



# A Trade-off Relationship between Lexical and Prosodic Means in Expressing Subjective and Objective Causality: Evidence from English and Mandarin

Na Hu<sup>1</sup>, Aoju Chen<sup>1</sup>, Fang Li<sup>2</sup>, Hugo Quené<sup>1</sup>, Ted J.M. Sanders<sup>1</sup>

<sup>1</sup> Utrecht institute of Linguistics OTS, Utrecht University, Utrecht, the Netherlands

<sup>2</sup> Beijing International Studies University, Beijing, China

n.hu@uu.nl, aoju.chen@uu.nl, lifangbisu@126.com, h.quene@uu.nl, tjmsanders@uu.nl

## Abstract

Prosody carries a variety of information in spoken communication. However, how it is involved in expressing the structure of discourse remains less well understood. This study addressed this issue by investigating the use of prosody in expressing two types of causality: subjective and objective causality. Specifically, we explored a trade-off relationship between prosody and lexical means, i.e., specialized causal connectives. We hypothesized, based on several theories, that the prosodic distinctions between subjective and objective causality would be less salient when the two types of causality are expressed by specialized causal connectives (e.g., *kejian* (so) and *yushi* (so) in Mandarin) than by a general causal connective (e.g., *so* in English). We conducted two production experiments to test this hypothesis, with Experiment 1 investigating the prosodic realizations of subjective and objective causality in English and Experiment 2 in Mandarin. We used a self-designed dialogue task to elicit speech samples and adopted a Bayesian approach to evaluate the effects of subjectivity on several acoustic measurements. The results of these two experiments support our hypothesis.

**Index Terms:** causal relations, discourse, subjectivity, prosody, Mandarin, English

## 1. Introduction

Prosody conveys important information in speech communication. It marks information structure, conveys speakers' social-indexical identities, and reveals speakers' emotions and attitudes (see [1] and [2] for extensive reviews). However, whether and how prosody is involved in expressing the structure of discourse, particularly the relation between discourse units, remains less well understood. So far, we only know that prosody provides information about the location of an utterance in discourse and distinguishes between causally related and non-causally related utterances [3]. This paper aims to shed more light on this issue by investigating a trade-off relationship between prosody and lexical cues in expressing different types of causal relations.

## 2. Research background

### 2.1. Subjective vs. objective causality

Human minds distinguish (at least) two types of causality. One type describes the cause-consequence relationship between actual events taking place in the physical world (as in (1)). The other type involves the argument-claim relationship between real-world events and the opinions, evaluations, or conclusions of a thinking subject, be it a speaker or the

narrator in discourse (as in (2)) [4]. In the discourse literature, these two types of causality have been referred to as content versus epistemic [5], or semantic versus pragmatic relations [4], respectively. The current study follows [6] to refer to the first type of causality as objective causality and the second type as subjective causality.

(1) It was a lovely day so I went swimming.

(2) Their car is not there so they are not at home.

### 2.2. The role of causal connectives in expressing subjective and objective causality

In some languages of the world, subjective and objective causality is often expressed by specialized causal connectives. For example, in Mandarin, objective causality expressed in (1) is often expressed by the causal connective *yushi* (so), which is a specialized objective causal connective, whereas subjective causality expressed in (2) is commonly expressed by the connective word *kejian* (so), which is a typical subjective causal connective [7]. One crucial function of these specialized causal connectives is that they make the type of causality explicit. However, not all languages favor using such specialized causal connectives to express causality. A case in point is English, where subjective and objective causality is commonly expressed by a general causal connective, e.g., *so* (as seen in (1) and (2)) [8], which leaves the type of causality underspecified.

### 2.3. The role of prosody in expressing subjective and objective causality

Spoken language involves not only lexical information, but also non-lexical information, such as prosody, facial expressions, and gestures. Do speakers use these communicative means to distinguish between subjective and objective causality? The answer, according to several theoretical frameworks, might largely depend on whether or not these two types of causality are already specified by lexical means, e.g., specialized causal connectives. The Functional Hypothesis [9] proposes a trade-off relationship between prosodic means and morphosyntactic means in expressing meanings or functions. According to this theory, it is more necessary to use prosody to express meaning or function when lexical cues for the meaning or function are absent than when they are present. A similar idea is expressed in the Signal Hypothesis or the Uniform Information Density Hypothesis, both of which hold that speakers tend to maintain a constant rate of information transfer [10, 11]. The ambiguity at the lexical level gives the speaker room to disambiguate using prosodic means. In the case of expressing subjective and objective causality, these theories suggest that these two types of causality are more likely to be expressed by prosody when

they are expressed by a general causal connective (such as *so* in English) than by specialized causal connectives (such as *kejian* and *yushi* in Mandarin).

The discourse literature long presumes that subjective and objective causality expressed by a general causal connective would be produced as two intonation units while objective causality as one intonation unit [12, 5]. Subsequent research observed this pattern in everyday conversation data [13, 14]. However, these two studies do not use statistical methods to test their observations. Hence, the extent to which these observations can be generalized across speakers remains questionable. Moreover, no study tested whether prosody still plays a role when specialized causal connectives are in use.

## 2.4. The current study

The present study aims to explore the trade-off relationship between using specialized causal connectives (frequently used in Mandarin, absent in English) and "specialized" prosody, i.e., different prosodic patterns. We hypothesize, based on the theories mentioned previously, that the prosodic distinctions between subjective and objective causality are less salient when the two types of causality are expressed by specialized causal connectives than when they are expressed by a general causal connective. To test this hypothesis, we conducted two production experiments. Experiment 1 compared the prosodic realizations of subjective and objective causality in English, where subjective and objective causality are often expressed by a general causal connective, and Experiment 2 in Mandarin, where the two types of causality are often expressed by specialized causal connectives. Both experiments focus on causality in forward order, i.e., CAUSE *SO* CONSEQUENCE and ARGUMENT *SO* CLAIM.

# 3. Experiment 1

## 3.1. Method

### 3.1.1. Participants

Fifteen native speakers of American English (11 females, 4 males; age range: 22-25 years; mean age: 23 years) participated in the experiment. The participants were recruited from Utrecht University through its international office and the first author's social network. The participants all claimed to have normal hearing and speaking abilities and declared to have no acting experience. They received €10 after completing the experiment.

### 3.1.2. Materials

Thirty target items were designed in pairs, each pair containing one objective causality sentence (see (3)) and one subjective causality sentence (see (4)). The sentences were comparable in length and syntactic complexity.

- (3) [Jim got his nose pierced]<sub>C1</sub> [so he bled a lot]<sub>C2</sub>.  
 (4) [Jim got his nose pierced]<sub>C1</sub> [so he wants attention]<sub>C2</sub>.

For the two items in each pair, the first clauses (denoted in (3) and (4)) by a subscript "C1" were the same, introducing a real-world event (e.g., "Jim got his nose pierced."), and the second clauses (denoted by "C2") were different. In the objective condition, C2 stated the consequence of the event described in C1 (e.g., "He bled a lot" as in (3)); in the

subjective condition, it expressed an opinion regarding that event (e.g., "He wants attention" as in (4)). These items were validated by several native speakers of American English.

The target items were divided into two lists, each containing approximately the same number of items from both causal categories but only one item from each pair. To conceal the research goal, we added ten filler items to each list, the fillers all conveying concessive relations involving the adverb *however*. The order of the items on each list was randomized across participants.

### 3.1.3. Task

These items were elicited using a dialogue task, which created a naturalistic conversational setting reminiscent of daily conversations.

In the task, the participants first read PowerPoint slides (see Figure 1 for an example), on which three kinds of information were presented automatically in sequence at 10-second intervals. First, the participants saw a background story (texts in italics in Figure 1) and an image, which set up the conversational context and primed the participants for upcoming information [15]. After ten seconds, four short sentences appeared simultaneously on the slide, the second and third sentences being the first and the second clauses of a target item (with the connective removed). Finally, a green button appeared in the bottom-right corner, prompting the participants that the dialogue was to start with the experimenter asking questions.

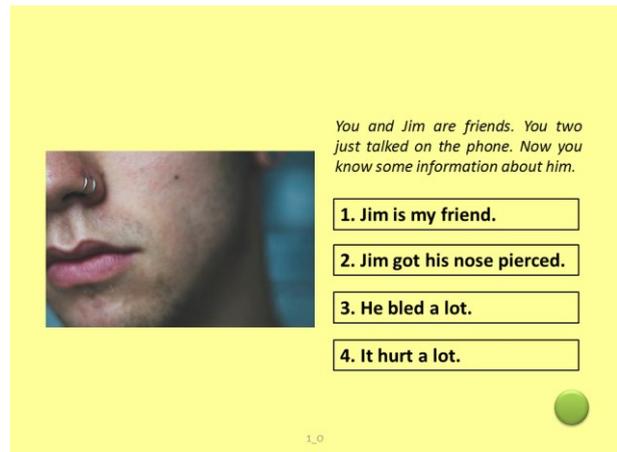


Figure 1: An example slide.

After reading each slide, the participants received three questions from the experimenter, the first question initiating the conversation, the second eliciting a target item (or a filler utterance expressing concession), and the third ending the conversation. In the subjective condition, the target question asked for an opinion on the topic under discussion. In contrast, in the objective condition, the second question only inquired about actual events.

Prior to the task, the participants received instructions on how to answer these questions. They needed to use the first and last sentences to answer the first and third questions, respectively, and the second and third sentences to answer the second question. They had to combine these two sentences into one sentence using *because* or *however*. The participants independently processed the relationship between these two sentences and received no feedback from the experimenter

regarding the correctness of the answer. They were strongly encouraged to treat the sentences as their own words and produce them as naturally as they would in a real conversation rather than reading off the screen.

### 3.1.4. Procedure

The participants were tested individually in a sound-treated booth in the phonetics lab at the Utrecht Institute of Linguistics OTS according to the following procedure.

First, the experimenter (the first author) informed the participant, without disclosing the real research goal, that the experiment collected natural speech samples for teaching purposes. Then, the experimenter introduced the interlocutor, a female native speaker of American English, to the participant. The participant was asked to treat the interlocutor as a friend and interact with her naturally during the experiment as if they were in real-life conversations. Third, the participant listened to the interlocutor explain the dialogue task, the design of the slides, and the instructions, and then they performed two practice trials, which were to familiarize them with the task.

The testing started when the participant was ready. The participant completed Lists 1 and 2 (or in reverse order) with a five-minute break in between. The conversations were recorded in a PCM (.WAV) audio format using a ZOOM 1 portable digital recorder (sampling rate of 44.1 kHz, 16 bit, stereo). The experiment took about 1 hour to complete.

### 3.1.5. Data annotation

The recordings obtained from the dialogue task were processed using Praat [16] as follows. First, the target utterances, i.e., the utterance elicited by the second question, were extracted, resulting in 450 utterances, four of which were discarded because they contained disfluencies. Second, each remaining utterance was annotated for the boundaries of three regions of interest: the first clause, the second clause, and the connective. For region-initial words starting with plosives, the initial boundary was set at the burst of the plosives. Third, the  $f_0$  contour of each utterance was inspected and octave jumps were manually corrected. Fourth, several prosodic measurements were extracted, including the  $f_0$  maximum and  $f_0$  minimum [in semitones (ST) relative to 1 Hz] in each clause, the duration of the connective (in seconds), and the duration of the pause (the silent interval) preceding the connective. We also calculated the speech rate of each clause, i.e., the number of syllables produced per second [17].

### 3.1.6. Statistical analysis

Data analysis was performed in R (version 3.6.3, [18]) using a Bayesian approach with the package `brms` [19], the wrapper package of the probabilistic programming language `Stan` [20]. Specifically, we evaluated the presence or absence of the effect of `subjectivity` on each acoustic measurement using Bayes factors. Bayes factors compare a full model ( $m_1$ ) containing the effect of `subjectivity` (contrast-coded, with subjective causality as +0.5 and objective causality as -0.5) with a null model ( $m_0$ ) not containing this effect. The analysis yields a value called  $BF_{10}$  (with the subscripts 1 and 0 referring to  $m_1$  and  $m_0$ , respectively), which indicates the extent to which the evidence supports  $m_1$  over  $m_0$ , akin to the ratio of likelihoods of the two models [21]. For evaluation purposes,  $BF_{10}$  values of 10-30, 3-10, 2-3, and close to 1

indicate strong evidence, weak evidence, very weak evidence, and no evidence supporting  $m_1$ , respectively; and  $BF_{10}$  values less than 0.1 indicate that the evidence is in favor of  $m_0$  [22]. Because the  $BF_{10}$  for an effect is sensitive to the prior distribution of the effect, we computed Bayes factors multiple times using priors with an increasingly larger standard deviation, following [23] and [24]. The relevant hyperparameters for the priors were derived from previous studies on the prosodic correlates of discourse structure [3]. The Bayes factors were computed by bridge sampling with four chains and 10,000 iterations, 2000 of which were in the warm-up phase.

Unless otherwise stated, the posterior distributions reported in the following result sections were all estimated by the model involving the prior with the largest standard deviation (i.e., the least informative prior).

## 3.2. Results

*The duration of the pause:* for all incorporated priors, the Bayes factors evaluating the effect of `subjectivity` on pause duration were around 0.1, in favor of the model without the effect, showing that `subjectivity` did not affect the pause duration between the two clauses.

*The duration of the connective:* the Bayes factors for `subjectivity` were greater than 100 for all incorporated priors, indicating very strong support for the presence of the effect. The model showed that the duration of the connective was greater in subjective causality than in objective causality (Estimate: 50 ms, 95% CrI [30 ms, 70 ms]).

*The  $f_0$  maximum and minimum in the first clause:* there was no evidence that `subjectivity` affected the  $f_0$  maximum and minimum in the first clause, as the  $BF_{10}$  for the effect was close to 0.1.

*The  $f_0$  maximum and minimum in the second clause:* the  $BF_{10}$  for the effect of `subjectivity` on the  $f_0$  maximum in the second clause was greater than 10, strongly favoring the presence of the effect of `subjectivity` on this measurement. The model estimated that the  $f_0$  maximum in the second clause was 0.71 st higher in subjective causality than in objective causality (Estimated: 0.71 st, 95% CrI [0.27 st, 1.16 st]). As for the  $f_0$  minimum in the second clause, the  $BF_{10}$  for the effect of `subjectivity` was less than 1, not in support of the presence of the effect.

*The speech rate in the first and second clauses:* the  $BF_{10}$  evaluating the effects of `subjectivity` on the speech rate in the first and second clauses was close to 0.1, in favor of the model without the effect.

## 4. Experiment 2

### 4.1. Method

#### 4.1.1. Participants

Thirty native speakers of Mandarin (15 females, 15 males; age range: 20-22 years; mean age: 21 years) took part in this experiment. The participants were recruited from Beijing International Studies University, China. All of them spoke a Beijing variety of Mandarin Chinese.

#### 4.1.2. Materials, task, and procedure

Thirty target items expressing subjective or objective causality in Mandarin were used in this experiment. The items were translated from the items used in Experiment 1 (see Section 3.1.2 for details). The items were elicited using the same task following the same procedure as used in Experiment 1. The optional connectives were *kejian (so)*, *yushi (so)*, and *danshi (but)*.

#### 4.1.3. Data and statistical analysis

The acoustic measurements analyzed in this experiment were identical to those used in Experiment 1. We normalized the duration of the connectives *kejian* and *yushi* using Z-scores.

## 4.2. Results

*The f0 maximum and minimum in the first clause:* the  $BF_{10}$  for the effect of `subjectivity` on the f0 maximum and minimum in the first clause was less than one, even under very informative priors, suggesting that there was no evidence in our data for the effect being present on the two f0 extremes over the course of the first clause.

*The f0 maximum and minimum in the second clause:* the model showed that subjective causality had a lower f0 maximum (Estimate = -0.66 st, CrI [-1.16 st, -0.16 st]) and a lower f0 minimum in the second clause (Estimate = -0.65 st, CrI [-1.19 st, -0.11 st]), compared to objective causality. Considering that the posterior 95% credible interval did not include zero, these results seemingly showed that subjectivity affected these two measurements. However, the Bayes factors were only 1-to-3 even when the prior was very constrained ( $Normal(0, 1)$ ), indicating very weak support for the model containing the effect of `subjectivity`.

*The speech rate in C1 and C2:* the Bayes factors evaluating the effect of `subjectivity` on these two measurements were less than 1, even under very constrained priors. The evidence indicated that in our data, `subjectivity` had no an effect on the speech rate of the two clauses.

*The duration of the pause:* The effect of `subjectivity` on the duration of the pause preceding the connective was estimated at a value of -0.05 s with a 95% CrI of [-0.09 s, -0.01 s], which seems to suggest that `subjectivity` had an effect on pause duration. However, the Bayes factor, even calculated under a very well calibrated prior ( $Normal(0, 0.1)$ ), only indicated a small likelihood (4.86) of the effect of `subjectivity` being present. When computed under less well calibrated priors ( $Normal(0, 0.5)$ ),  $BF_{10}$  dropped to 1.15, indicating that the model containing `subjectivity` was not any better than the model without the effect. Thus, there is no concrete evidence in our data that `subjectivity` affected the duration of the pause.

*The duration of the connective:* the  $BF_{10}$  for the effect of `subjectivity` remained smaller than 1 under different priors. Thus, the evidence did not support the presence of the effect on the duration of connective in our data.

## 5. General discussion

This study examined a trade-off relationship between prosody and specialized causal connectives in expressing subjective and objective causality using two experiments. Experiment 1 compared the prosodic realizations of subjective and objective

causality in English, where a general causal connective was often used, and Experiment 2 focused on Mandarin, where specialized causal connectives were more often used. The English results showed very strong evidence that subjective and objective causality was different in terms of prosody. The Mandarin results, in contrast, only showed very weak support for subjectivity having an effect on prosody.

Considered together, the results of the two experiments reveal that prosody is more rigorously used to distinguish subjective causality from objective causality in English, where a general causal connective is used, than in Mandarin, where specialized causal connectives are used. This finding supports the functional trade-off relationship between using specialized causal connectives and prosody in expressing subjective versus objective causality suggested by the Functional Hypothesis [9] and the Uniform Information Density Hypothesis [11]. It suggests that speakers deploy different communicative resources in an economical way, at least in the kind of conversational situation used in the current experiment, i.e., casual chats between acquaintances with comparable language and cognitive levels. Several studies share the same view, concerning other pragmatic domains. For example, it is argued in [25] that “speakers do not bother with prosodic cues if other cues are present to disambiguate structure.” It is proposed in [26] that if speakers can resolve the scope ambiguity of the adverb *now* by means of word order, then they might not use prosody at all, or use it to a lesser extent. It is argued in [27] that the need for a speaker to use prosodic devices to indicate the organization of the text would be eliminated by the presence of the explicit markers of discourse structures of a text.

Our finding also raised further questions about the role of prosody in communicating subjectivity in causality. The first question is the extent to which our findings can be generalized to other languages. The second question is whether or not the prosodic distinctions between subjective and objective causality in English can help listeners distinguish between these two types of causality. Future research can explore these directions.

## 6. Conclusions

In conclusion, we found across-linguistic evidence supporting a functional trade-off relationship between lexical cues and prosodic cues in expressing subjective and objective causality in English and Mandarin. The prosodic contrast between subjective and objective causality is stronger in English (along 2 of the 4 tested prosodic dimensions), where a general causal connective is used to express these two types of causality, than in Mandarin (only 1 of the 4 dimensions), where specialized causal connectives are in use. Moreover, the evidence in our data in favor of the effect of subjectivity is much stronger in English than in Mandarin.

## 7. Acknowledgements

We thank Orlando Jousset, Julie Leijtens, and Carla Kist for helping recruit the participants, and the interlocutors in the experiments, Aelish Hart (Experiment 1) and Jessica Wray (Experiment 2). This study was supported by a scholarship granted to the first author by the Chinese Scholarship Council (grant number 201709110154). The author also thanks the UiL-OTS phonetics lab for technical support.

## 8. References

- [1] J. Cole, "Prosody in context: a review," *Lang. Cogn. Neurosci.*, vol. 30, no. 1–2, pp. 1–31, 2015.
- [2] J. Hirschberg, Š. Beňuš, A. Gravano, and R. Levitan, "Prosody in Discourse and Speaker State," in *The Oxford Handbook of Language Prosody*, 2020.
- [3] H. den Ouden, L. Noordman, and J. Terken, "Prosodic realizations of global and local structure and rhetorical relations in read aloud news reports," *Speech Commun.*, vol. 51, no. 2, pp. 116–129, 2009.
- [4] T. J. M. Sanders, W. P. M. Spooren, and L. G. M. Noordman, "Toward a taxonomy of coherence relations.pdf," *Discourse Process.*, no. 15, pp. 1–35, 1992.
- [5] E. Sweetser, *From Etymology to Pragmatics*. 1990.
- [6] T. J. M. Sanders and E. Sweetser, "Introduction: Causality in language and cognition - what causal connectives and causal verbs reveal about the way we think," *Causal Categ. Discourse Cogn.*, pp. 1–18, 2009.
- [7] F. Li, J. Evers-Vermeul, and T. J. M. Sanders, "Subjectivity and result marking in Mandarin," *Chinese Lang. Discourse Chinese Lang. Discourse An Int. Interdiscip. J.*, vol. 4, no. 1, pp. 74–119, 2013.
- [8] M. Andersson, "The architecture of result relations: Corpus and experimental approaches to result coherence relations in English (Unpublished doctoral dissertation)," Stockholm University, Sweden, 2016.
- [9] J. Haan, *Speaking of questions. an exploration of Dutch question intonation*, vol. 52. 2001.
- [10] M. Aylett and A. Turk, "The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech," *Lang. Speech*, vol. 47, no. 1, pp. 31–56, 2004.
- [11] R. Levy and T. F. Jaeger, "Speakers optimize information density through syntactic reduction," in *Advances in Neural Information Processing Systems*, 2007, pp. 849–856.
- [12] W. E. Rutherford, "Some Observations concerning Subordinate Clauses in English," *Language (Baltim.)*, 1970.
- [13] E. Couper-Kuhlen, "Intonation and clause combining in discourse: The case of because," *Pragmatics*, 1996.
- [14] S. Günthner, "From subordination to coordination? verb-second position in German causal and concessive constructions," *Pragmatics Pragmatics. Q. Publ. Int. Pragmat. Assoc.*, 1996.
- [15] M. Scholman and V. Demberg, "Crowdsourcing discourse interpretations: On the influence of context and the reliability of a connective insertion task," pp. 24–33, 2017.
- [16] P. Boersma and D. Weenink, "Praat: doing phonetics by computer [Computer program]. Version 6.0.43," *retrieved 8 September 2018*, 2018. .
- [17] W. J. Hardcastle, J. Laver, and F. E. Gibbon, *The Handbook of Phonetic Sciences: Second Edition*. 2010.
- [18] R Core Team, "R: A language and environment for statistical computing.," *R Foundation for Statistical Computing, Vienna, Austria*. 2020.
- [19] P. C. Bürkner, "brms: An R package for Bayesian multilevel models using Stan," *J. Stat. Softw.*, vol. 80, no. 2017, 2017.
- [20] B. Carpenter *et al.*, "Stan: A probabilistic programming language," *J. Stat. Softw.*, 2017.
- [21] M. D. Lee and E.-J. Wagenmakers, *Bayesian cognitive modeling: A practical course*. Cambridge University Press, 2014.
- [22] H. Jeffreys, *The theory of probability*, Third. Oxford University Press.
- [23] S. Vasishth, B. Nicenboim, M. E. Beckman, F. Li, and E. J. Kong, "Bayesian data analysis in the phonetic sciences: A tutorial introduction," *J. Phon.*, vol. 71, pp. 147–161, Nov. 2018.
- [24] B. Nicenboim, S. Vasishth, and F. Rösler, "Are words pre-activated probabilistically during sentence comprehension? Evidence from new data and a bayesian random-effects meta-analysis using publicly available data," *Neuropsychologia*, vol. 142, no. February 2019, 2020.
- [25] J. Snedeker and J. Trueswell, "Using prosody to avoid ambiguity: Effects of speaker awareness and referential context," *J. Mem. Lang.*, vol. 48, no. 1, pp. 103–130, 2003.
- [26] B. Braun and A. Chen, "Intonation of 'now' in resolving scope ambiguity in English and Dutch," *J. Phon.*, vol. 38, no. 3, pp. 431–444, 2010.
- [27] C. L. Smith, "Topic transitions and durational prosody in reading aloud: Production and modeling," *Speech Commun.*, vol. 42, no. 3–4, pp. 247–270, 2004.