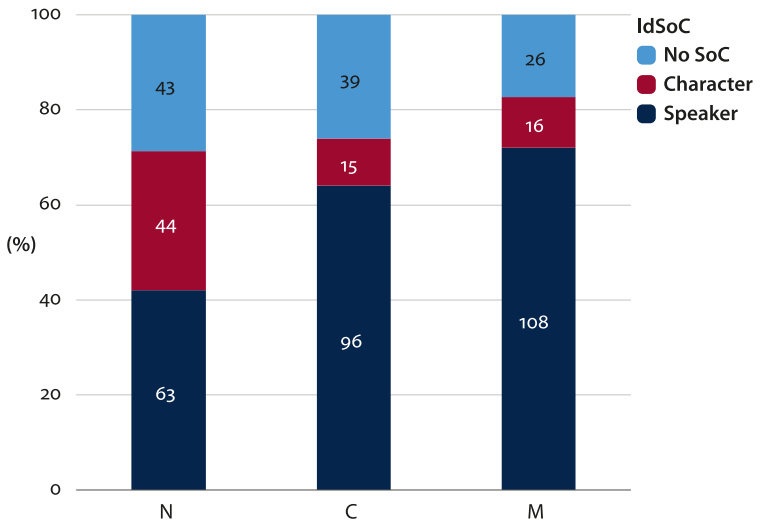

Figure 5. Frequencies and percentages of each IdSoC category observed in the corpus of Newspaper (N), Conversation (C) and Microblog (M)

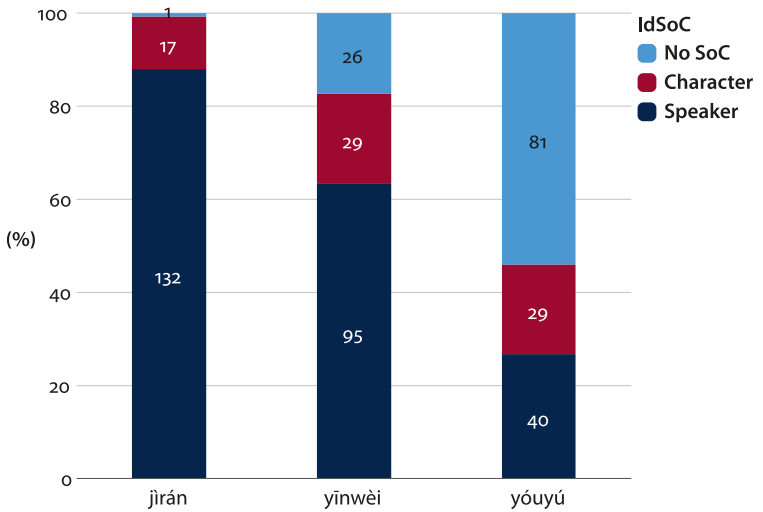

Figure 6. Frequencies and percentages of each PropAtt category co-occurring with *jìrán, yīnwèi*, and *yóuyú*

The log-linear analysis resulted in a third model containing two significant two-way interactions. The fit of the resulting model is adequate, $\chi 2\,(12, N = 450) = 16.03$, $p = .19$. The first significant interaction is between *Corpus* and *IdSoC*, $\chi 2$ (4,

$N = 450) = 37.54$, $p < .001$. The other is between *Connective* and *IdSoC*, $\chi 2$ (4, $N = 450) = 163.86$, $p < .001$. Tables 9 and 10 present the cross-table analysis.

Table 9. Distribution of *IdSoC* types in each corpus (frequencies and standardized residuals)

Corpus		**Identity of *SoC***			
		Speaker	**Character**	**No *SoC***	**Total**
Newspaper	Count	63	44	43	150
	Std. Res.	−2.8	3.8	1.2	
Conversation	Count	96	15	39	150
	Std. Res.	.7	−2.0	.5	
Microblog	Count	108	16	26	150
	Std. Res.	2.0	−1.8	−1.7	
Total		**267**	**75**	**108**	**450**

As shown by the standardized residuals, the interaction *Corpus*IdSoC* is mainly caused by significantly fewer *Speaker SoC* (s.r. = −2.8) yet more *Character SoC* (s.r. = 3.8) in the corpus of *Newspaper*; in *Microblog*, by contrast, there are more *Speaker SoC* (s.r. = 2.0) yet slightly less *Character SoC* (s.r. = −1.8) and *No SoC* (s.r. = −1.7); *Character SoC* in *Conversation* is less frequent than expected (s.r. = −2.0). See Examples (28) and (29) respectively for typical fragments with *Speaker SoC* "I" in microblog posts and *Character SoC* "people engaged in corruption" in newspaper texts.

Table 10. Co-occurring of each *IdSoC* type and connective (frequencies and standardized residuals)

Connective		**Identity of *SoC***			
		Speaker	**Character**	**No *SoC***	**Total**
jìrán	Count	132	17	1	150
	Std. Res.	4.6	−1.6	−5.8	
yīnwèi	Count	95	29	26	150
	Std. Res.	.6	.8	−1.7	
yóuyú	Count	40	29	81	150
	Std. Res.	−5.2	.8	7.5	
Total		**267**	**75**	**108**	**450**

The interaction *Connective*IdSoC* is well demonstrated by the high frequency of *jìrán* co-occurring with *Speaker SoC* (s.r. = 4.6) but very low frequency with *No SoC* (s.r. = −5.8), and from very high frequency of *yóuyú* with *No SoC* (s.r. = 7.5) and low frequency with *Speaker SoC* (s.r. = −5.2).

Fragments (27) and (28) are both typical use of *jìrán* with *Speaker SoC*. The only difference is that the SoC remains implicit in (27) but is explicit (in the accusative case "me") in (28). Example (30) illustrates a *yóuyú* instance without SoC involved. The causal relation between "(her) contact information has long been deleted before" and "now I only have her friends" holds regardless of the intentionality of either the speaker "I" or anyone else.

4.4 Linguistic realization of SoC

Data with respect to the last subjectivity feature, *LingReal-SoC*, are summarized in Figure 7. Although in the newspaper corpus, three cells have expected counts less than 5, which marginally violates assumptions in the chi-square test and log-linear analysis (Field 2011:692, 710), for the sake of comparability, we still used the log-linear procedure.

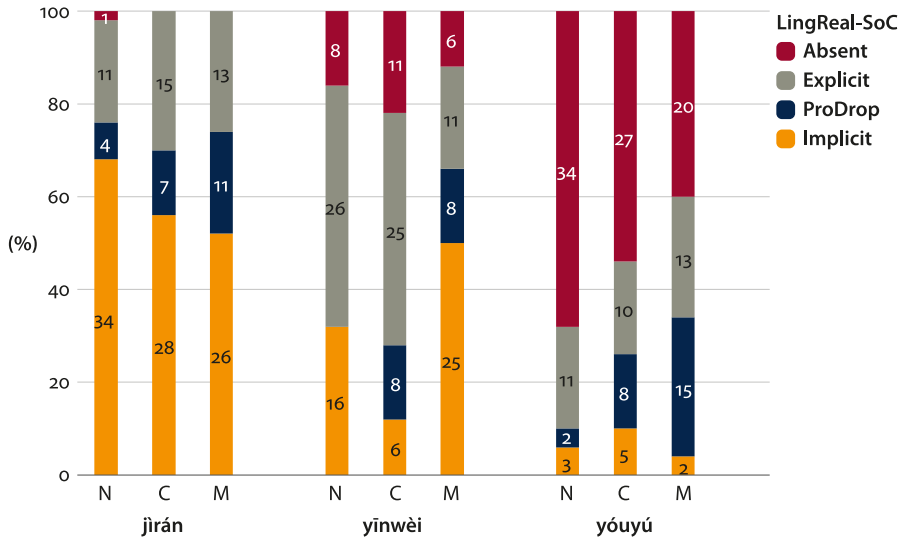

Figure 7. Frequencies and percentages of each LingReal-SoC category co-occuring with *jìrán, yīnwèi*, and *yóuyú* in Newspaper (N), Conversation (C) and Microblog (M)

The log-linear analysis generated a model containing a significant three-way interaction between *Connective, LingReal-SoC*, and *Corpus*, $\chi 2$ (12, N = 450) = 25.37,

$p = .01$. In other words, the overall pattern of *LingReal-SoC* co-occurring with connectives differs across corpora. Details can be seen in Table 11.

Table 11. Co-occurring of *LingReal-SoC* categories and connective in each corpus (frequencies and standardized residuals)

		LingReal-SoC								
		Implict		**Pro-drop**		**Explicit**		**Absent**		
		Count	**s.r.**	**Count**	**s.r.**	**Count**	**s.r.**	**Count**	**s.r.**	
	jìrán	34	3.9	4	1.4	11	−1.3	1	−3.5	
Newspaper*	*yīnwèi*	16	.4	0	−1.4	26	2.5	8	−1.7	
	yóuyú	3	−3.5	2	.0	11	−1.3	34	5.2	
Total		**53**		**6**		**48**		**43**		**150**
	jìrán	28	4.2	7	−.2	15	−.4	0	−3.6	
Conversation**	*yīnwèi*	6	−1.9	8	.1	25	2.0	11	−.5	
	yóuyú	5	−2.2	8	.1	10	−1.6	27	4.0	
Total		**39**		**23**		**50**		**38**		**150**
	jìrán	26	2.0	11	−.1	13	−.2	0	−2.9	
Microblog***	*yīnwèi*	25	1.7	8	−1.0	11	−.4	6	−.9	
	yóuyú	2	−3.7	15	1.1	13	.2	20	3.8	
Total		**53**		**34**		**37**		**26**		**150**

* χ^2 (6, N = 150) = 82.99 ** χ^2 (6, N = 150) = 62.19, $p < .001$ *** χ^2 (6, N = 150) = 47.57, $p < .001$

Seen from the frequencies and the standardized residuals, *LingReal-SoC* with regard to *jìrán* is not sensitive to discourse types: *jìrán* co-occurs predominantly with *Implicit* in all three corpora; the three-way interaction mainly comes from the different behaviors of *yīnwèi* and *yóuyú* across corpora.

The relevant category involved in *yóuyú* is *Pro-dro*p in particular, the frequency of which is relatively low in the newspaper corpus (count = 2) while high in microblogs (count = 15). Example (18), repeated here as (33), illustrates a *yóuyú* fragment in microblogs having a *null* subject "I", the SoC responsible for the intentional act "(I) put on an extra jacket […]".

The categories involved in *yīnwèi* are *Implicit* and *Explicit*. Relatively few implicit SoCs were found in the corpus of *Conversation* (count = 6; s.r. = −1.9), while relatively many occur in *Microblog* (count = 25; s.r. = 1.7). The occurring of *Explicit*, however, is the opposite: mainly in *Conversation* (count = 25; s.r. = 2.0) rather than in *Microblog* (count = 11, s.r. = −.4). Numerous explicit cases were found in *Newspaper*, too (count = 26; s.r. = 2.5), yet no occurrence of *Pro-drop* (count = 0). Example (26), repeated here as (34), illustrates a typical use of *yīnwèi* having an implicit SoC in microblog posts: the blogger suggests "please smile while picking up the phone" because "people on the other side can feel it", but he/

she remains implicit in the utterance. Example (24), repeated below as (35), illustrates an explicit SoC co-occurring with *yīnwèi* in conversational exchanges.

(33) ***Yóuyú*** *P huódòng chǎngdì kàojìn hǎibiān, wǎnshàng tèbié lěng,* ***suǒyǐ*** *Q chūmén de shíhòu Ø duō jiā le jiàn yīshang.*
Yóuyú *P* the activity venue was nearby the sea, it would be very cold at night, *suǒyǐ* 'so' *Q* (I) put on an extra jacket while leaving home (for the activity).

(34) *Q Nàqǐ diànhuà de shíhòu qǐng wéixiào,* ***yīnwèi*** *P duìfāng néng gǎnjuédào!*
Q Please smile while picking up the phone, ***yīnwèi*** *P* people on the other side can feel it!

(35) *Yīnwèi P zuòwéi xuéshēng Ø méiyǒu zìjide diànshì, suǒyǐ Q néng zà nǎ kàn (NBA) wǒ jiù zàinǎ kàn.*
Yīnwèi *P as a student, (I) don't have my own TV, suǒyǐ 'so' Q I then watch (NBA) wherever I can.*

4.5 Forward-linking and backward-linking *yīnwèi*

Yīnwèi stands out among the three connectives for often linking both the forward and the backward causal relations. Presented in Figure 8 is the distribution of each subjectivity feature in the two orders.

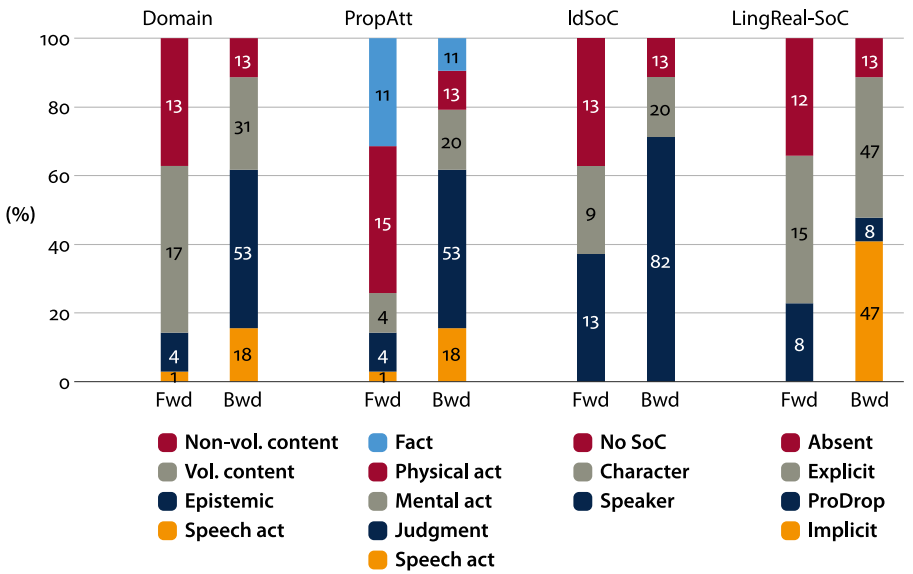

Figure 8. Frequencies and percentages of categories of each feature co-occuring with *Fwd/Bwd yīnwèi*

The log-linear analyses for the relationship between causality order, subjectivity, and corpus are summarized in Table 12. For each of the four subjectivity features, we obtained a model containing three sets of two-way interactions. The fit for each model is adequate ($p \geq 0.38$).

Table 12. Yīnwèi: Relationships between corpus, subjectivity and causality order

Effects	Chi-Square	df	Sig.	Goodness-of-Fit (Pearson)
Corpus*Domain	17.01	6	.009	
CausalityOrder*Corpus	13.32	2	.001	$\chi^2(6, N=150) = 6.45, p = .38$
CausalityOrder*Domain	23.15	3	.000	
Corpus*PropAtt	17.71	8	.024	
Causalityorder*Corpus	11.90	2	.003	$\chi^2(8, N=150) = 8.29, p = .41$
Causalityorder*PropAtt	29.37	4	.000	
Corpus*IdSoC	15.85	4	.003	
CausalityOrder*Corpus	17.97	2	.000	$\chi^2(4, N=150) = 4.20, p = .38$
CausalityOrder*IdSoC	15.23	2	.000	
Corpus*LingReal-SoC	26.14	6	.000	
CausalityOrder*Corpus	9.47	2	.009	$\chi^2(6, N=150) = 3.96, p = .68$
CausalityOrder*LingReal-SoC	29.13	3	.000	

Directly relevant to RQ3 are the two-way interactions between the causality order and each subjectivity feature. They show that the subjectivity characteristics of *yīnwèi* used in the two positions differ significantly from each other. The lack of three-way interactions demonstrates that this relationship is not dependent on *corpus*. A cross-table analysis on the interaction between *Fwd/Bwd yīnwèi's* and subjectivity is singled out and presented in Table 13.

Table 13. Co-occurring of *Fwd/Bwd yīnwèi* and each category of the four subjectivity features (frequencies and standardized residuals)

		Fwd yīnwèi		*Bwd yīnwèi*		
		Count	**Std. Res.**	**Count**	**Std. Res.**	**Total**
Domain	Speech act	1	−1.6	18	.9	150
	Epistemic	4	−2.6	53	1.4	
	Vol. content	17	1.7	31	−1.0	
	Non-vol. content	13	2.8	13	−1.6	
PropAtt	Speech act	1	−1.6	18	.9	150
	Judgment	4	−2.6	53	1.4	
	Mental act	4	−.7	20	.4	
	Physical act	15	3.3	13	−1.8	
	Fact	11	2.6	11	−1.4	

Table 13. *(continued)*

		Fwd yīnwèi		*Bwd yīnwèi*		
		Count	**Std. Res.**	**Count**	**Std. Res.**	**Total**
IdSoC	Speaker	13	−1.9	82	1.1	150
	Character	9	.8	20	−.5	
	No SoC	13	2.8	13	−1,6	
LingReal-SoC	Implicit	0	−3.3	47	1.8	150
	Pro-drop	8	2.2	8	−1.2	
	Explicit	15	.1	47	.1	
	Absent	12	2.6	13	−1.4	

Fwd yīnwèi displays a clear objective profile in that it combines significantly more often with the objective categories of each feature, i.e. *Non-vol. content* (s.r. = 2.8), *Physical act* (s.r. = 3.3), *Fact* (s.r. = 2.6), and *No SoC* (s.r. = 2.8). By contrast, *Bwd yīnwèi* has, though not significantly, a slight preference for the more subjective categories: *Epistemic* and *Judgment* (both, s.r. = 1.4), a large number of *Speaker SoC* (82 out 150, s.r. = 1.1) remaining *Implicit* (s.r. = 1.8); however, it rarely combines with the objective categories like *Physical act* (s.r. = −1.8), *Non-vol. content* or *No SoC* (both, s.r. = −1.6).

Example (36) below illustrates the typical use of *Fwd yīnwèi* signaling a non-volitional content relation: the fact that "(they) live close to each other" leads to the fact that "(they) often encounter on the way to market"; the Q-segment "(they) often encounter on the way to market" is a physical fact; and there is no need to refer to an SoC for the interpretation of this causal relation. (31), repeated here as (37), illustrates *Bwd yīnwèi* signaling an epistemic relation: the fact that "you know everything ([…])" leads to the conclusion "[…], you should be able to […]", which is perceived by the speaker who is put implicit in the utterance.

(36) ***Yīnwèi*** *P Ø zhùde jìn, Q Ø chánghuì zài mǎicài lùshàng xiāngyù.*
Yīnwèi *P* (they) live close to each other, *Q* (they) often encounter (each other) on the way to market.

(37) *Q Rúguǒ nǐ yǒu- yǒu zhè yībǎi wàn, Ø yìng gāi néng zài zìjǐ de zǔguó shēnghuó dé gèng- bǐ dào wàiguó qù- qù gèng hǎo.* ***Yīnwèi*** *P nǐ shénme dōngxī dū zhīdào.*
Q If you have, have a million (Rmb), (you) should be able to live a better, compared to living abroad, better life in your home country. ***Yīnwèi*** *P* you know everything (in your home country).

Another significant interaction that catches our attention is that between *causality order* and *corpus*. Previous studies by Biq (1995) and Song & Tao (2009) both have found different distributions of the two sequences in written *vs* spoken dis-

course. For a close look at the results from our data, we made a cross- table analysis (Table 14) to illustrate this specific interaction.

Table 14. Distribution of *Fwd/Bwd yīnwèi* in each corpus (frequencies, percentages and standardized residuals)

	Fwd yīnwèi			*Bwd yīnwèi*			**Total**
	Count	**Percentage**	**Std. Res.**	**Count**	**Percentage**	**Std. Res.**	
Newspaper	8	16%	−1.1	42	84%	.6	100%
Conversation	22	44%	3.0	28	56%	−1.7	100%
Microblog	5	10%	−2.0	45	90%	1.1	100%

Pearson χ^2 (2, N = 150) = 18.41, p < .001

Across the three corpora, backward sequence occurs more frequently than forward sequence. The overwhelming predominance was found in the corpora of *Newspaper* and *Microblog*. In *Conversation*, however, the occurrence of forward sequence is significantly more than expected (s.r. = 3.0), a typical instance of which is (36) above. Below, (38) exemplifies a backward sequence in microblogs (90%).

(38) *Q Zhíchǎng yǒngyuǎn bùhuì tàipíng,* ***yīnwéi*** *P rénxīn yǒngyuǎn dōu bùhuì tàipíng.*
Q There will never be peace in career, ***yīnwéi*** *P* there will never be peace in people's mind.

4.6 The relative importance of subjectivity features

To measure the importance of the four subjectivity features, together with the causality order, in predicting the use of *jìrán, yīnwèi*, or *yóuyú*, we carried out a conditional inference trees analysis and random forests analysis.[9] These are non-parametric tree-structure models of regression and classification, with the advantage of unbiased predictor selection and the returned p values presenting the respective confidence level (Levshina 2015: 291–300).

Figure 9 is the conditional inference tree that represents the decision rules determining the use of *jìrán, yīnwèi* or *yóuyú*. Splits were made upon the random selection of the pre-specified predictors. *Causality order* (Node 1) proved to have the strongest association with the split: backward causality directly predicted almost only *yīnwèi* (Node 9). Among the forward structures, *PropAtt* (Node 2) is the most important predictor. If it is *Speech act* or *Judgment*, the predicted con-

9. Both analyses were enabled by R-package "party" (Levshina 2015: 291–300).

nective is predominantly *jìrán* (about 85% of the 147 fragments, Node 3). If it is *Physical act* or *Fact*, the *IdSoC* (Node 6) plays a role for a finer distinction. In the case of *Speaker SoC* or *No SoC*, seen in Node 7, *yóuyú* accounts for 80% of the 126 cases, while there is hardly any *jìrán*. *Character SoC*, as shown in Node 8, predicts mainly *yóuyú (60%)*, but also *jìrán* and *yīnwèi* (both roughly 20%). *Mental act* predicts moderately equal chances of *jìrán* and *yóuyú* (both slightly over 40%, Node 5). No impact of *domain* or the *LingReal-SoC* was detected. The predictive value of this model is substantial ($C = 0.84$).

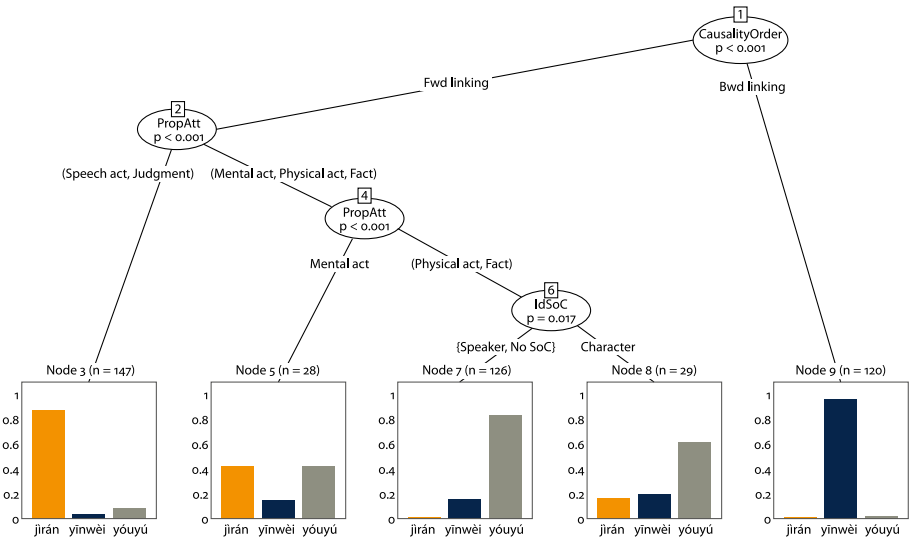

Figure 9. Conditional inference tree (the outcome of the binary recursive partitioning)

Figure 10 is the outcome of the random forests analysis representing the conditional impact of variables on the connectives' profiles. In our case, *Causality order* has the strongest impact, followed in turn by *PropAtt* and *Domain*. The *IdSoC* and *LingReal-SoC* have very weak impact on the predictions. The predictions of this model are substantial as well ($C = 0.85$).

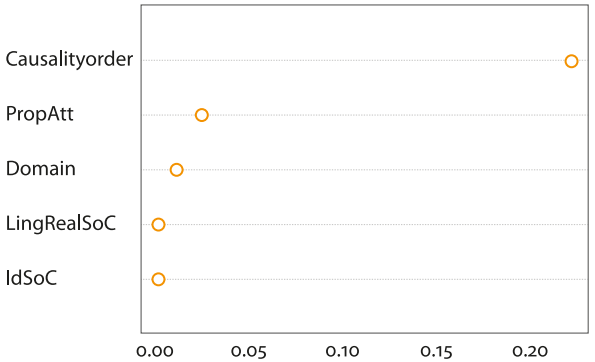

Figure 10. Conditional importance of variable

5. Summary of results and discussion

The statistical analysis consisted of two parts. First, we carried out a general log-linear analysis for each subjectivity feature, which provided us with the best-fit model for the relationship between *connective, subjectivity* and *corpus*; the follow-up crosstable analyses enabled us to examine the interactions contributing to the resulting models. Second, to measure the relative importance of the independent factors, i.e. the four subjectivity features and the causality order, in predicting the *meaning and use* of one connective other than another, we employed the conditional inference trees and random forests analyses. § 5.1 through § 5.3 present the summary of results and a discussion of research questions 1 to 4.

5.1 Research questions one and two

The contextual difference typical of discourse in each corpus is, to a large extent, supported by our data. More specifically, it follows the significant two-way interactions between *corpus* and *domain, PropAtt,* and *IdSoC,* respectively. Interestingly, we did not find a corpus effect on the interactions between *connective* and *domain/PropAtt/IdSoC*. This strongly supports the analysis that these three subjectivity characteristics are intrinsic to *jìrán, yīnwèi*, and *yóuyú*. Analysis of the standardized residuals showed that across corpora, (1) *jìrán* systematically and prototypically signals relations of the subjective domains, i.e. *speech act* and *epistemic* relations, in which Q segments are predominantly *speech act* and *judgment*; the subjective causal relations and propositional attitudes are conceptualized predominantly by *speaker SoC*. (2) *Yóuyú*, by contrast, is found to systematically and prototypically signal relations in the objective domains, i.e. *volitional* and *non-*

volitional content relations, generally involving *no SoC* and an extremely high amount of Q segments expressing *physical act* and *fact*. (3) *Yīnwèi* co-occurs roughly equally with both the subjective and objective relations, with a slight inclination to *epistemic* and *volitional content* domains, to propositional attitude of *judgment, mental act* and *fact,* and to *speaker SoC*.

The interaction between *connective* and *LingReal-SoC,* however, appears sensitive to differences between corpora. This is definitely worth special attention, because it is directly related to research question 2: does the semantic-pragmatic profile of a given connective remain robust or does it vary across corpora?

It might be possible that this result is an accidental issue due to the low expected counts in three cells of *jìrán* and *yóuyú* (violation of log linear assumptions, § 4.4). However, genre effect on the interaction between *connective* and *LingReal-SoC* was also found in Li et al. (2016: 31) using written discourse, which adds to our confidence that this particular feature associated with the connectives under analysis tends to be influenced by the context that they are used in.

The standardized residuals in Table 11 (§ 4.4) provide us with insight into the specific interactions. *Jìrán* co-occurs predominantly with *implicit SoC* in all three corpora, which corresponds with the subjective profile formulated by the interactions between *jìrán* and the other three aspects *domain/PropAtt/ IdSoC*. As for *yóuyú*, it is *pro*-drop that appears sensitive to discourse types (rarely occurring in newspapers yet relatively often in microblogs), which might result from the formal-informal style contrast between newspaper and microblog discourse. To our knowledge, no literature has yet associated *pro*-drop in Chinese with subjectivity. Its influence on the subjectivity profile of *yóuyú* remains unclear, which might be a topic worth exploring in the future. Nevertheless, the most objective category, i.e. *absent SoC*, takes up the great majority of the occurrences in each corpus, which confirms the objective profile of *yóuyú* as established with regard to *domain/PropAtt/IdSoC*.

Things become more complicated with *yīnwèi*: the frequency of *implicit SoC* is relatively high in microblogs and low in spontaneous conversations, whereas the pattern with *explicit SoC* is just the opposite. In other words, the construal of SoC appears more subjective via microblogging than via oral communication. Note that this agrees with our findings that the microblog texts tend to have relatively more subjective characteristics than the conversations (see Table 5, Table 7, and Table 9 for the distribution of *domain/PropAtt/IdSoC* across corpora). Seen from our data, causal relations constructed in conversations do feature the description of everyday experience, as can be seen in (32), repeated here as (39). Bloggers, while constructing causality, seem to aim at conveying their personal opinions, which gives rise to a more subjective context characteristic of argumentation, e.g. (38), repeated here as (40).

(39) *Āiyā, nèitiān zǎoshang,* ***yīnwèi*** *P tóuyītiān wǎnshàng wǒ chīde tàiduōle, chī bùxià, suǒyǐ Q Ø chīle xiāngjiāo qīng qīng chángzi.*
Oops! That morning, ***yīnwèi*** *P* the night before that morning, I had too much food and couldn't eat any more next morning, *suǒyǐ 'so' Q* (I) first had a banana for digestion.

(40) *Q Zhíchǎng yǒngyuǎn bùhuì tàipíng,* ***yīnwéi*** *P rénxīn yǒngyuǎn dōu bùhuì tàipíng.*
Q There will never be peace in career, ***yīnwéi*** *P* there will never be peace in people's mind.

The above results and discussion provide answers to research questions 1 and 2. First, *jìrán, yīnwèi*, and *yóuyú* are distinctively categorized in terms of subjectivity: *jìrán* is the most subjective connective, *yóuyú* the least subjective; *yīnwèi* is more general. Similar as its counterpart *because* in English, *yīnwèi* has a hypernym-status as a causal connective (Knott & Sanders 1998), capable of marking causal relations that are prototypically expressed by either *jìrán* or *yóuyú*. Second, the prototypical subjective profile of *jìrán* and objective profile of *yóuyú* are robust across discourse of either newspaper, spontaneous conversation, or microblog. *Yīnwèi* is basically neutral across all the three discourse types, especially with respect to the features of domain, *PropAtt* and the identity of SoC, although the linguistic construal of SoC seems to be more subjective (implicit) in microblog texts while less subjective (explicit) in spoken and written texts.

5.2 Research question three

Zooming in on *yīnwèi,* we analyzed the subjectivity features of *Fwd yīnwèi* and *Bwd yīnwèi* separately. It was found that the "*two*" *yīnwèi's* differ systemically from each other with respect to each feature and that the distinctions are all independent of corpus variation. Specifically, *Fwd yīnwèi* proves to be very objective, signaling prototypically *volitional* and *non-volitional content* relations with dominant Q segments of *physical act* and *fact*. The SoC (either speaker or character) is *explicitly* referred to or is simply *absent (no SoC)* in the consequent. *Bwd yīnwèi*, by contrast, shows a tendency to co-occur with more subjective categories: *epistemic* relations and Q segments expressing *judgment* which are conceptualized predominantly by *speaker SoC,* either *implicit* or *explicit*.

Moreover, we also found a significant distributional variation of *Fwd/Bwd* causal constructions across corpora. In all three corpora, backward sequence "*Q, yīnwèi P*" is more frequent, which conforms to the reports in literature using written and spoken data (Biq 1995; Wang 2002; Song & Tao 2009).

Unlike the above-mentioned studies, the extreme inclination towards backward construction was found in our newspaper and (especially) microblog corpora, rather than in the corpus of conversations over telephone and TV talk-show programs. We found many instances of *"yīnwèi P, Q"* in conversation data (44%, s.r. = 3.0). This, however, could be explained in line with Young's (1994) argument that the reasoning model of "*yīnwèi P, (suǒyǐ 'so') Q*" in *conversational exchanges* comes naturally and is rational to Chinese people. From the socio-cultural perspective, this holistic attention to causality, in other words, the preference for inductive reasoning from premise to conclusion, e.g. (39), is considered more collaborative and considerate. The analytic causality, where conclusions are directly imposed upon the addressee, e.g. (40), could be a *face-threatening act* (Brown & Levinson 1987) to both the speaker and the addressee. Another possible concern of this premise-first preference might be that, by attributing causality to the circumstances first, speakers avoid taking much personal agency. This might also explain why the frequency of *implicit SoC* is considerably low in the conversation corpus while high in microblogs (Table 11). The non-face-to-face communication via blogging could save, to some extent, the blogger's concern of expressing personal opinions in a more subjective way (without putting him-/herself *on the stage* to be argued for or against).

5.3 Research question four

The analyses of conditional inference trees and random forests provided a clear picture of the importance status of the five independent factors. Specifically, causality order is the most effective feature in the model. However, since we have barely any case of backward constructions signaled with *jìrán* or *yóuyú* in our data, this effect applies only to *yīnwèi*. This has been confirmed by the distinctive profile of *yīnwèi* defined in the two causality orders. *PropAtt* is the second important feature. *Speech act* or *judgment* predict predominantly the occurrence of *jìrán*, whereas *physical act* or *fact* predict that of *yóuyú*. *IdSoC* comes in the third place. Among the Q segments that express *physical act* or *fact*, when the *Speaker SoC* or *No SoC* is involved, the connective observed is substantially *yóuyú*; when it is *Character SoC*, the observations contain mainly *yóuyú*, but also some *jìrán* and *yīnwèi*. The predicted results have reflected, to a large extent, the profiles of connectives generated through log-linear analyses. Given the fact that each feature is selected automatically by the program, the subjectivity features defined in the present study (and the previous studies alike) are reliably justified in their significance of categorizing causal connectives.

6. Conclusion

In this article, we aim to study the cognitive system underlying people's categorization of causal relations in Mandarin discourse by focusing on the *semantic-pragmatic* profile (the *meaning and use*) of Mandarin REASON connectives *jìrán, yīnwèi*, and *yóuyú*. Through systematic statistical analysis, we succeed in categorizing the connectives using an integrated subjectivity account, that is, through analyzing the degree of subjectivity encoded in a causal construction in terms of four features. The modification we have made to the existing model of analysis (Li et al. 2016), specifically, a finer distinction of categories within the features of *PropAtt* and *LingReal-SoC,* enabled us to evaluate precisely the subjectivity profile of each connective. More importantly, our statistical analysis is built upon systematically selected discourse that is as distinct as printed newspapers, spontaneous conversations, and microblog posts. This range of discourse types provide a solid basis for investigating the subjectivity profiles of the connectives. Furthermore, the relative strength of *domain, PropAtt, IdSoC, LingReal-SoC,* and *causality order* in distinguishing one connective from another is also measured. From both a theoretical and a methodological perspective, our study adds Mandarin Chinese, which is typologically different from European languages, to the database of languages that demonstrate the significance of the integrated account for the cognitive mechanism underlying the coherence relations in discourse, specifically here, causal coherence.

This study is also prone to some limitations. First, the objectivity bias we have taken for ambiguous cases contributes to the reliability of the analyses, but it may affect their validity (over-coding of objective features). Nonetheless, as the total number of ambiguous cases is rather low (for details, see § 3.3), we believe the bias will not have a significant effect on the results. Secondly, following our coding rules, there is a high correlation between a relation domain and the corresponding *PropAtt*. The correlation only breaks down in the case of *volitional content domain* in which the *PropAtt* is either *mental act* or *physical act*, both involving the *intentionality* of SoC. This explains why *domain* is not present in the conditional inference trees, whereas *PropAtt* has displayed great importance in the model. An interesting issue for follow-up research is whether both *domain* and *PropAtt* are needed as separate factors in a parsimonious but cognitively plausible account of causal coherence.

The present study has generated some topics that are worth studying in the future. It might be interesting to further explore the possible pragmatic and cognitive implications underlying the linguistic device *pro*-drop in discourse and the linguistic reference to SoC (implicit vs explicit) in *yīnwèi* fragments which varies across our corpora. Last, our explanation for the relatively frequent occurrences of forward *yīnwèi* construction in our spoken data, which is now stated in socio-

cultural terms, is tentative at best; systematically selected and more varied spoken data resources are definitely needed for a clearer discussion of this topic.

Acknowledgements

This research is supported by Fujian province social science planning project (FJ2016B258) awarded to Hongling Xiao. We thank the editors and two anonymous reviewers for their constructive suggestions on a previous version of this paper. Part of the content in this paper has been presented in the 7th International Conference of BeNeCLA in Nijmegen (2017), the 14th International Cognitive Linguistics Conference (ICLC-14) of ICLA in Tartu (2017), and the workshop "Perspectives on Subjectivity" in Utrecht University (2018). We thank the colleagues at these meetings for their comments. Needless to say, all remaining errors and shortcomings are ours.

Appendix. Relative frequency of *jìrán, yīnwèi* and *yóuyú* across corpora

The table below presents the relative frequencies of *jìrán, yīnwèi* and *yóuyú* occurring in the corpus of newspaper, spontaneous conversations and microblog discourse (per million characters). Analysis of the standardized residuals shows that *yóuyú* is typical in newspaper articles and atypical in spontaneous conversations (on telephone and TV programs); whereas for *yīnwèi*, which overall is the most frequently used connective, the opposite holds.[10] *Jìrán* is also atypical in newspapers. In microblogs, none of these connectives is found to occur significantly more or less than expected.

The occurrence of *jìrán, yīnwèi* and *yóuyú* in each corpus per million characters (relative frequencies and standardized residuals)

		Jìrán	*Yīnwèi*	*Yóuyú*
Newspaper*	Count	10.25	145.47	228.98
	Std. Res.	−2.53	−9.58	26.64
Conversation *&**	Count	65.93	1640.76	49.28
	Std. Res.	1.26	4.09	−11.50
Microblog***	Count	17.62	276.23	27.37
	Std. Res.	−0.21	0.78	−1.87

χ^2 (4) = 962.42, $p < .001$
* from Linguistic Data Consortium (LDC)
** from Media Language Corpus (MLC)
*** from BCC

10. These frequencies are based on string searches in the corpora; the frequencies therefore disregard the fact that *yīnwèi* and *yóuyú* can also be used as prepositions. For a brief view of the frequencies of the *connective* vs *prepositional uses* of *yīnwèi* and *yóuyú*, please refer to § 3.1 Corpus and data collection.

References

Austin, John L. 1962. *How to do things with words*. Oxford: Oxford University Press.

Benveniste, Émile. 1971. *Problems in general linguistics*. Coral Gables: University of Miami Press.

Biq, Yung-O. 1988. From focus in proposition to focus in speech situation: *cai* and *jiu* in Mandarin Chinese. *Journal of Chinese Linguistics* 16(1). 72–108.

Biq, Yung-O. 1995. Chinese causal sequencing and *yīnwèi* in conversation and press reportage. In Bilmes, Leela & Liang, Anita C. & Ostapirat, Weera (eds.), *Proceedings of the twenty-first annual meeting of the Berkeley Linguistics Society: Special session on discourse in south-east Asian languages*. 47–60. Berkeley: Berkeley Linguistics Society.

Brown, Penelope, & Levinson, Stephan C. 1987. *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.

Chafe, Wallace. 1982. Integration and involvement in speaking, writing, and oral literature. In Tannen, Deborah (ed.), *Spoken and written language: Exploring orality and literacy*, vol. 6, 35–53. Norwood: ABLEX Publishing Corporation.

Chafe, Wallace. 1984. Integration and involvement in spoken and written language. In Borbe, Tasso (ed.), *Semiotics unfolding: Proceedings of the second Congress of the International Association for Semiotic Studies, Vienna, July 1979*, 1095–1102. Berlin: Mouton.

Degand, Liesbeth, & Pander Maat, Henk. 2003. A contrastive study of Dutch and French causal connectives on the Speaker Involvement Scale. In Verhagen, Arie & van de Weijer, Jeroen (eds.), *Usage-based approaches to Dutch: Lexicon, grammar, discourse*, 175–199. Utrecht: LOT, Netherlands Graduate School of Linguistics.

De Smet, Hendrik & Verstraete, Jean-Christophe. 2006. Coming to terms with subjectivity. *Cognitive Linguistics* 17(3). 365–392.

Ducrot, Oswald. 1980. Pragmatique linguistique: II. Essai d'application: Mais–les allusions à l'énonciation–délocutifs, performatifs, discours indirect [Application test: BUT –allusions to the utterance –talkative, performative, indirect speech]. In Parret, Herman (ed.), *Le langage en context: Etudes philosophiques et linguistiques de pragmatique* [Language in context: Philosophical and linguistic studies of pragmatics], 487–575. Amsterdam: John Benjamins.

Eifring, Halvor. 1995. *Clause combination in Chinese*. Leiden: E.J.Brill.

Evers-Vermeul, Jacqueline. 2005. *The development of Dutch connectives: Change and acquisition as windows on form-function relations*. Utrecht: LOT.

Ferrara, Kathleen & Brunner, Hans & Whittemore, Greg. 1991. Interactive written discourse as an emergent register. *Written Communication* 8(1). 8–34.

Field, Andy. 2011. *Discovering statistics using SPSS: (and sex and drugs and rock 'n' roll)*. Los Angels: SAGE Publications.

Finegan, Edward. 1987. On the linguistic forms of prestige: Snobs and slobs using English. In Boardman, Phillip C. (ed.), *The legacy of language: A tribute to Charlton Laird*, 146–161. Reno: University of Nevada Press.

Fung, Pascale & Huang, Shudong & Graff, David. 2005. *LDC2005S15/T32*. Philadelphia: Linguistic Data Consortium (https://www.ldc.upenn.edu/) (Accessed 2015-12-3) (Web Download.)

Glenn, Meghan & Lee, Haejoong & Strassel, Stephanie & Maeda, Kazuaki. 2013–2015. *GALE Phase 3–4 Chinese Broadcast Conversation Transcripts LDC2013T08; 2014T28; 2015T09*. Philadelphia: Linguistic Data Consortium. (https://www.ldc.upenn.edu/) (Accessed 2015-12-1) (Web Download.)

Günthner, Susanne. 1993. "… weil–man kann es ja wissentschaftlich untersuchen" –Diskurspragmatische Aspekte der Wortstellung in WEIL-Sätzen ["… because –you can investigate it scientifically" –Discourse-pragmatic aspects of word order in WEIL sentences]. *Linguistische Berichte* [Linguistic Reports] 143. 37–55.

Guo, Jimao. 2008. "Yīnwèi suoyi" ju he "jìrán nàme" ju de chayi [A contrastive analysis between sentences with *yīnwèi* and *jìrán*]. *Hanyu Xuexi* [Chinese Language Learning] 2008(3). 22–29.

Herring, Susan C. 2011. Computer-mediated conversation, Part II: Introduction and overview. *Language@Internet* 8(2). (https://www.languageatinternet.org/articles/2011/Herring) (Accessed 2015-10-20).

Hole, Daniel P. 2004. *Focus and background marking in Mandarin Chinese: System and theory behind cái, jiù, dōu and yě*. London: RoutledgeCurzon.

Huang, Yan. 2006. Speech acts. In Brown, Keith. (ed.), *Encyclopedia of language and linguistics*, 2nd edn., 656–665. Amsterdam: Elsevier.

Huang, Wenlong. 1998. "Jiran p, jiu q" ju zhiyi [Questions on "Jiran p, jiu q" construction]. *Guizhou Shifan Daxue Xuebao (Shehui Kexue Ban)* [Journal of Guizhou Normal University (Social Sciences)] 1998(4). 87–91.

Jin, Lixin & Du, Jiajun. 2014. "Jiu" yu "cai" zhuguanliang duibi yanjiu [*Jiu* and *cai*: A contrastive study of subjective quantity]. *Yuyan Kexue* [Linguistic Sciences] 13(2). 140–153.

Kay, Paul. 1977. Language evolution and speech style. In Blount, Ben G. & Sanches, Mary (eds.), *Sociocultural dimensions of language change*, 21–33. New York: Academic Press.

Keller, Rudi. 1995. The epistemic weil. In Stein, Dieter & Wright, Susan (eds.), *Subjectivity and subjectivisation: Linguistic perspectives*, 16–30. Cambridge: Cambridge University Press.

Knott, Alistair, & Sanders, Ted. 1998. The classification of coherence relations and their linguistic markers: An exploration of two languages. *Journal of Pragmatics* 30(2). 135–175.

Langacker, Ronald. 1985. Observations and speculations on subjectivity. In Haiman, John (ed.), *Iconicity in syntax: Proceedings of a Symposium on Iconicity in Syntax, Stanford, June 24–26, 1983*, 109–150. Amsterdam: John Benjamins.

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

Levshina, Natalia. 2015. *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam: John Benjamins.

Li, Charles N. & Thompson, Sandra A. 1989. *Mandarin Chinese: A functional reference grammar*. Berckley: University of California Press.

Li, Fang & Evers-Vermeul, Jacqueline & Sanders, Ted. 2013. Subjectivity and result marking in Mandarin. *Chinese Language and Discourse* 4(1). 74–119.

Li, Fang & Sanders, Ted & Evers-Vermeul, Jacqueline. 2016. On the subjectivity of Mandarin REASON connectives: Robust profiles or genre-sensitivity? In Stukker, Ninke & Spooren, Wilbert & Steen, Gerard (eds.), *Genre in language, discourse and cognition*, 15–50. Berlin: De Gruyter Mounton.

Li, Jinxia. 2011. Lun "yóuyú" yu "yīnwèi" de chayi [On the differences between *yóuyú* and *yīnwèi*]. *Shijie Hanyu Jiaoxue* [Chinese Teaching in the World] 25(4). 490–496.

Li, Jinxia & Liu, Yun. 2004. "Yóuyú" yu "jìrán" de zhuguanxing chayi [The differences of *yóuyú* and *jìrán* in subjectivity]. *Zhongguo Yuwen* [Studies of the Chinese Language] 2004(2). 123–128.

Liu, Chuqun. 2002. "Yīnwèi" he "yóuyú" chayi chutan [A preliminary study on the difference between *yīnwèi* and *yóuyú*]. *Journal of Anhui Institute of Education* 20(1). 89–92.

Liu, Yuehua & Pan, Wenyu & Gu, Wei. 2001. *Shiyong xiandai Hanyu yufa (Zengding ben)* [A practical grammar of modern Chinese (Expanded edition)]. Beijing: The Commercial Press.

Lǚ, Shuxiang (ed.) 1999. *Xiandai Hanyu babai ci (Zengding ben)* [Eight hundred words of modern Chinese (Expanded edition)]. Beijing: The Commercial Press.

Lyons, John. 1977. *Semantics*. Cambridge: Cambridge University Press.

Lyons, John. 1982. Deixis and subjectivity: Loquor, ergo sum? In Jarvella, Robert J. & Klein, Wolfgang (eds.), *Speech, place, and action: Studies in deixis and related topics*, 101–124. Chichester: John Wiley & Sons.

Pander Maat, Henk, & Degand, Liesbeth. 2001. Scaling causal relations and connectives in terms of speaker involvement. *Cognitive Linguistics* 12(3). 211–245.

Pander Maat, Henk & Sanders, Ted. 2000. Domains of use or subjectivity? The distribution of three Dutch causal connectives explained. In Couper-Kuhlen, Elizabeth & Kortmann, Bernd (eds.), *Cause, condition, concession, contrast: Cognitive and discourse perspectives*, 57–82. Berlin: Mouton de Gruyter.

Pander Maat, Henk & Sanders, Ted. 2001. Subjectivity in causal connectives: An empirical study of language in use. *Cognitive Linguistics* 12(3). 247–273.

Sanders, Ted. 1997. Semantic and pragmatic sources of coherence: On the categorization of coherence relations in context. *Discourse Processes* 24(1). 119–147.

Sanders, Ted & Spooren, Wilbert P.M. & Noordman, Leo G.M. 1992. Toward a taxonomy of coherence relations. *Discourse Processes* 15(1). 1–35.

Sanders, Ted & Sanders, José & Sweetser, Eve. 2009. Causality, cognition and communication: a mental space analysis of subjectivity in causal connectives. In Sanders, Ted & Sweetser, Eve (eds.), *Causal categories in discourse and cognition*, 19–59. Berlin: Mouton de Gruyter.

Sanders, Ted & Spooren, Wilbert. 2009. Causal categories in discourse: Converging evidence from language use. In Sanders, Ted & Sweetser, Eve (eds.), *Causal categories in discourse and cognition*, 205–246. Berlin: Mouton de Gruyter.

Sanders, Ted & Spooren, Wilbert. 2013. Exceptions to rules: A qualitative analysis of backward causal connectives in Dutch naturalistic discourse. *Text & Talk* 33(3). 377–398.

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

Sanders, Ted & Sweetser, Eve (eds.). 2009. *Causal categories in discourse and cognition*. Berlin: Mouton de Gruyter.

Searle, John R. 1969. *Speech acts: An essay in the philosophy of language*. Cambridge: Cambridge University Press.

Shen, Jiaxuan. 2003. Fuju san yu "xing, zhi, yan" [Complex sentences in three conceptual domains: Acting, knowing, and uttering]. *Zhongguo Yuwen* [Studies of the Chinese Language] 2003(3). 195–204.

Song, Zuoyan & Tao, Hongyin. 2009. A unified account of causal clause sequences in Mandarin Chinese and its implications. *Studies in Language. International Journal sponsored by the Foundation "Foundations of Language"* 33(1). 69–102.

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

Stukker, Ninke & Sanders, Ted & Verhagen, Arie. 2008. Causality in verbs and in discourse connectives: Converging evidence of cross-level parallels in Dutch linguistic categorization. *Journal of Pragmatics* 40(7). 1296–1322.

Sweetser, Eve. 1990. *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure*. Cambridge: Cambridge University Press.

Tagliamonte, Sali & Baayen, Harald. 2012. Models, forests and trees of York English: *Was/were* variation as a case study for statistical practice. *Language Variation and Change* 24(2). 135–178.

Traugott, Elizabeth Closs. 1982. From propositional to textual and expressive meanings: Some semantic-pragmatic aspects of grammaticalization. In Lehmann, Winfred Philipp & Malkiel, Yakov (eds.), *Perspectives on historical linguistics*, 245–271. Amsterdam: John Benjamins.

Traugott, Elizabeth Closs. 1989. On the rise of epistemic meanings in English: An example of subjectification in semantic change. *Language* 65(1). 31–55.

Traugott, Elizabeth Closs. 1995. Subjectification in grammaticalisation. In Stein, Dieter & Wright, Susan (eds.), *Subjectivity and subjectivisation: Linguistic perspectives*, 31–54. Cambridge: Cambridge University Press.

Traugott, Elizabeth Closs. 2010. (Inter)subjectivity and (inter)subjectification: A reassessment. In Davidse, Kristin & Vandelanotte, Lieven & Cuyckens, Hubert (eds.), *Subjectification, intersubjectification and grammaticalization*, 29–74. Berlin: Mouton De Gruyter.

Van Dijk, Teun A. 1979. Pragmatic connectives. *Journal of Pragmatics* 3(5). 447–456.

Voiskounsky, Alexander E. 1997. Telelogue conversations. *Journal of Computer-Mediated Communication* 2(4). (https://doi.org/10.1111/j.1083-6101.1997.tb00194.x) (Accessed 2016-02-26.)

Walker, Kevin & Caruso, Christopher & Maeda, Kazuaki & DiPersio, Denise & Strassel, Stephanie. 2013–2015. *GALE Phase 3–4 Chinese Broadcast Conversation Speech LDC 2013S04; 2014S09; 2015S06*. Philadelphia: Linguistic Data Consortium. (https://www.ldc.upenn.edu/) (Accessed 2015-11-30) (DVD/Web Download.)

Wang, Chun-hui. 2015. Tianjianju zhong de "na/name" [*Na/name* in conditional sentences]. *Hanyu Xuexi* [Chinese Language Learning] 2015(2). 41–48.

Wang, Yufang. 2002. The preferred information sequences of adverbial linking in Mandarin Chinese discourse. *Text & Talk* 22(1). 141–172.

Wu, Zhibiao. 1995. *Mandarin Chinese News Text LDC95T13*. Philadelphia: Linguistic Data Consortium. (https://www.ldc.upenn.edu/) (Accessed 2015-10-06) (Web Download.)

Xing, Fuyi. 2001. *Hanyu fuju yanjiu* [A study of Chinese complex sentences]. Beijing: The Commercial Press.

Xing, Fuyi. 2002. "Yóuyú" de yuyi pianxiang bian [On the semantic preference of the pattern introduced by *yóuyú*]. *Zhongguo Yuwen* [Studies of the Chinesee Language] 2002(4). 337–342.

Xun, Endong, Rao, Gaoqi, Xiao, Xiaoyue & Zang, Jiaojiao. (2016). Da shuju beijingxia BCC yuliaoku de yanzhi [The construction of the BCC Corpus in the age of Big Data]. *Yuliaoku Yuyanxue* [Corpus Linguistics] 3(1), 93–109.

Young, Linda W.L. 1994. *Crosstalk and culture in Sino-American communication*. Cambridge: Cambridge University Press.

Yus, Francisco. 2011. *Cyberpragmatics: Internet-mediated communication in context*. Amsterdam: John Benjamins.
Zhao, Zongsa & Yao, Shuangyun. 2016. Cong yuti shijiao kan "yīnwèi" "yóuyú" de chayixing [On the differences between *yīnwèi* and *yóuyú* from medium perspective]. *Dangdai Xiucixue* [Contemporary Rherotic] 2016(1). 62–71.
Zhong, Xiaoyong & Zhang, Lin. 2013. "Jìrán" ju he "yīnwèi" ju zhuguanxing chayi tan [On the subjectivity differences between sentence *jìrán* and sentence *yīnwèi*]. *Hanyu Xuexi* [Chinese Language Learning] 2013(4). 35–40.
Zufferey, Sandrine. 2012. "Car, parce que, puisque" revisited: Three empirical studies on French causal connectives. *Journal of Pragmatics* 44(2). 138–153.

Authors' addresses

Hongling Xiao (corresponding author)
Centre for Language Studies
Radboud Universiteit Nijmegen
Erasmusplein 1
6525 HT Nijmegen
The Netherlands
h.xiao@let.ru.nl
xhling@xmu.edu.cn

Publication history

Date received: 22 December 2017
Date accepted: 30 September 2018