



Testing the Cognitive Processing Model of Chinese Scalar Implicatures

Si Liu¹, Chunmei Wang²

1. Lanzhou University; Mingdao Building, 199 W Donggang Rd, School of FLL, Lanzhou University
zsmjyk@126.com

2. Shaanxi Zhongtian Rocket Technology Co., Ltd., Academy of Aerospace Propulsion Technology;
Xiangyang Road, Tianhong Street, Baqiao District, Xi'an, Shaanxi
Wangchm212@163.com

ABSTRACT

This paper reports our experimental study of the cognitive processing of scalar implicature in the Chinese language. Our experiments tested the two processing models: the Default model proposed by the neo-Gricean theorists and the Context-Driven model supported by the post-Gricean, which has long been debated in the pragmatic research field. Our Experiment 1 investigated whether scalar implicature might be generated in neutral contexts (neither enabling nor canceling the processing of scalar implicature). By manipulating the scalar trigger *yixie* (some) in sentence-initial and sentence-final conditions, the reading times on the target phrase *qiyude/qitade* (the rest/the others) showed facilitation in the sentence-initial. In Experiment 2, a picture-sentence verification of under-informative utterance was performed in order to test the cost of generating scalar implicature. The results showed the significant effects of the two factors (Quantifiers <all, some> and Picture-Match/NoMatch). The yes/no responses showed that native speakers of Chinese were sensitive to scalar implicature. However, the role of context in processing scalar implicature was not supported. Experiment 3 was designed to test the processing of scalar implicature in Upper-Bound (enabling the processing) and Lower-Bound (canceling the processing) contexts. The result of this online experiment was slightly in favor of the Context-Driven model. The findings of the three experiments were significantly in agreement with neither the CM nor the DM. We presumed that the processing of scalar implicature might be accounted for with an alternative model. This was the first study to use this experimental paradigm in the Chinese language, whose findings were expected to highlight a need for further studies in order to investigate the meaning processing models with various languages and cultures.

Keywords: experimental pragmatics; scalar implicature; Default model; Context-Driven model; generalized conversational implicature

Academic Discipline And Sub-Disciplines

Linguistics

SUBJECT CLASSIFICATION

Experimental Pragmatics

TYPE (METHOD/APPROACH)

Experiments

1. INTRODUCTION

Utterance meaning not only contains the truth-conditional meaning of sentences but also the meaning that arises by considering the speaker's intention. In addition to the basic postulation of the theory of conversational implicature, the Griceans also classify the implicatures into two categories: the generalized conversational implicature (GCI) and the particularized conversational implicature (PCI).

Horn and Levinson etc. mainly follow a neo-Gricean line of study about conversational meaning, although some revisions and modifications have been made to Grice's original theory of conversational implicatures (Liu, 2008). Horn proposes that scale <S, W> forms a Q- or Horn-scale, as in <all, some>, <and, or>. S stands for "semantically strong expression" and W stands for "semantically weak expression." S and W are equally lexicalized from the same register, at about the same semantic relation or from the same semantic field. The first one is more informative than the second and it is asymmetrically entailed. As a result, such entailment scales or Horn scales give rise to scalar implicature (SI).

In the discussion of generalized quantity implicatures, Levinson accounts for scalar quantity implicatures as important sub-cases. Given his Q heuristic ("What you do not say is not the case"), Levinson proposed that SI is "a special case of a whole family of implicatures based on salient alternates (mostly, but not necessarily, ranked as informationally weaker or stronger)" (Levinson, 2000: 36). Under the Q-heuristics, Levinson held that the choice of an under-informative expression in a "Q-contrast sets" implies the assertion of the weaker items and the negation of the stronger items. Levinson proposed treating GCIs as the result of "default pragmatic inferences which may be cancelled by specific assumptions, but otherwise go through" (Levinson, 2000: 38). He claimed that the scale <all, some> is automatically activated by the use of "some," and the default inference to "not all" goes through the implicature.

However, this account has been challenged by the post-Griceans. Based on a cognitive account, the Relevance Theory regards the speaker and hearer as capable agents who have primary interest in keeping one's processing effort to a minimum when providing and decoding messages. The account of pragmatic inference by Sperber and Wilson (1995) and Carston (2004) implies that the on-line generation of SI only occurs when the context warrants it.



The processing model derived from the account of post-Griceans, represented by Sperber and Wilson (1995) and Carston (1991, 1998), assumes that SI is context-dependent or context-sensitive, which means that SI is understood by the relevance-theoretical mechanism.

We have reviewed the "Default inference and the Context-Driven view" (Breheny et al., 2006: 434), which are the two recent competing views. These are our concerns of focus, and we use the Default model (DM) and Context-Driven model (CM) for testing the SI processing in our study.

2. LITERATURE REVIEW

2.1. Some experimental studies testing the two models

The problem of how SI are processed on-line is closely examined by proponents of both the CM and DM through experimental investigations.

In the attempt to examine the role of minimal propositions played in understanding utterance, Bezuidenhout and Cutting (2002) designed a series of experiments to test the predictions of three different processing models: the Literal-First Serial (LFS) Model (based on Grice's views), the Local Pragmatic Processing (LPP) Model (based on the relevance) and the Ranked Parallel (RP) Model (neo-Griceans). The results of this study conform partially to the RP model. Bezuidenhout and Cutting (2002) attributed the lengthened reading times of sentences, in which the literal meaning is preferred, to the inhibition of the easy accessibility of non-literal meaning in the initial stage of processing.

The study of Chierchia's (2004) structural approach and the grammar and context interaction investigation by Katsos et al. (2005) have been reviewed and experimentally addressed by Garrett and Harnish (2009). They concluded that I-phenomena, in which SI is typical, are supposed to depend on "stereotypical background information." Partially in line with the mechanism of "default heuristics," they further discussed and concluded that "standardization" was justified.

Noveck and Posada (2003) conducted their experiments in order to test the two accounts in terms of processing resources, and with a sentence verification paradigm by using under-informative utterances, such as "Some elephants are mammals." It is found that the time taken by the participants answering "false" was significantly longer than those answering "true" in under-informative utterance. This suggests that when respondents based their answers on the plain meaning, it was not the case that the SI was first generated and then canceled but rather that the implicature was not generated in the first place.

Bott and Noveck (2004) replicated these findings in an experiment where an additional layer of narration is introduced. The stimuli were preceded either by the declaration that "Mary says the following sentence is true" or that "Mary says the following sentence is false." They found that the number of responses based on the SI increased as the permitted response time increased. All of the results suggest that SI are costly inferences in terms of how much time they need in order to be generated and that they manifest none of the automaticity or speed that the DM would expect.

Bezuidenhout and Morris (2004) employed the eye-movement tracking to address the controversy of the DM and underspecification model. They concluded that the DM is unacceptable for the readers who would abandon the default in the face of potentially conflicting information (the word "all" in our materials) rather than waiting until forced to do so (the "them were/did" region).

Papafragou and Musolino (2003) designed two sets of experiments to investigate the acquisition of SI. They compared the performance of 30 5-year-old native speakers of Greek and 30 adults on three different scales. Their results show that children are not as sophisticated as adults while defining the subtle aspects of the semantics-pragmatics interface. It has also emerged that the generation of SI is psychologically real and the success in children's computing scalar inferences is in accordance with most of work, in that the children's performances are largely dependent on the verbal instruction; thus, the role of context is more obvious.

Huang and Snedeker (2009) employed the visual-world eye-tracking paradigm to investigate scalar quantifiers by presenting pictures with different situations on them and displaying commands to the subjects. The result showed a significant delay between the semantic and pragmatic processing. Thus, they demonstrated there was "temporal lag between semantic processing and the initiation of pragmatic processing" (p.15).

2.2. Experimental studies of katsos and his colleagues

According to Katsos et al. (2005), the structural conditions are not sufficient but necessary in the on-line processing of SI, and thus Katsos et al. proposed that there are "further contextual constraints" in the pragmatic approach to detect the processing issue. In pursuit of such a consideration, they postulated a typology of contexts, that is, the "Upper Bound context" (UBC), "Lower Bound" context (LBC) and "Neutral" context (NC).

In order to examine GCI through the phenomenon of SI, Breheny, et al. (2006), with the UB, LB and N contexts, conducted four experiments as a serial exploration into the problem. The main purpose of this serial experiment is to test the "context-dependency or autonomy" of the generation of SI. They found a significantly longer reading time in the Upper-Bound contexts, which fit well into the picture portrayed in the CM (p.446). Yet, after discussing their findings with regard to "the modular versus interactive nature of the human parser," Breheny, et al. (2006) introduced the possibility of some "interactionist" versions of the DM. Katsos (2007) proposed a psychologically relevant criterion, which is that of the primary or secondary role of context. As a result, reaction time methods have been used by most of the studies and their results are variable.

Katsos (2009) examined both the classic scale (some < all) and the poset scale (triangle < {triangle, square}) between adults and children to compare the Context-Driven and Default accounts. He employed the experimental mechanism of the Truth Value Judgment Task. The result shows more preference to the Context-Driven account. Katsos and Boshop (2011) conducted a further study based on the 2009 work, and similar results were found.

3. OUR INVESTIGATION FOR CHINESE EQUIVALENT OF “SOME”

The language materials used in studies of SI processing were all in English, and we conducted the study in Chinese. In order to choose a test material equivalent of "some," we conducted a survey of questionnaires before the experiments.

There are several alternative equivalents for "some" in Chinese. For example, "yixie/youde/youxie" represents the meaning of "part of a whole" as "some" in English. The first part of the questionnaire contained ten English sentences with multiple choices of yixie, youde or youxie in Chinese, in which the critical blank of "some" was left. The participants of Chinese-English bilingual speakers were asked to choose the most appropriate word to complete the sentences. The

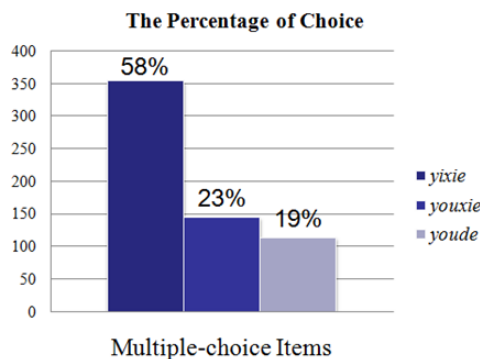


Figure 1. The percentage of choice (Part I of the Questionnaire)

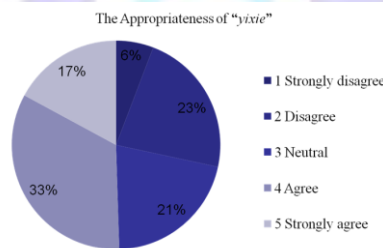


Figure 2. The percentage of Likert Scale (Part II of the Questionnaire)

second part was a Likert scale questionnaire, which consisted of twenty English sentences. The participants were asked to grade the degree of appropriateness or fitness of meaning in "some" with "yixie."

The results of the questionnaire, carried out by 30 participants, revealed that the choice of "yixie" for "some" is natural and appropriate (see Figure 1 and Figure 2).

4. EXPERIMENT 1: SCALAR IMPLICATURE GENERATION IN NEUTRAL CONTEXTS

4.1 Research Questions

As previously reviewed, the DM and CM make different predictions concerning what may happen in neutral contexts. Whether SI could be generated in this context is a central issue in demarcating the two processing models. So the first experiment was designed mainly to explore the generation of scalar expression yixie (some) in Chinese in neutral contexts.

The logic of the experiment is that if the inference bushi quanbu (not all) could be processed on-line by scalar triggers yixie (some) in the first sentence, then the reading time on the target phrase qitade (the others) could be facilitated.

The research questions in the present experiment are as follows:

- (1) Can SI be generated in neutral contexts?
- (2) Does the manipulation of scalar terms in the topic and non-topic position affect the processing of SI?
- (3) Is the processing of SI dependent on the context or autonomous?

4.2 Design of Experiment

A total of 20 sets of texts, each in four conditions, comprised the test materials. The texts were between 20 and 30 Chinese characters in length; the texts of the same set did not differ in size by more than 4 characters. The trigger phrase



yixie Xs (some of the Xs) always occurred in the first sentence. The target always appeared in the first segment of the second sentence and was followed by a segment that was identical in all four conditions. For example:

Yixie youke bei baoyu linshi-le. Qitade chumenqian dai-le yusan.

Some tourists were soaking wet from the substantial rain; others brought an umbrella when they went out.

Zhiyou-yixie youke bei baoyu linshi-le. Qitade chumenqian dai-le yusan.

Only some tourists were soaking wet from the substantial rain; others brought an umbrella when they went out.

Baoyu linshi-le yixie youke. Qitade chumenqian dai-le yusan.

The substantial rain made some tourists soaking wet; others brought an umbrella when they went out.

Baoyu linshi-de zhiyou-yixie youke. Qitade chumenqian dai-le yusan.

The substantial rain made only some tourists soaking wet; others brought an umbrella when they went out.

Furthermore, these 20 sets of texts for tests were supplemented with 20 filler texts, which were similar in length and size. In order to motivate participants' comprehension, and 25% of all the texts (both test and fillers) were followed by a yes/no comprehension question. The order of the presentation is randomized by conditions and contents.

4.3 Predictions

On the basic assumption that the topic position could create "implicit contextual expectations," while the non-topic position could not, proponents of the DM and CM make different predictions on the reading time on the target phrase.

According to the Default model, SI could be automatically processed in default reasoning by default rules (Levinson, 2000: 46). Since the scalar implicature "not all" could be available in neutral contexts, the reading time on the target phrase "qiyude/qitade" would be the same, regardless of the sentence position of the scalar trigger "some of the Xs." Besides, the reading time on the target phrase in both the "some" and "only some" condition should be comparable.

By contrast, the proponents of the Context-Driven model predict that the reading time on the target segment in the sentence-initial condition should be faster than that in the sentence-final condition because the contextual issues could be provided by the topic of the sentence. That is, sentence position should have an effect on the reading time on the target segment. When the trigger "some of the Xs" is in the sentence-initial position, the reading time should be facilitated.

In the two "only some" conditions, the inference "not all" is made explicit by the presence of the operator "only" (*zhiyou*). So another important prediction of the Context-Driven model is that if SI would be generated in the sentence-initial condition, the reading time on the target phrase should be as fast as possible in the sentence-initial "some" and sentence-initial "only some" condition.

Thus, if the results show any difference of reading time for the target phrase in a different sentence position, the influence of the linguistic context or sentence structural constraint on the generation of SI would be confirmed. Otherwise, it is not safe to conclude whether SI is generated on-line in a neutral context. In other words, "asymmetrical reading time" for the target phrases indicates that the confirmation of the CM and "comparable reading time" leads to verification of the DM. Besides, other divergent results may require us to resort to other explanations.

4.4. Participants

The participants were 24 native speakers of Chinese, aged between 19 and 26 years old, with the mean age of 22. There were 11 males and 13 females. They were students at the Lanzhou University, studying a range of science and arts subjects (but none were taking linguistics courses). Normal and corrected-to-normal vision is required. At last, all the participants were given a gift worth 5 RMB for their participation.

4.5. Procedures

4.5.1 Pilot Study

Prior to the full-scale research experiment, a pilot study was conducted with the aim to evaluate the feasibility of the experiment and improve the language materials used in the experiment. After analyzing the data collected in the pilot study and interviewing the participants, some revisions had been made in the language materials: the sentences were adjusted for the fluency and naturalness of Chinese, and each sentence had been revised and simplified into four segments in order to reduce the effect of word length.

4.5.2 The On-line Experiment

The experiment was run on a computer with DMDX software. All the language materials were in Chinese. The segments of the sentence stimuli were displayed from left to right by segments, with the previous segments remaining on the screen. After reading each segment, the participants had to press the button on the keyboard to see the next segment in a limited time. The reading time (RT) and yes/no answers (the left shift button means "no" and the right shift button means "yes"), in which the right answer is termed "accuracy rate" (AR), were recorded by the software.



The post-experiment interview revealed that no participants were aware of the purpose of the experiment. After discarding two participants' data, whose AR were lower than 80%, the final data we would use for analysis were settled. Participants' accuracy on the comprehension questions was 90% in total, with the highest AR of 100% and the lowest of 80%, which indicates that the participants focused their attention on understanding the language materials.

4.6 Results

The raw data were treated in the same way as Breheny et al. (2006). The raw data were trimmed by excluding some outliers, and the value for exclusion was set at 50ms. Overall 2.44% of the data were removed. In the four conditions, 2.9% of the data were removed from the sentence-initial "some" condition, 2.0% of the data were removed from the sentence-initial "only some" condition, 2.5% of the data were removed from the sentence-final "some" condition and 2.3% of the data were removed from the sentence-final "only some" condition.

The results show that there was no main effect on "sentence position" ($F=1.86, P=0.173 > 0.05$) or "explicitness factor" ($F=0.39, P=0.532 > 0.05$), which was mainly in agreement with the results of Breheny et al. (2006). In contrast to the results of Breheny et al. (2006), there was no interaction effect between sentence position and explicitness, $F=0.12, P=0.732 > 0.05$.

The mean reading times on the target phrase in the "only some" condition were both faster than that in the "some" condition but not statistically significant. In other words, the mean reading time on the "some" condition was as fast as the "only some" condition without considering the factor of the sentence position. The mean reading times on the target phrase in the sentence-final condition were slower than that in the sentence-initial condition but not statistically significant.

The same as the results of Breheny et al. (2006), the slowest mean reading time in the present experiment also appeared in the sentence-final "some" condition. A discrepancy came up in the fastest mean reading time. The fastest mean reading time of the experiment of Breheny et al. (2006) occurred in the sentence-final "only some" condition, which was unexpected. However, the fastest mean reading time was obtained in the sentence-initial "only some" condition in the present experiment, which roughly met the expectation.

However, the results of testing the segments following the target phrases showed no main effect of the explicitness factor $P=0.96 > 0.05$ and no interaction effect between sentence position and explicitness, $F=0.38, P=0.54 > 0.05$. But a significant effect of the sentence position, $F=4.23, P=0.04 < 0.05$, was revealed (see Table 1). This difference might be explained by the delayed effect carrying over from preceding segments. In other words, during the participants' reading of the last segment, they may think over the previous information in order to fully understand the whole sentence and may assign more reading time to the last segments that may have been originally assigned to the previous segments.

Therefore, the prolonged reading time on the last segments in the sentence-final condition might indicate that the reading times on the target phrase in the sentence-initial condition had possibly been facilitated (see Table 2). As previous data analysis revealed, the mean reading time on the target phrase in the sentence-final condition was also slower than the sentence-initial. This tendency became significant in the mean reading time on the segment following the target phrase. So far, some supporting evidence for the generation of SI in the sentence-initial position had been found.

Table 1. Main effect of sentence position

Source	Df	Mean Square	F	Sig.
Corrected Model	3	642008.098	1.54	.20
Position	1	1769127.525	4.23	.04
Explicitness	1	858.766	.002	.96
Position * Explicitness	1	160285.443	.38	.54
Error	1703	418227.183		

Table 2. Mean reading time on the segment following the target phrase

Factor	Conditions	Reading time (milliseconds)
Position	Sentence initial-'some' & 'only some'	706.01
	Sentence final-'some' & 'only some'	770.40
Explicitness	Sentence initial- & final- 'some'	738.91
	Sentence initial- & final- 'only some'	737.49



4.7 Discussion

The results of this SI generation experiment in neutral contexts by the self-paced reading task were in partial agreement with the results of Breheny et al. (2006). Although no significant difference occurred in the reading time of the target phrase in different conditions, a significant effect of the sentence position was found in the reading time on the following segment of the target phrase. On the assumption that the prolonged reading time may have been carried from the previous segment to the segment following the target, a tentative conclusion was drawn that the SI might be generated in the sentence-initial condition but not in the sentence-final condition. Thus, the result was slightly biased toward the CM.

Therefore, when “some” was triggered in the sentence-final or non-topic position, in which no contextual information can be accommodated, no effect on the facilitation was revealed in the reading time on the target phrase or the segment following the target. This indicated that SI may not be generated in neutral contexts, in which the inference “not all” is irrelevant.

It was only when the contextual information was manipulated by positioning the trigger at the topic condition that an effect of acceleration could be revealed in the reading time on the segment following the target not yet directly revealed in the reading time on the target phrase.

In this aspect, the processing of SI could be dependent on the context on a certain degree as the CM suggested. The generation of SI is a resource-demanding process instead of an autonomous or independent process as the DM claimed.

All in all, manipulating the scalar trigger in the sentence position may create some background or contextual information, but the effect was limited and indirect. Since the present experiment was conducted in the neutral context, the problem of whether the Chinese participants are sensitive to the scalar terms in the UBC and LBC and whether generation of SI is costly was still unsolved. These considerations would be taken into the present study in later experimental exploration.

5. EXPERIMENT 2: TESTING PICTURE-SENTENCE VERIFICATION

5.1 Research Questions

We referred the experiment of Tavano (2010) and adopted the method of the truth-value judgment paradigm. Also, we used pictures as experimental stimuli in order to create an extra-linguistic context based on which participants responded.

The primary research questions are as follows:

- (1) Is the processing of SI costly and effortful, as advocated by the CM, or effortless and costless, as claimed by the DM?
- (2) Are the Chinese participants sensitive to the phenomenon of scalar implicature when they encounter the scalar terms, especially in the under-informative utterances?

5.2 Design

We investigate the SI mainly through under-informative utterance by using pictures to create scenes in different degrees of “informativity.” Our basic strategy is to compare the participants’ accuracy rate and response time across the different conditions.





The pictures in the experimental items were composed of two categories: (1) all the objects in the picture were in the same color (Picture-All), while two groups of objects were in different color (Picture-Some). The sentences were descriptions of the pictures; and (2) All the sentences in the test items began with the Quantifier-All (QA) or Quantifier-Some (QS) in the scale <all, some> and were in the same sentence pattern. For example, Quanbu pingguo shi hongse-de (All the apples are red); and Yixie pingguo shi hongse-de (Some apples are red) (see Table 3).

We received four conditions by intercrossing about four factors: QAPA, QSPS, QAPS and QSPA. In the QAPA and QSPS conditions, the pictures agreed with the quantifiers; thus, the sentences were good/right descriptions of the pictures, and then we termed them QAPA Match and QSPS Match. In the QAPS and the QSPA conditions, the pictures did not agree with the quantifiers, and we termed them QAPS NoMatch and QSPA NoMatch (see Table 3).

All the pictures were of 520*440 pixels and displayed on the central of the computer screen. For the test items, two layouts of either 5 or 7 objects appeared in the pictures.

Except for the 20 test items, another 30 filler items in the same style but in different layouts were added. The sentences used to describe the filler items were statements about the color, location and state of the objects in the pictures. Besides, no under-informative sentence would be involved in the filler items.

Table 3. The four conditions and AR predictions

Conditions	Example Sentences	Picture Examples	Expected Answers
Quantifier-All-Picture-All (QAPA/Match)	<i>quanbu pingguo shi hongse de.</i> All the apples are red.		Yes
Quantifier-Some-Picture-Some (QSPS/Match)	<i>yixie pingguo shi hongse de.</i> Some of the apples are red.		Yes
Quantifier-All-Picture-Some (QAPS/NoMatch)	<i>quanbu pingguo shi hongse de.</i> All the apples are red.		No
Quantifier-Some-Picture-All (QSPA/NoMatch)	<i>yixie pingguo shi hongse de.</i> Some of the apples are red.		?

5.3 Predictions

5.3.1 Predictions on the accuracy rate

The experiment was designed to elicit an approximately equal number of “Yes” and “No” responses. Since the sentences are right and good descriptions to the pictures in QAPA Match and QSPS Match conditions, “Yes” responses are required. But the sentences in QAPS NoMatch are wrong or bad descriptions to the pictures; thus, “No” responses should be elicited QSPA NoMatch (see Table 3).

The response to the QSPA condition may be complicated. If the participants process *yixie* (some) semantically or logically as “possibly all,” a “Yes” response should be elicited because “some” is not contradicted with “all” when one interprets it as “possibly all” on the semantic sense. If participants process “some” pragmatically as “some but not all,” a “No” response should be enlisted and SI “not all” is generated. In the predictions of the DM, the proportion of a “No” response should be much higher because of the default and automatic generation of SI. Meanwhile, the Context-Driven model predicates did not differ significantly in the “yes/no” response, and more divergence of predictions would be revealed in the comparison of the response time.

In the predictions of the DM, the proportion of “No” responses would be much higher because of the default and automatic generation of SI. On the contrast, the CM predicates no significant difference in the AR. (More divergence of predictions will be revealed next in the comparison of the response time.)

5.3.2 Predictions on the RT

In the QSPA condition, the sentence is under-informative relative to the picture. The RT in this condition would be varied according to different processing models.

The predictions with the DM are as follows:

- (1) The RT might not differ across the conditions, since the default mechanism of SI leads to costless and automatic processing of under-informative utterance.
- (2) The RT of the Quantifier-All and Quantifier-Some would nearly be the same.

The predictions of the CM are as follows:

- (1) If participants cannot calculate an inference from the Quantifier-Some in the QSPS condition, the RT would be rapid. And the response time would be longer only in the condition of QSPA, in which the SI is relevant and necessary.
- (2) The RT of the other three conditions (QAPA, QSPA and QSPS condition) would be patterned; thus, no SI is needed.

5.4 Participants

In total, 36 participants took part in the present experiment. All of the participants were students who studied science and engineering, medical science and liberal arts at the Lanzhou University. All of them were native speakers of Chinese, aged between 19 to 26 years old, with the mean age of 22 years old. Besides, participants all had normal or corrected-to-normal vision and no colorblindness. After the experiment, each participant was given a gift.

5.5 Procedures

5.5.1 Pilot Study

At first, a pilot study was conducted. Twenty-seven students from the Lanzhou University participated in the pilot experiment. At first, the experiment was conducted with 110 sets of pictures and sentences comprising 80 target items



and 30 filler items. Twenty kinds of objects presented in the pictures appeared across the four conditions. After the experiment, a detailed interview was conducted with each of the participants. An experimental task effect was found during the interview. Most participants reported that they had taken some strategies to make a response since the same objects appeared twice but in different Picture-Match or NoMatch conditions. After data analysis, such an effect of task-specific strategies was found. Therefore, revisions had been made in order to address this shortcoming. At last, the pictures of repeated objects were deleted and a total of 20 pictures of the test items were employed in the on-line experiment. Besides, more participants were recruited in the formal on-line experiment.

5.5.2 On-line Experiment

The experiment was run on a computer with DMDX software. At first, a serial number was displayed on the top of the computer screen. After the picture had stayed on the screen for approximately 2 seconds in order for the participants to observe the objects, the sentence describing the picture appeared. Participants were required to press the button on the keyboard within a limited time (3000 ms). Then the responses and time could be recorded by DMDX software.

A Latin square is employed in the design of the presentation order with each item rotated in the five experimental blocks in four different conditions across the presentation list.

The task for participants was the picture verification. Participants were instructed to decide whether the sentences match the picture or not. Participants responded by pressing buttons on the keyboard, indicating “Yes” (right/good description) or “No” (wrong or bad description). The complete experiment session on each participant lasted approximately 25 minutes, including an instruction and training session prior to the on-line experiment and an interview after the on-line experiment.

5.6 Results

5.6.1 Results of AR

The results of the AR were mainly in accordance with the predictions. Nearly 100% of responses to the Quantifier-All-Picture-All and Quantifier-Some-Picture-Some conditions, also known as the Picture-Match conditions, were “Yes,” and they were “No” for the Quantifier-All-Picture-Some condition (see Figure 3). Furthermore, most of the participants’ (78%) responses were relatively consistent in response. Also, these results also agree with the pilot study, even though the proportion of the “No” responses declined slightly and the task-strategic effect reduced.

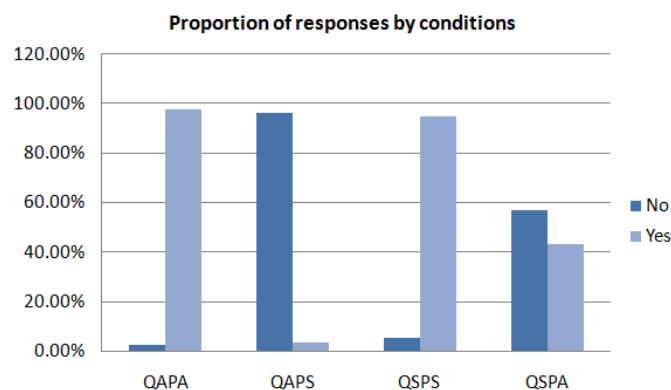


Figure 3. The accuracy rate for different groups

The results for Quantifier-Some-Picture-All (QSPA) items were mixed. With an agreement to the prediction, there were more pragmatic interpretation “No” responses (99 of 180 responses, 56%) than logical-interpretation “Yes” responses (80 of 180 responses, 44%) (see Figure 3). Most of the participants responded with consistency. Twenty-eight of the 36 participants provided the same answer in the QSPA trials, with 17 consistently answering “No” (the pragmatic response) and 11 consistently answering “Yes” (the logical response).

The results of AR in QSPA were contrary to Tavano’s (2010) original experiment, in which such results were satisfactory but were not obtained. In addition, this split was in agreement with the study in Noveck (2001) and Noveck and Posada (2003). In the interview after the experiment, no participants were aware of the experimental purpose. They also reported no difficulty in identifying the pictures and seldom made errors when pressing the keyboard.

5.6.2 Results of Response Times

The results of AR included just one parameter in the present experiment. Much information can be revealed in comparison of the response time. The two cases of over-time and some wrong responses were excluded from the raw data. A two-way Analysis of Variance was conducted to analyze the response time data with the factors of Quantifier (Some/All) and Picture Match (Match/NoMatch).



Table 4. ANOVA for the RT differences

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	13085752.367 ^a	3	4361917.456	31.27	.000
Picture Match	5981221.170	1	5981221.170	42.88	.000
Quantifier	8025912.893	1	8025912.893	57.54	.000
Picture * Quantifier	2742356.777	1	2742356.777	19.66	.000
Error	85637751.554	614	139475.165		

Table 5. Independent Samples Test for the test condition QSPA

		N	mean	SD	Levene's Test for Equality of Variances		t-test for Equality of Means		
					F	Sig.	t	df	Sig. (2-tailed)
RT	Yes (logical)	80	1514.60	465.82					
	No (pragmatic)	90	1792.71	443.45					
	Equal variances assumed				.258	.612	-4.08	177	.000
	Equal variances not assumed						-4.06	166	.000

As shown in Table 4, the main effect of Quantifier and Picture-Match was revealed. There was also an interaction effect. The results of ANOVA mainly agreed with the experiment of Tavano (2010).

Figure 4 shows RT by condition. It illustrates that RT in QSPA was significantly slower than RT in QAPA. However, almost no difference was revealed between QAPS and QSPS. Therefore, in the same Picture-All scene, participants took longer to respond to the under-informative utterances. In the same Picture-Some scene, however, participants responded to the different quantifiers in nearly the same speed, which caused a variation from the results of Tavano (2010). Furthermore, RT in QSPA was much longer than RT in QSPS. Above all, response times were similar among QAPA (Match), QAPS (NoMatch) and QSPS (Match) conditions, but a significantly faster response time appeared in QSPA.

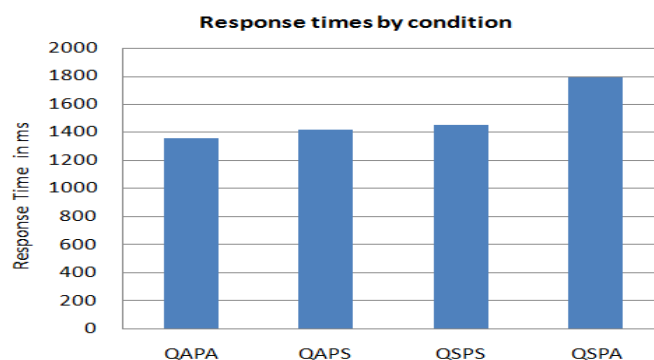


Figure 4. Response times by condition

Last but not the least, the target test condition QSPA (Quantifier-Some-Picture- All) was split by AR. The response times of the two interpretations (logical interpretation “possibly all” and pragmatic interpretation “not all”) were also compared with an independent samples t-test. As shown in Table 5, the RT of the “No” response was significantly slower than the RT of the “Yes” response in the QSPA condition ($F = 0.258, p = 0.05$).

5.7 Discussion

The Context-Driven model was supported after the analyses described above. The data results were not only in agreement with the results of Tavano (2010) but also in accordance with most of the previous experimental studies, which had been reviewed in chapter three.



In general, the split AR shows that native speakers of Chinese are sensitive to scalar implicature when they encounter the scalar terms, especially in the under-informative utterances. The significant effect of Picture-Match and Quantifier and their interaction effect suggest the generation of scalar implicature in the QSPA condition. Also, the obvious slow-down when encountering scalar terms in relevant scenes, which makes it under-informative, suggests additional processing resources, and the cost is required in generation of SI. Further comparison between RT and AR also confirmed that effort is necessary in generating scalar implicature. All of these results were biased toward the Context-Driven model, which was also obtained in some previous experimental studies (Noveck & Posada, 2003; Bott & Noveck, 2004; Huang & Snedeker, 2009). Other experiments, in which the rate of SI generation declines along with the increase of cognitive load, were also standard for the costly SI processing.

Besides, the shortest response time appeared in the QAPA condition, in which the informativity of the sentence was rightly suitable to match the picture. Both the sentence and picture were "all" and had no inference at all. Therefore, the response time of it is shortest in the four conditions.

However, it cannot be ignored that a main effect on Picture-Match was shown in the results. Some explanations might be needed. The picture scenes used in the present experiment were close set. Participants responded with a confirmation or disagreement in matching the picture and sentence and could not avoid referring to other experimental items. Nonetheless, this cannot be a denial to the generation of SI in under-informative utterance. In addition, the longer response to the QSPA condition than the QSPS response might be attributed to the difference in relevance between the sentence and the picture, since the QSPS response times were also significantly slower than the QAPS times. It could be concluded that SI was not generated on-line in the QSPS condition, although there was the quantifier "some" in the sentence.

Above all, the results of the present experiment demonstrated considerable support to the CM of SI processing. Furthermore, the role of context in processing implicature still needs to be explored, which will be further described in the next experiment.

6. EXPERIMENT 3: SCALAR IMPLICATURE GENERATION IN UBC AND LBC

6.1 Research Questions

As previously presented, the first repeated experiment of Breheny et al. (2006) was conducted in the sentence without any relevant context, and the second experiment was a verification of the processing cost and sensitivity of SI in picture-created scenes. It was verified that the generation of SI was context-dependent, but how the role of context plays in the processing of scalar implicature was not fully investigated. So our third experiment was designed: the SI generation in UBCs and LBCs in Chinese.

The research question of the present experiment is as follows:

Do the Upper-Bound and Lower-Bound contexts affect the SI processing?

6.2 Design of Experiment

The experiment comprised 36 sets of texts with three additional trail items. A total of nine pairs of texts constituted the test items of the experiment, and one item in each pair differed in the given contexts (the Upper-Bound and the Lower-Bound contexts). Correspondingly, there are another 10 texts comprised of the control items. Also, eight other texts, which are similar in size to the test items, serve as fillers of the experiment and disperse in the different types of texts. All of the texts were between 45 and 50 characters in length. The two items of each pair are nearly the same in length. The experiment lasted for nearly 20 minutes with a pre-experiment instructions and a post-experiment interview.

Each of the texts consisted of three lines. The first sentence assigned a role to the participants, and a particular purpose was provided. The second sentence provided the condition that a reward should be granted. The third sentence was an utterance said by the other person involved. The participants were asked to decide whether a reward should be granted to the speaker.

6.3 Predictions

In terms of the relevance of informativity, the UBC in the SI "not all" is highly relevant in deciding whether the reward should be granted. If the participants calculate the SI, they are supposed to give a "No" response, indicating no reward should be granted. If the SI are not inferred by the participants, a "Yes" response is predicted. In the LBC, the reward should be granted for the request if the first sentence is met (see Table 6).

Besides, in terms of response time, if the SI is generated in this condition, the response time should be longer than that in the UBC, given that the generation of SI is more effortful. While if the generation of SI has not occurred, the response time in the LBC might be shorter than in the UBC, according to the CM. However, the predictions by the DM are right in contrast with these (see Table 6). In terms of RT, the only call for the matching of semantic content without any calculation of implicatures leads to no obvious shorter RT in these conditions (see Table 7).



Table 6. Predictions on AR

	Test items	Control items
Upper-Bound context	No (SI generated) Yes (SI not generated/ cancelled)	Yes
Lower-Bound context	Yes	No

Table 7. Predictions on Response time

	Test items
Default model	RT(Upper-Bound) < RT(Lower-Bound) RT(Upper-Bound) ≈ RT(Lower-Bound) (cancellation of the activated inference is costly)
Context-Driven model	RT(Upper-Bound) > RT(Lower-Bound) (generation of SI is costly)

6.4 Participants

There were 25 students from the Lanzhou University who participated in this experiment. All of them were native speakers of Chinese, aged between 21 to 26 years old, with the mean age of 23. Normal or corrected-to-normal vision was required. Participants majored in various disciplines of liberal arts, science and technology and medical science. All the participants were given a gift worth 5 RMB as a reward for participation.

6.5 Procedures

6.5.1 Pre-experiment Questionnaires

A pre-experiment questionnaires survey was conducted. After analysis, the contexts were reasonable in that the discrimination of the Upper-Bound and Lower-Bound contexts was confirmed by the ratio of the answers enlisted from 30 participants (see Figure 5). At last, the materials are revised and programmed to be tested on-line by DMDX software on the computer.

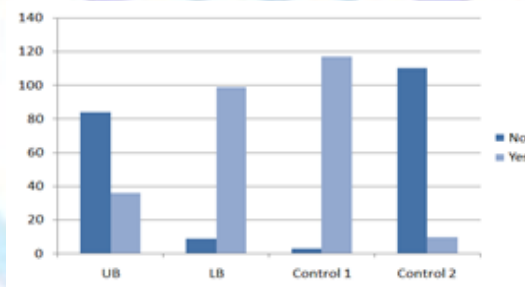


Figure 5. Results of questionnaires

6.5.2 Pilot study

Before the main experiment, a pilot study was made in order to address the problem of the stimuli's validity and fidelity of the response time assigned. Eighteen native speakers of Chinese took part in the pilot study. After analyzing the results of the pilot study and revising, the final version of the language materials and programming was settled.

6.5.3 The On-line Experiment

The experiment was run on a computer with DMDX software. The texts were automatically displayed line by line on the central position of the computer screen with a 2-second delay to the previous line. The trigger phrase "some" always occurred in the final line of the texts. The presentation of the experimental texts was randomized. Responses were enlisted from the participants when they encountered the final line of the text.

In addition, short interviews were conducted. Several questions were designed in the semi-structural interview. The post-experiment interview revealed that no participants were aware of the purpose of the experiment.

6.6 Results

6.6.1 Results of Accuracy Rate

The data of AR were processed; the result of AR mainly conformed to the predictions. The accuracy of the control items is nearly 95%. The particular high accuracy of controlling items is justified, although some slips or over-time occurred.



Table 8. AR in control items

	Right	Wrong	Over-Time
Situation Type 1	96.8%	0.8%	2.4%
Situation Type 2	94.4%	1.6%	4.0%

Table 9. AR in test items

	Yes	No	Over-Time
Upper-Bound context	14.2%	78.7%	7.1%
Lower-Bound context	95.6%	2.2%	2.2%

It should also be noted that the “Yes” response of granting the reward was over 95% (see Table 8), in which the under-informative expression “some” met the demands given in the first and second line. SI was not necessary in this context, for no relevance of the contextual factors is called for. And in this condition, it was difficult to say whether the SIs were generated or not.

The result of the Upper-Bound contexts was split as predicted, that is, 78.7% of “No” response and 14.2% “Yes” response (see Table 9). In addition, the over-time occurred at a particular high rate 7.1%, which will be discussed in a later section. The AR showed that the scalar expression “some” led to some uncertainty and confusion, which is termed “semantic-unspecification” by the relevance theorists. Except for this, no more information could be calculated by the AR on its own, and the results of RT may reveal something more.

6.6.2 Results of Response Time

The response time of whether granting a reward during reading the final line of texts can show far more information than the single AR by comparing it in different contexts and conditions.

The data were processed with Analysis of Variance. The RTs of the four conditions: the UBC and LBC of the test items and the semantically right and wrong items of controls were compared. The RT of the UBC of the test items was the longest. One-way ANOVA revealed a significant effect of the test items; $F=5.09$, $P=0.002$ (see Table 10).

Table 10. ANOVA for a comparison between the test items and control items

	Test items	Control items	F	Sig.
Upper-Bound context	1324	1157	5.09	0.002
Lower-Bound context	1174	1274		

Table 11. ANOVA for a comparison of response time between UBCs and LBCs in test items

	Mean	SD (ms)	F	Sig.
Upper-Bound context	1234.43	449	6.93	0.26
Lower-Bound context	1144.76	377		

Notes: The data of time-over and the errors were eliminated.

The response time of the test items is compared in the two different contexts: the Upper-Bound context and the Lower-Bound context after the over-time and the errors were eliminated. For the reading times for the last line of the texts, in which the trigger segment “some” was involved, a significant effect was revealed, $F=6.93$, $P=0.26$ (Equal Variances Not Assumed) (see Table 11).

Furthermore, as we noticed in the analysis of the AR, the relatively large proportion of timeout occurred in the Upper-Bound contexts, which further illustrated the hesitation of the participants in calculating SI in this context. According to the predictions of the first section of this chapter, the data results of the test items appear in favor of the Context-Driven model in this respect.

The time of “Yes” or “No” responses in the UBC was also compared. A significant difference was shown. The time of “No” response implies the negation of granting the reward, and the time of “Yes” response implies affirmation of granting the reward (see Table 12). As predicted by the two rival processing models, SI could be generated in the Upper-Bound context, which accounted for the high rate of “No” responses. This is because the “Yes” response means that the participants would like to grant the reward to the people involved who did not meet the demands of the “mention-all” expression provided. It took a little bit longer to accept the under-informative expression “some,” although no significance on the statistical sense has been shown.



Table 12. ANOVA for a comparison of the response times between YES or NO in the UBC

	Upper-Bound context	SD(ms)	F	Sig.
Yes	1298.25	456	0.332	0.02
No	1178.09	411		

6.7 Discussion

As analyzed above, the results of this on-line experiment were in favor of the CM, since there was a longer response time in the UBC than in the Lower-Bound context and control items. The calculation of SI was effortful from this respect. In the LBC, the scalar inference “not all” was irrelevant and interpretation was unnecessary. The response time was shorter in general. In the relevance theorists, the LBC encourages no implicature because of the lack of the contextual effect.

The results of the control items show no significant acceleration of response time, especially in the condition that no reward was granted, which was not predicted at first. This might be explained because the metalinguistic negation took a longer time.

7. GENERAL DISCUSSION AND CONCLUSION

In Experiment 1, SI generation in neutral contexts offers initiative support to the CM of SI processing. The manipulation of the contextual effect by the sentence position might affect the processing of SI; however, the effect is limited since the significant difference between sentence positions is found in the segment following the target phrase. So the conclusion could be made that SI could not be generated in the linguistic neutral context (sentence-final position), except if the trigger “some” is positioned in sentence-initial, and if a certain degree of contextual effect has been created, then SI could be generated.

In Experiment 2, the picture-sentence verification shows that the participants are sensitive to scalar implicature, especially when they encounter scalar terms in under-informative utterances. It is a pragmatic process of enrichment or expansion of semantic indeterminacy, which requires cognitive effort rather than automatic process.

Experiment 3, the on-line experiment of scalar implicature in Upper-Bound and Lower-Bound contexts, demonstrates that scalar implicature might be generated in Upper-Bound but not in Lower-Bound contexts; and the generation of SI is also costly.

As a general conclusion for the three experiments in this study, the generating of scalar implicature might not just be subject to some structural properties or the linguistic context of the utterance and not just be an interaction between the grammar and linguistic-context constraint, as supported by Katsos et al. (2005). Neither CM nor DM was supported in this study. Some other model (for example, standardization model) could be tested to determine if it is a more acceptable account for the processing.

ACKNOWLEDGMENTS

Our thanks go to the support by the Funds of Humanities and Social Science Planning Projects of China Ministry of Education (11YJA740053).

REFERENCES

- [1] Liu, S. 2008. A Review on Experimental Pragmatics. *Journal of Modern Linguistic*, 10(3), 246-256.
- [2] Levinson, S. 2000. *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. Cambridge, MA: MIT Press.
- [3] Sperber, D., & Wilson, D. 1995. *Relevance: Communication and cognition*. 2nd Ed. Oxford: Blackwell.
- [4] Carston, R. 2004. Relevance theory and the saying/implicating distinction. In L. Horn & G. Ward (Eds.), *The handbook of pragmatics* (pp. 633–656). Oxford: Blackwell.
- [5] Carston, R. 1991. Implicature, explicature, and truth-theoretic semantics. In S. Davis, (Eds), *Pragmatics:A reader* (pp. 33-51). New York: Oxford University Press.
- [6] Carston, R. 1998. Informativeness, relevance and scalar implicature. In R. Carston & S. Uchida (Eds.), *Relevance theory: Applications and implications* (pp. 179-236). Amsterdam: John Benjamins.
- [7] Breheny, R., Katsos, N., & Williams, J. 2006. Are generalized scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition*, 100(3), 434-463.
- [8] Bezuidenhout, A., & Morris, R. 2004. Implicature, relevance and default inferences. In I. Noveck, & D. Sperber (Eds.), *Experimental pragmatics*. Basingstoke: Palgrave Macmillan. p.257-282
- [9] Chierchia, G. 2004. Scalar implicatures, polarity phenomena, and the syntax/pragmatic interface. In A. Belletti (Eds), *Belletti structures and beyond*. Oxford: Oxford University Press.



- [10] Katsos, N., Breheny, R., & Williams, J. 2005. Interaction of structural and contextual constraints during the on-line generation of scalar inferences. Paper presented at Proceedings of the 27th annual conference of the cognitive science society, Mahwah, NJ: Erlbaum.
- [11] Garrett, M., & Harnish, R. M. 2009. Q-phenomena, I-phenomena and implicature: some experimental pragmatics. *International Review of Pragmatics*, 1(1), 84-117.
- [12] Noveck, I. A., & Posada, A. 2003. Characterizing the time course of an implicature: An evoked potentials study. *Brain and Language*, 85(2), 203-210.
- [13] Bott, L., & Noveck, I. A. 2004. Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language*, 51(3), 437-457.
- [14] Bezuidenhout, A., & Cutting, J. C. 2002. Literal meaning, minimal propositions, and pragmatic processing. *Journal of Pragmatics*, 34(4), 433-456.
- [15] Papafragou, A., & Musolino, J. 2003. Scalar implicatures: Experiments at the semantics- pragmatics interface. *Cognition*, 86(3), 253-282.
- [16] Huang, Y. T., & Snedeker, J. 2009. Semantic meaning and pragmatic interpretation in 5-year-olds: Evidence from real-time spoken language comprehension. *Developmental psychology*, 45(6), 1723.
- [17] Katsos, N. 2007. *The Semantics/Pragmatics Interface from an Experimental Perspective: the Case of Scalar Implicature*, *Synthese*, 165 (3), 385 - 401.
- [18] Katsos, N. 2009, Evaluating under-informative utterances with context-dependent and contextindependent scales: experimental and theoretical implications, In Sauerland, U. & Yatsushiro, K., *Experimental Semantics and Pragmatics*. 51-73.
- [19] Katsos, N. & Bishop, D. V. M. 2011. Pragmatic tolerance: Implications for the acquisition of informativeness and implicature, *Cognition*, 2011.
- [20] Tavano, E. M. 2010. *The Balance of Scalar Implicature*. (Doctoral dissertation). University of Southern California .
- [21] Noveck, I. A. 2001. When children are more logical than adults: Experimental investigation of scalar implicatures. *Cognition*, 78(2), 165-188.
- [22] Huang Yan. 2001. Reflections on theoretical pragmatics. *Journal of Foreign Languages*, 131 (1) 2-14.

Author' biography with Photo

Liu, Si, currently a full professor at the School of Foreign Languages and Literature of Lanzhou University, China. She was a postdoctoral fellow in the Department of Psychology of the University of Arizona, USA, specializing in the field of experimental pragmatics, from 2007 to 2009. She was awarded a PhD of linguistics at the University of Arizona in 2004. She earned an MA of applied linguistics at Guangdong University of Foreign Studies in 1993, after having acquired graduate course credits at Beijing Foreign Studies University, China.

Wang, Chunmei, MA, graduated from the School of Foreign Languages and Literatures, Lanzhou University, works for Shaanxi Zhongtian Rocket Technology Co., Ltd, and takes charge of international business of the company.