

**THE EFFECTS OF SYNTACTIC KNOWLEDGE
AND EXTRA-GRAMMATICAL FACTORS ON SENTENCE
ACCEPTABILITY JUDGMENT:
AN EMPIRICAL STUDY ON
CHINESE TOPIC CONSTRUCTIONS**

Chun-Chieh Hsu¹ and Shelley Ching-Yu Hsieh²

¹*National Tsing Hua University*

²*National Cheng Kung University*

ABSTRACT

In this study, we explore how different components like syntactic knowledge and discourse context may interact with each other in sentence acceptability judgment performance. We examined the acceptability judgment patterns on Chinese topic constructions in different contexts and we also looked at how different types of topic NPs (definite topic NPs vs. bare topic NPs) may affect the judgment patterns. Our findings show that island-obeying topic sentences were consistently judged better than their island-violating counterparts, whether presented in a null context or in a relevant context. This contrast suggests that discourse context and syntactic knowledge do not interact in sentence acceptability judgment performance, and favors a categorical approach to grammar. In addition, our findings also show that topic sentences with a definite topic NP were consistently rated better than those with a bare topic NP, and that canonical sentences which served as the baseline were also consistently judged to be better than grammatical topic sentences. These patterns suggest that acceptability judgment is continuous with a wider range of variations due to the ambiguity, frequency, and processing complexity involved in the given sentences. Overall, our study not only suggests that the traditional acceptability judgment task can elicit good data about grammaticality, but it also recognizes that extra-grammatical factors can affect acceptability judgment performance and should therefore be taken into consideration to ensure the quality of the collected data.

Key words: Acceptability judgment, grammaticality, Chinese topic constructions

1. INTRODUCTION

Acceptability judgment has long been the main source of data that linguists depend on to formulate syntactic rules and theories. It is standard practice to obtain sentence acceptability data from language informants who apply their native intuition to judge whether the given sentences are acceptable or unacceptable. Linguists, especially syntacticians, prefer sentence judgment data provided by the informants over naturalistic data from the existing corpora, because the judgment data on the specifically-designed sentences allow them to test their theoretical proposals directly. In these recent fifteen years, this widely-practiced method has received criticism and challenges. The debate centers on the issue of whether the informally judged data are reliable for making solid syntactic claims (Bard, Robertson, & Sorace, 1996; Cowart, 1997; Culicover & Jackendoff, 2010; Edelman & Christiansen, 2003; Featherston, 2005b, 2007; Ferreira, 2005; Gibson & Fedorenko, 2010a, 2010b; Myers, 2009a; Phillips, 2009; Phillips & Lasnik, 2003; Schütze, 1996; Sprouse & Almeida, 2012, to appear, etc.). One cause for this controversy is related to the different assumptions between grammaticality and acceptability. The working assumption for syntacticians is that grammaticality is based on our syntactic knowledge, which is part of linguistic competence, and it is categorical – either grammatical or ungrammatical. Therefore, it is common that linguists strive to understand syntactic competence by eliciting native speakers' intuitive judgments because they assume that judgment data reflect native speakers' syntactic knowledge which would *not* be affected by extra-grammatical factors such as input frequency, processing complexity, and pragmatic context (Chomsky, 1965, 1972, 1975). However, one worry that has arisen from this assumption is that syntactic competence is elicited via acceptability judgment, which is actually a complex behavioral process that could potentially involve different components other than syntactic knowledge. Sentence acceptability judgment is a kind of linguistic behavior and belongs to linguistic performance. Unlike grammaticality which is categorical, judgment performance is continuous with intermediate levels along a spectrum, because it could be affected by extra-grammatical factors,

such as input frequency, working memory limitations, and plausibility (Sprouse, 2007b; also see Fanselow, Féry, Schlesewsky, & Vogel, 2004; Featherston, 2005a; Keller, 2000; Sorace & Keller, 2005, etc.). For example, a lot of syntacticians have the experience of getting uncertain responses from their informants like “*The sentence sounds strange, and people usually don’t say it this way.*” or “*This sentence is kind of weird, and it needs a special context.*” Therefore, it is sometimes difficult to determine what exactly the informants’ judgments are based upon (i.e. upon their grammatical knowledge, extra-grammatical factors, or both). Such difficulty is especially obvious when dealing with complex sentences. To control the quality, consistency, and reliability of the judgment data, some studies have adopted formal judgment experiments for testing grammatical hypotheses (Ambridge & Goldberg, 2008; Bernstein, Cowart, & McDaniel, 1999; Clifton, Fanselow, & Frazier, 2006; Cowart, 1997; Featherston, 2005a; Myers, 2009b, 2012, etc.).

The goal of this paper is not to discuss how reliable the informally judged data can account for grammaticality. Instead, we focus on acceptability, and examine how different components may or may not interact with each other in sentence judgment performance. It is worth exploring how syntactic knowledge and extra-grammatical factors may affect people’s linguistic judgment behavior because it not only helps us further understand how linguistic judgment behavior works, but also provides insightful implications for understanding grammaticality.

1.1 Extra-grammatical Factors and the Role of Context

When judging the acceptability of any given sentence, both syntactic knowledge and extra-grammatical factors are involved. Syntactic knowledge is considered the competence factor that plays a pivotal role in determining the acceptability of a sentence. Extra-grammatical factors, on the other hand, belong to performance factors that may affect judgment results. These factors include the frequency and the processing complexity of the given sentences, and the pragmatic context in which the sentences could be interpreted properly. The effects of input frequency and processing complexity on off-line sentence judgment have been proven empirically. For example, for the effect of frequency,

studies show that participants rated sentences as more acceptable if they had read them earlier and that speakers' previous experience with the sentences can improve their acceptability rating of similar types of sentences (Luka & Barsalou, 2005; Dąbrowska, 2010). For the effects of processing complexity, past studies show that sentences involving complex structures like *wh*-movement would cause more difficulty in processing and thus reduce the acceptability of both grammatical and ungrammatical sentences (Casasanto, Hofmeister, & Sag, 2010; Sag, Hofmeister, & Snider, 2007).

As for the effect of context, while abundant studies show that felicitous context can facilitate on-line sentence comprehension (Altmann, Garnham, & Henstra, 1994; Altmann, 1989; Crain & Steedman, 1985; Grodner, Gibson, & Watson, 2005; Kaiser & Trueswell, 2004; Spivey-Knowlton, Trueswell, & Tanenhaus, 1993; van Berkum, Brown, & Hagoort, 1999, etc.), its role in off-line sentence acceptability judgment performance is less clear. Sprouse (2007a) examined whether the presence or absence of context may affect the acceptability of island violation sentences in English. He compared the acceptability judgment results of *wh*-question sentences like the sentences in (1) when they were situated in a null context with the results of these same sentences when they were preceded by a felicitous context sentences as those in (2), which provided an appropriate answer (underlined) to the target *wh*-question in (1)

- (1) a. Who_i *t_i* denied [_{NP} the fact that you could afford the house?]
b. *What_i did you deny [_{NP} the fact that you could afford *t_i*?]
(2) a. Context: You denied the fact that you could afford the house.
b. Context: You denied the fact that you could afford the house.

Sprouse (2007a) found that the judgment results of the *wh*-questions proceeded by a relevant context sentence were not different from the results of the *wh*-questions presented in isolation. His finding suggested that a felicitous context that provides the intended meaning of an island violation sentence does not affect its off-line acceptability in English. Although Sprouse (2007b) did not find the effect of context on off-line sentence acceptability judgment in English, it is possible that the effect

of context may be found on a different construction in a typologically different language. The hypothesis here is that context may affect acceptability judgment performance when the syntactic construction needs a context for interpretation and when the language is highly dependent on discourse context for interpretation. Thus, the goal of the present study is to further examine whether the roles of syntactic knowledge and discourse context may interact with each other in the off-line sentence acceptability judgment performance, and we look at topic constructions in Mandarin Chinese to investigate this. In the next section, we explain why topic constructions in Chinese were chosen as our target test sentences.

1.2 Chinese Topic Constructions and The Present Study

Chinese is chosen as the target language in our study because it has been considered as a language that depends heavily on context for sentence interpretation for several reasons. First, it lacks inflectional morphology: no case-markings to indicate the grammatical role of the nouns, no subject-verb agreement in terms of number, gender or person, and no overt markers to indicate clause boundaries. Basically, the grammatical relations between major constituents are linked by far fewer overt morphological devices than many other languages. Second, Chinese is a pro-drop language without any explicit case markers, and it has been classified as a “cool” language for its frequent use of null pronouns (Ross, 1982). It allows both null subjects and null objects whenever they could be appropriately identified in the relevant context. Thus, understanding a sentence in Chinese may naturally require some effort on the receiver’s side, and sentence interpretation often involves context and world knowledge (Huang, 1984). Third, in addition to the canonical SVO word order, Chinese permits other word order variations in different pragmatic situations, such as SOV, OSV, and VOS (Li & Thompson, 1981), and these different surface word orders occur frequently in daily conversations. These characteristics of Mandarin Chinese make it a language worth testing the effect of context on acceptability judgment performance.

Importantly, we test topic constructions in Chinese because they not only have to meet syntactic locality constraints but also require some kind of presupposed discourse. “Topic” is a discourse notion (Li & Thompson, 1979; Tsao, 1979, etc.), and a topic NP usually represents an entity that has been mentioned in the previous discourse and is being discussed again in the current sentence to add new information (Shi, 2000: 386). Since the identity of the topic NP is known to all interlocutors concerned, it has to be definite. Take examples in (3) for illustration. In (3a), though the NP *Yí-jìan shì* ‘one-CL matter’ represents something to be talked about, it cannot be introduced as a topic NP because it has unclear identity and violates the definiteness condition. In (3b), the topic NP *Zhè-jìan shì* ‘this-CL matter’ is definite, referring to the previously-mentioned event “my brother’s school-skipping”, and thus can be introduced at the beginning as the topic to be talked about in the following comment clause.

(3) (From Shi 2000: 387)

- a. ***Yí-jìan shì**, wó xiǎng gàosù mama t_i .
one-CL matter I want tell mother
‘One thing, I want to tell mother about it.’
- b. Dìdi jīntian táo-xué.
younger.brother today skip-school
Zhè-jìan shì, wó xiǎng gàosù mama t_i .
this-CL matter I want tell mother
‘My younger brother skipped school today.
As for this matter, I want to tell mother about it.’

Sometimes, generic NPs, usually represented by bare NPs, are used as topics when the speaker and the listener have shared world knowledge about the class of entities represented by the generic NP (Dahl, 1974; Givón, 1983). Thus, the topic, **Gǒu** ‘dog’, in (4) is a bare NP, and is ambiguous between a definite reading (i) and a generic reading (ii). The assumed relevant discourse contexts for these two different readings are presented in (4b) and (4c) respectively. (4b) describes a specific context where the referent is clearly introduced and can be referred to by the

topic, whereas (4c) is about general knowledge of the world and is usually not stated explicitly in conversations.

(4) (From Shi 2000: 387)

- a. **Gǒu**_i wó jiàn-guò t_i..
dog I see-ASP
(i) ‘The dog, I have seen.’
(ii) ‘Dogs, I have seen.’
- b. Ta bàbà mǎi-le yì-tiáo **gǒu**, yì-zhī mao.
Her father buy-ASP one-CL dog one-CL cat
‘Her father bought a dog and a cat.’
- c. Cháng-jiàn de chóngwù yǒu **gǒu**, mao hé tùzǐ.
often-see DE pet have dog cat and rabbit
‘Common pets include dogs, cats and rabbits.’

Syntactically, topic NPs are structurally related to a position inside the comment clause. In other words, the comment clause is an open clause with a gap position inside and the co-referential relation between the topic and the gap is subject to locality constraints (Shi, 2000). Within the framework of government-binding theory, previous research has shown that topic structures in Chinese are derived via *wh*-movement which obeys syntactic island constraints (Huang, 1982; Huang, Li, & Li, 2009; Li, 1990; Ning, 1993; Shi, 2000; Shyu, 1995), etc.).¹ For example, (5a) is a sentence with canonical word order in Chinese. (5b) is a topic-comment structure where the object, *lishi-xiaoshuo* ‘history-novel’, is topicalized to the sentence-initial position. When the topicalized NP is related to a position inside of a complex NP formed by a relative clause,

¹ It has been suggested that there are two types of topics in Chinese: moved topics (traces) and base-generated topics (PRO/pro) (Huang, Li, Li 2009). Some have argued that there are no dangling topics in Chinese and all topics are derived via movement (Shi, 2000; Huang & Ting, 2006, etc.), but others have argued for the existence of dangling topics and proposed a semantic-pragmatic account (Pan & Hu, 2008, etc.). Our study focused on topic sentences that have a clear empty position, and these topics are derived via syntactic movement, which is constrained by locality conditions.

as shown in (5c), the sentence is ungrammatical because it violates locality constraints.

(5) a. Canonical word order (no movement)

Jiaoshou zhidao jige xuesheng cengjing du-guo
Professor know several student ever read-ASP
 lishi-xiaoshuo.
history-novel

‘The professor knows that several students have read history novels before.’

b. Topicalization

Lishi-xiaoshuo_j, jiaoshou zhidao [jige xuesheng cengjing
History-novel professor know several students ever
 du-guo *t_j*].
read-ASP

‘As for history novels, professor knows that several students have read them before.’

c. Topicalization with CNPC violation

*Lishi-xiaoshuo_j, jiaoshou renshi [jige [*t_i* cengjing du-guo *t_j*
History-novel professor know several ever read-ASP
 de]_{CP} xuesheng_i]_{NP}
DE students

‘As for history novels, professor knows several students who have read them before.’

However, the degree of well-formedness between (5b) and (5c) is not so evident. Many native speakers actually consider both (5b) and (5c) to be worse than (5a), as people may have trouble understanding what the topic NP, *Lishi-xiaoshuo* ‘history-novel’, refers to. This is because the topic sentences in (5b) and (5c) appear in a null context, and they do not fulfill the discourse requirement that topic NPs are usually taken as old information referring to some previous-mentioned NP in the discourse context. Thus, the acceptability of these topic sentences could be affected due to the lack of a proper context.

In this study, an experiment was conducted to examine whether the presence or absence of context affects the acceptability judgment on topic sentences such as (5) in Mandarin Chinese. We empirically tested whether there is a contrast between (5b-5c) and (5a), and whether the contrast remains the same when they are situated in different contexts (null context vs. relevant context). The prediction is that if the same contrast is found consistently across different context conditions, this would suggest that discourse context does not interact with syntactic knowledge in acceptability judgment performance. On the other hand, if context plays a role in affecting the judgment performance, then the acceptability patterns for these sentences would vary in different context conditions. In addition, we also compare bare topic NPs and definite topic NPs to see if their difference plays a role in affecting judgment patterns. As shown in (4), bare topic NPs, but not definite topic NPs, are ambiguous as to whether they are of a definite reading or a generic reading. We predict that such ambiguity may reduce the acceptability of topic sentences, because people may exhibit more difficulty in constructing a coherent context and interpretation when the status of the topic NP is ambiguous.

2. EXPERIMENT METHODOLOGY

2.1 Participants

Eighty students from a public university participated in this experiment. In order to avoid the same participant being exposed to very similar target sentences repeatedly, we adapted a between-group design, and randomly assigned forty participants to the bare topic NP condition and forty participants to the definite topic NP condition. They were all native speakers of Mandarin Chinese spoken in Taiwan, and were either undergraduate or graduate students between 19 and 25 years of age. Each participant was paid NT \$50 for his/her participation, and the testing session lasted about 25 minutes.

2.2 Design and Materials

The experiment was a mixed design. The between-group factor was the type of topic – definite topic NP and bare topic NP. For each group, five conditions were constructed, including a control condition serving as the baseline and four other conditions with a 2 x 2 factorial design (context: null/relevant; topic: island-obeying/island-violating). A sample set of test paradigm is presented in Table 1.

Acceptability Judgment on Chinese Topic Constructions

Table 1. Test Paradigm

Conditions	Bare topic NP Condition	Definite topic NP Condition
Condition A Canonical Sentence (Control)	教授 知道 幾個 學生 曾經 讀過 歷史小說。 Jiaoshou zhidao jige xuesheng cengjing du-guo lishi-xiaoshuo Professor know several student ever read-ASP history-novel 'The professor knows that several students have read history novels before.'	教授 知道 幾個 學生 曾經 讀過 那部 小說。 Jiaoshou zhidao jige xuesheng cengjing du-guo na-bu xiaoshuo Professor know several student ever read-ASP that-CL novel 'The professor knows that several students have read that history novel before.'
Condition B -Null Context -Island-obeying	歷史小說 _{<i>t</i>} , 教授 知道 [幾個 學生 曾經 讀過 <i>t</i>] _{<i>i</i>} 。 Lishi-xiaoshuo _{<i>i</i>} , jiaoshou zhidao [jige xuesheng cengjing du-guo <i>t</i>] _{<i>i</i>} . History-novel professor know several students ever read-ASP 'As for history novels, the professor knows that several students have read them before.'	那部 小說 _{<i>t</i>} , 教授 知道 [幾個 學生 曾經 讀過 <i>t</i>] _{<i>i</i>} 。 Na-bu xiaoshuo _{<i>i</i>} , jiaoshou zhidao [jige xuesheng cengjing du-guo <i>t</i>] _{<i>i</i>} . That-CL novel professor know several students ever read-ASP 'As for that novel, the professor knows that several students have read it before.'
Condition C -Null Context -Island-violating	*歷史小說 _{<i>t</i>} , 教授 認識 [幾個 [<i>t</i> _{<i>i</i>} 曾經 讀過 <i>t</i> _{<i>j</i>} 的] 學生] _{<i>i</i>} 。 *Lishi-xiaoshuo _{<i>i</i>} , jiaoshou renshi [jige [<i>t</i> _{<i>i</i>} cengjing du-guo <i>t</i> _{<i>j</i>} de] xuesheng] _{<i>i</i>} . History-novel professor know several ever read-ASP. DE students 'As for history novels, the professor knows several students who have read them before.'	*那部 小說 _{<i>t</i>} , 教授 認識 [幾個 [<i>t</i> _{<i>i</i>} 曾經 讀過 <i>t</i> _{<i>j</i>} 的] 學生] _{<i>i</i>} 。 *Na-bu xiaoshuo _{<i>i</i>} , jiaoshou renshi [jige [<i>t</i> _{<i>i</i>} cengjing du-guo <i>t</i> _{<i>j</i>} de] xuesheng] _{<i>i</i>} . That-CL novel professor know several ever read-ASP DE students 'As for that novel, the professor knows several students who have read it before.'
Condition D -Relevant Context -Island-obeying	Context sentence 圖書館 上個月 添購了 許多 不同 種類的 小說。 Tushuguan shanggeyue tiangoule xudou butong zhonglei-de xiaoshuo. Library last-month order-LE many different kind novel 'The library ordered many different kinds of novels last month.' Target sentence 歷史小說 _{<i>t</i>} , 教授 知道 [幾個 學生 曾經 讀過 <i>t</i>] _{<i>i</i>} 。 Lishi-xiaoshuo _{<i>i</i>} , jiaoshou zhidao [jige xuesheng cengjing du-guo <i>t</i>] _{<i>i</i>} . History-novel professor know several students ever read-ASP 'As for history novels, the professor knows that several students have read them before.'	Context sentence 圖書館 上個月 添購了 一部 著名的 經典 小說。 Tushuguan shanggeyue tiangoule yi-bu zhumingde jingdian xiaoshuo. library last-month order-LE one-CL famous classics novel 'The library ordered one famous classical novel last month.' Target sentence 那部 小說 _{<i>t</i>} , 教授 知道 [幾個 學生 曾經 讀過 <i>t</i>] _{<i>i</i>} 。 Na-bu xiaoshuo _{<i>i</i>} , jiaoshou zhidao [jige xuesheng cengjing du-guo <i>t</i>] _{<i>i</i>} . that-CL novel professor know several students ever read-ASP 'As for that novel, the professor knows that several students have read it before.'
Condition E -Relevant Context -Island-violating	Context sentence 圖書館 上個月 添購了 許多 不同 種類的 小說。 Tushuguan shanggeyue tiangoule xudou butong zhonglei-de xiaoshuo. Library last-month order-LE many different kind novel 'The library ordered many different kinds of novels last month.' Target sentence *歷史小說 _{<i>t</i>} , 教授 認識 [幾個 [<i>t</i> _{<i>i</i>} 曾經 讀過 <i>t</i> _{<i>j</i>} 的] 學生] _{<i>i</i>} 。 *Lishi-xiaoshuo _{<i>i</i>} , jiaoshou renshi [jige [<i>t</i> _{<i>i</i>} cengjing du-guo <i>t</i> _{<i>j</i>} de] xuesheng] _{<i>i</i>} . History-novel professor know several ever read-ASP. DE students 'As for history novels, the professor knows several students who have read them before.'	Context sentence 圖書館 上個月 添購了 一部 著名的 經典 小說。 Tushuguan shanggeyue tiangoule yi-bu zhumingde jingdian xiaoshuo. Library last-month order-LE one-CL famous classics novel 'The library ordered one famous classical novel last month.' Target sentence *那部 小說 _{<i>t</i>} , 教授 認識 [幾個 [<i>t</i> _{<i>i</i>} 曾經 讀過 <i>t</i> _{<i>j</i>} 的] 學生] _{<i>i</i>} 。 *Na-bu xiaoshuo _{<i>i</i>} , jiaoshou renshi [jige [<i>t</i> _{<i>i</i>} cengjing du-guo <i>t</i> _{<i>j</i>} de] xuesheng] _{<i>i</i>} . That-CL novel professor know several ever read-ASP. DE students 'As for that novel, the professor knows several students who have read it before.'

The test targets with bare topic NPs are illustrated on the left column (*lishi-xiaoshuo* ‘history-novel’), and the test targets with definite topic NPs are illustrated on the right column (*na-bu xiaoshuo* ‘that novel’). Condition A is the control condition that contained a canonical sentence in a null context. Conditions B and C were null-context conditions. Condition B obeyed island constraints and Condition C violated island constraints. Conditions D and E were relevant-context conditions, and they each included a context sentence that introduced a possible referent

to be referred to by the topic NP in the target sentence. Condition D obeyed syntactic islands and Condition E did not.

For the bare topic NP conditions (left column, Table 1), in the relevant-context conditions (D/E), as shown in Table 1, the context sentence introduced the noun phrase, *butong zhonglei-de xiaoshuo* ‘different kinds of novels’. This provided a relevant context to help the readers interpret the bare topic NP, *lishi-xiaoshuo* ‘history-novel’, in the target sentence. Yet, the link between the context sentence and the topic NP was less direct and might need some kind of world knowledge to build up the connection. As for the definite topic NP conditions (right column, Table 1), the context sentence in the relevant-context conditions (D/E) introduced a specific noun phrase, *yi-bu zhumingde jingdian xiaoshuo* ‘one famous classical novel’, and it served as a clear referent for the definite topic NP, *na-bu xiaoshuo* ‘that-CL novel’, which appeared at the beginning of the target sentence.

Twenty sets of sentences with five conditions as shown in Table 1 were constructed. The test targets were distributed among five lists in a Latin-Square design such that each list contained twenty items from each condition. Forty filler sentences of similar length and complexity were constructed. They involved various types of legal and illegal structures. Half of them coupled with a context sentence and half without. Each participant saw one of the lists with 20 target items intermixed with the 40 filler items in a random order.

2.3 Procedure

The experiment was conducted on a laptop using the Psychology Software Tools “E-Prime 2.0”. The participants were tested individually in a quiet room. The task was a 7-point scale acceptability judgment task, with “one” referring to the least natural and least acceptable, and “seven” referring to the most natural and most acceptable. The 7-point scale was chosen because it was sensitive enough to measure the differences in participants’ responses. For relevant-context conditions, both the context sentence and the target sentence appeared on the screen at the same time, and the target sentence was underlined for the participants to rank its acceptability. The participants were asked to first read the sentence(s)

presented on the screen carefully, and then make judgment on these sentences as soon as they could. They needed to rate the target sentence according to its acceptability and naturalness by pressing the corresponding number key on the keyboard. The response time for making the decision was also recorded. After the participants pressed the number key, the target sentence disappeared and a new sentence appeared on the screen immediately. The participants had to evaluate whether the new sentence shared the same meaning with the target sentence they just ranked. This comprehension test was to ensure that the participants attended to the stimuli and understood the target sentences. Precise instructions and a practice session were given to the participants prior to the real experiment. Both the media of the instruction and the instructor's explanation were all in Mandarin Chinese.

3. EXPERIMENT RESULTS

The results section consists of three parts: the comprehension accuracy, the judgment rating score, and the judgment response time.

3.1 Comprehension Accuracy

Table 2 shows the mean comprehension accuracy for all conditions. The overall comprehension accuracy was above 95%. Two statistical analyses were carried out. First, Conditions A, B, and C were compared and analyzed to see if there was a difference in the comprehension accuracy between canonical sentences and topic-comment sentences when they appeared in a null context (just like the set of examples in (5)). A two-way mixed model ANOVA with topic type (definite/bare) as the between-participants factor and sentence type as the within-participants factor show that there was no effect of topic type ($F(1, 156) = 1.714, p = .194$), no effect of sentence type ($F(2, 156) = 1.274, p = .262$), and no interaction of the two factors ($F(2, 78) = .000, p = 1$).

Table 2. The averaged comprehension accuracy for each condition

Condition Topic Type	Condition A	Condition B	Condition C	Condition D	Condition E
	<i>Control (Canonical)</i>	<i>Null context Island-obeying</i>	<i>Null context Island-violating</i>	<i>Relevant context Island-obeying</i>	<i>Relevant context Island-violating</i>
Bare Topic NP	96.88% (0.10)	96.88% (0.08)	95% (0.12)	98.13% (0.07)	96.88% (0.08)
Definite Topic NP	98.13% (0.07)	98.75% (0.06)	96.25% (0.11)	96.25% (0.11)	96.88% (0.08)

Second, Conditions B, C, D, and E were compared and analyzed to check if participants' comprehension of island-obeying topic sentences and island-violating ones differed when they were situated in a null context and in a relevant context. A three-way mixed model ANOVA shows that there was no effect of topic type ($F(1, 78) = .093, p = .761$), no effect of sentence type ($F(1, 78) = 1.516, p = .222$), no effect of context ($F(1, 78) = 0.095, p = .759$), and no interactions ($F_s < 1$). In sum, the participants' comprehension accuracy was overall similarly good across all conditions.

3.2 Judgment Rating Scores

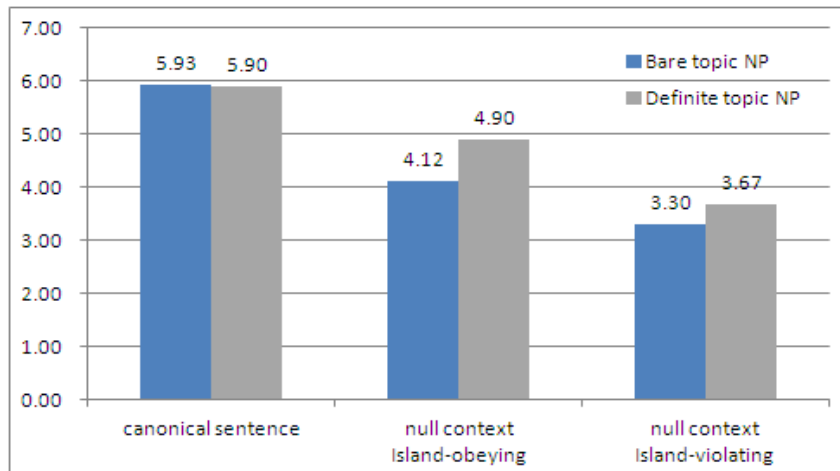
This section reports the results of the judgment rating scores. The incorrect responses in the comprehension test were removed from the following statistical analyses and the reported results are based on correct responses. The analyses were carried out for three sets of comparisons: null-context conditions vs. the control condition, relevant-context conditions vs. the control condition, and null-context condition vs. relevant-context conditions.

3.2.1 Null-context conditions vs. the control condition

This comparison aims to see how island-obeying/-violating topic sentences were rated in comparison to the canonical sentences when they appeared in a null context. Figure 1 presents the average rating scores for Condition A (Canonical sentence), Condition B (null context,

island-obeying) and Condition C (null context, island-violating), for both the bare topic NP group and the definite topic NP group.

Figure 1. The mean ratings for canonical sentence and null-context conditions



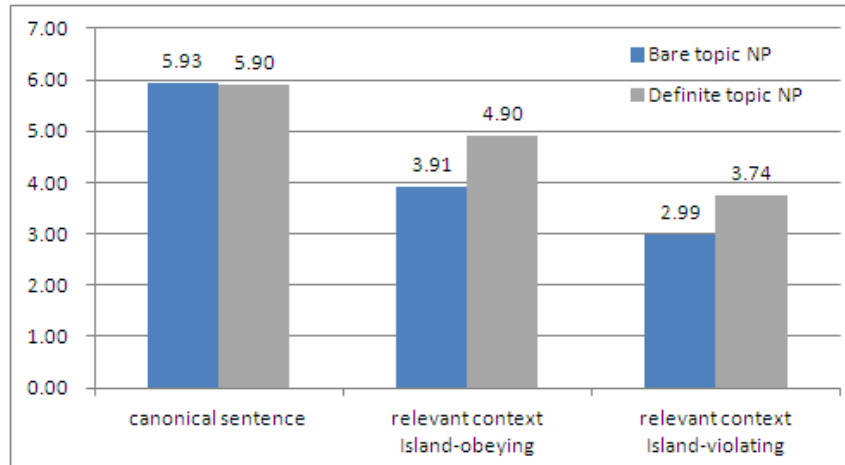
A two-way mixed model ANOVA with topic type as the between-participants factor and sentence type as the within-participants factor showed that there was significant interaction between topic type and sentence type ($F(2, 156) = 5.661, p < .004$), as suggested the different rating patterns for the three types of sentences in the bare topic NP group and in the definite topic NP group. First, the canonical sentences were rated similarly in both groups (5.93 vs. 5.90, $F(1, 78) = .041, p = .840$). Second, in a null context, the island-obeying topic sentences were rated higher in the definite topic NP group than in the bare topic NP group (4.90 vs. 4.12, $F(1, 78) = 10.692, p < .002$). Third, similarly, the rating for island-violating topic sentences was higher in the definite topic NP group than in the definite topic NP group, but the difference did not reach significance (3.30 vs. 3.67, $F(1, 78) = 2.294, p = .134$). Overall, the definite topic NP condition was rated significantly higher than the bare topic NP condition (4.82 vs. 4.45; $F(1, 78) = 4.673, p = .034$), and such effect suggests that the topic sentences with a

definite topic NP were rated better than those with a bare topic NP. Last but not least, there was a main effect of sentence type ($F(2, 156) = 204.05, p < .000$), and the planned pair-wise comparisons show that the canonical sentences were rated significantly higher than both the island-obeying topic sentences (5.91 vs. 4.51, $p < .000$) and the island-violating ones (5.91 vs. 3.48, $p < .000$). And, importantly, in the null context, island-obeying topic sentences were rated better than island-violating ones (4.51 vs. 3.48, $p < .000$).

3.2.2 Relevant-context conditions vs. the control condition

This comparison aims to see how island-obeying/-violating topic sentences were rated in comparison to the canonical sentence when they were situated in a relevant context. Figure 2 presents the average rating scores for Condition A (Canonical sentence), Condition D (relevant context, island-obeying) and Condition E (relevant context, island-violating). Overall, the patterns are very similar to the comparisons between the canonical sentence and the null-context conditions. The two-way mixed model ANOVA showed that there was significant interaction between topic type and sentence type ($F(2, 156) = 10.266, p = .000$). Within a relevant context, the island-obeying topic sentences were rated higher in the definite topic NP group than in the bare topic NP group (4.90 vs. 3.91, $F(1, 78) = 17.172, p < .000$); similarly, the island-violating topic sentences were rated higher in the definite topic NP group than in the bare topic NP group (3.74 vs. 2.99, $F(1, 78) = 8.81, p = .004$). That is, the topic sentences in the definite topic NP group were rated significantly higher than in the bare topic NP group (4.85 vs. 4.28; $F(1, 78) = 10.225, p = .002$).

Figure 2. The mean ratings for canonical sentence and relevant-context conditions



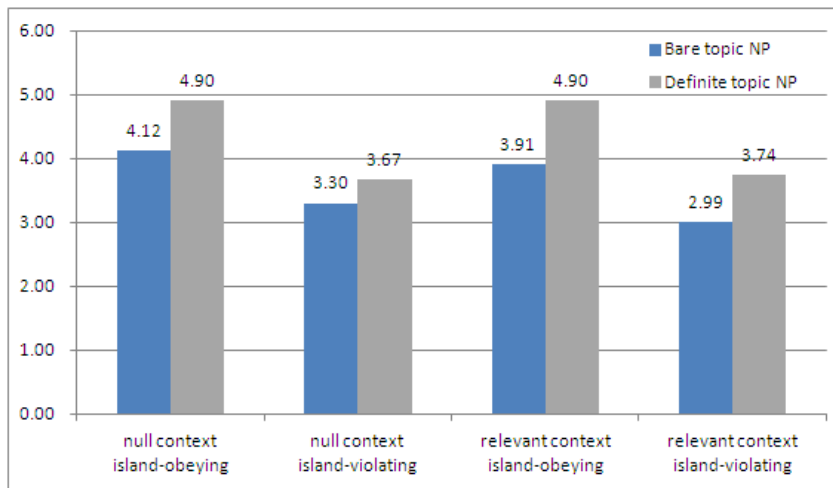
There was also a significant main effect of sentence type ($F(2, 156) = 375.78, p < .000$), and the planned pair-wise comparisons show that the canonical sentences were again rated significantly higher than the topic sentences in relevant context, both the island-obeying ones (5.91 vs. 4.41, $p < .000$) and the island-violating ones (5.91 vs. 3.37, $p < .000$). Importantly, in a relevant context, topic sentences that obey islands were still rated higher than those that violate islands (4.41 vs. 3.37, $p < .000$).

3.2.3 Null-context conditions vs. relevant-context conditions

This last comparison aims to verify whether the contrast between island-obeying topic sentences and island-violating topic sentences remains the same in different context conditions; in other words, this comparison allows us to see whether syntactic knowledge and discourse context interact with each other in affecting off-line sentence acceptability judgment performance. Figure 3 presents the average rating scores from Conditions B to E from both groups. A three-way mixed model ANOVA was conducted to test the effects of context (null/relevant), island violation (obey/violate), topic types (bare/definite),

and their potential interactions. The analysis shows there was no three-way interaction ($F(1, 78) = .563, p = .456$) and no two-way interaction between context and island-violation ($F(1, 78) = .023, p = .881$), but there were marginal interactions between context and topic type ($F(1, 78) = 2.914, p = .09$) and between island and topic type ($F(1, 78) = 3.105, p = .08$).

Figure 3. The mean ratings for null-context and relevant-context conditions



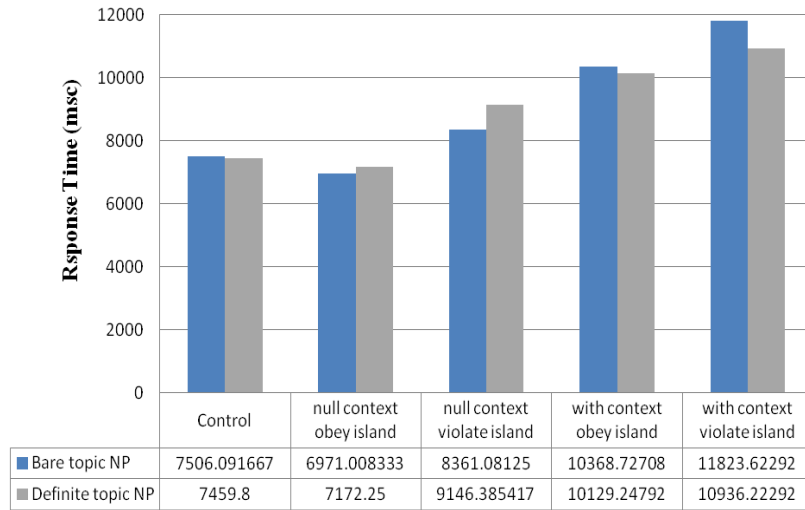
Further analyses were run to examine the interaction patterns. First, in both null-context and relevant-context conditions, the definite topic NP group was rated higher than the bare topic NP group ($F(1, 78) = 7.184, p = .009$; $F(1, 78) = 15.298, p = .000$). For the definite topic NP group, the ratings were similar in the null-context conditions and the relevant-context conditions ($F(1, 39) = .078, p = .782$). For the bare topic NP group, the ratings were higher in the null-context conditions than in the relevant-context conditions ($F(1, 39) = 5.519, p = .024$). Second, for the interaction between topic type and island, both island-obeying and island-violating topic sentences were rated better with a definite topic NP than with a bare topic NP ($F(1, 78) = 16.230, p < .000$; $F(1, 78) = 6.294, p = .014$). Moreover, in both topic types, the island-obeying topic

sentences were rated higher than the island-violating ones ($F(1, 39) = 66.066, p < .000$; $F(1, 39) = 56.211, p < .000$), with a larger contrast in the definite topic NP group than in the bare topic NP group (Mean difference: 1.198 vs. 0.898). To summarize, the island-obeying topic sentences were rated higher than the island-violating ones, and topic sentences with a definite topic NP were rated higher than those with a bare topic NP.

3.2 Judgment Response Time (RT)

The comparison of response time (RT) aims to see how syntactic knowledge and context may affect the process of a judgment behavior. Figure 4 presents the average response time of all conditions. Visually, the control sentences and the null-context, island-obeying topic sentences had a faster response time than the other three conditions.

Figure 4. The average response time of all test conditions.



A three-way mixed model ANOVA was run to test the effects of context (null/relevant), topic type (definite/bare), and island violation (obey/violate) on the RTs for judging topic sentences (Conditions B, C,

D, and E). The analysis shows that there was no three-way interaction ($F(1, 78) = 2.135, p = .148$), no two-way interaction between context and island ($F(1, 78) = 1.709, p = .195$), and no interaction between island and topic type ($F(1, 78) = 3.105, p = .082$). A significant interaction between context and topic type ($F(1, 78) = 3.876, p = .053$) was found. In the null context, the participants took longer to judge topic sentences with definite topic NPs than those with bare topic NPs ($F(1, 39) = 85.650, p < .000$). However, in the relevant context, the patterns were reversed. The topic sentences with bare topic NPs yielded higher RTs than those with definite topic NPs ($F(1, 39) = 37.368, p < .000$). This is probably related to the less coherent context created by the ambiguity in the bare topic NP conditions than in the definite topic NP conditions. Overall, the participants took longer to respond to the relevant-context conditions than the null-context conditions ($F(1, 78) = 116.901, p < .000$), and this is reasonable because it is necessary for people to take more time to integrate sentences within a relevant context. In addition, the island-obeying topic sentences were responded to faster than island-violating ones ($F(1, 39) = 25.653, p < .000$), suggesting that people take more time to judge sentences which violate syntactic knowledge.

4. DISCUSSION AND IMPLICATIONS

Our study empirically tested native speakers' acceptability judgments on topic sentences in Mandarin. The three main findings in our experiment are summarized in (6), and we discuss the implications of these findings in this section.

(6) a. Finding 1 about the effect of context:

No interaction was found between syntactic knowledge and discourse context in off-line acceptability judgment performance on topic sentences in Mandarin. Island-obeying topic sentences were consistently rated higher than their island-violating counterparts, no matter if the sentences were presented in a null context or a relevant context.

b. Finding 2 about the effect of topic type:

The topic sentences with a definite topic NP were consistently rated better than those with a bare topic NP. This pattern was found for both island-obeying and island-violating topic sentences, and in both null-context and relevant-context conditions.

c. Finding 3 about the comparisons with the canonical sentences:

The sentences with the canonical SVO word order served as the baseline in the experiment. They were consistently rated better than topic sentences, both the island-obeying and island-violating ones, and in both null and relevant contexts.

4.1 The Effect of Context on Syntactic Judgment

Finding 1 suggests that syntactic knowledge about island constraints for forming topic sentences and the availability of a relevant context for providing a clear identity for the topic NP do not interact with each other in off-line acceptability judgment on topic sentences in Mandarin Chinese. The original assumption is that, when presented in isolation, island-violating topic sentences are considered bad for two reasons: one is the violation of island constraints (syntactic knowledge) and the other is the lack of a clear referent mentioned in the previous discourse context. The reasoning of our experiment design is that by adding a relevant context which provides a clear identity of the topic NP, these island-violating sentences could be improved and judged better. However, as evidently shown in the results, the island-violating topic sentences were consistently judged worse relative to the island-obeying ones, no matter whether they were presented in a null context or embedded in a relevant context. In other words, island-violating topic sentences are judged worse because they do not conform to the syntactic knowledge for forming grammatical topic sentences, and adding relevant context does not improve their acceptability. This finding corroborates with Sprouse (2007a)'s study in English which shows that *wh*-questions that violate island constraints are consistently judged worse even when they are preceded by a relevant context. Thus, our finding adds another

piece of evidence to the claim that discourse context, an extra-grammatical factor, does not interact with syntactic knowledge in affecting the patterns of sentence acceptability judgment performance. This is so even with a construction that requires presupposition of a relevant context like topic constructions and in a language that heavily depends on context for sentence interpretation like Chinese.

In addition, our finding of the consistent contrast between island-obeying topic sentences and their island-violating counterparts across different types of contexts and different types of topic NPs supports the categorical approach to grammaticality. This is because such contrast suggests that participants are implicitly employing their grammatical knowledge and imposing categorical distinctions in judging given sentences. Supplying a relevant context with a clear referent to the topic NP does not alter the contrast between violation and non-violation sentences. That is to say, the syntactic island constraints for well-formed topic sentences could not be violated or else the resulting sentences would always be considered relatively worse than those that obey the constraint. Recently, some researchers have proposed gradience in grammar (Keller, 2000; Sorace and Keller, 2005; Fanselow, et al. 2004; Featherston, 2005a). They argue against using the binary judgment task, which assumes that grammaticality is categorical, and promote a seemingly more sensitive measure such as a magnitude estimation (ME) task to capture the gradience in linguistic data.² However, several recent studies have shown that the acceptability judgment data gathered by ME are not more informative than the data collected via the binary judgment task or the ordinal n-point scale judgment task (Bader & Häussler, 2010; Weskott & Fanselow, 2008, 2011). In addition, it has been shown that the cognitive assumptions of ME do not hold for participants in acceptability judgment experiments (Sprouse, 2011). Therefore, our data together with these previous findings seem to imply that even if

² In a magnitude estimation task, participants are presented with a pair of sentences for judgment every time. The first one is the reference sentence, and has a value associated with its acceptability (for example, 10). The acceptability of the second sentence can then be estimated using the acceptability of the first. If the sentence is two times more acceptable than the reference sentence, it would receive a value twice that of the reference (e.g. 20).

acceptability performance is a gradient and even when the assessment of sentences' well-formedness is measured on a gradient scale, the contrast between constraint-obeying sentences and constraint-violating sentences still emerges, suggesting that there is a categorical distinction in the grammar which participants implicitly utilize when judging sentences in general.

4.2 The Effect of Extra-grammatical Factors

Finding 2 and Finding 3 constitute evidence that, unlike grammaticality, which is categorical, acceptability performance has a wider range of variations which are associated with extra-grammatical factors like ambiguity, frequency, and processing complexity involved in the given sentences. Below, we discuss the implications of Finding 2 and Finding 3 separately.

First, Finding 2 shows that topic sentences with a bare topic NP are consistently rated worse than those with a definite topic NP. This suggests that grammar has limitation and that extra-grammatical factors like ambiguity and context coherence play a role in sentence acceptability judgment performance. In both definite topic NP and bare topic NP conditions, the same pattern of island-obeying topic sentences being rated better than island-violating ones were observed. This suggests that island constraints hold consistently, and any subtype of NPs that undergo a movement that violates the constraint would reduce the acceptability. However, grammar alone could not explain why topic sentences with a bare topic NP are consistently rated significantly worse than those with a definite topic NP. The pattern of Finding 2 is probably caused by the ambiguity and context coherence involved in the interpretation of the bare NP in Mandarin Chinese. As illustrated in (4) in the introduction, when the topic is a bare NP, it is ambiguous as to whether it is a definite reading or a generic reading, depending on the presupposed context where the topic sentence is uttered. When presented in a null context, the ambiguity of the bare topic NP is likely to cause difficulty for interpretation. To resolve the ambiguity, the participants have to construe a specific context for a definite reading or come up with some shared world knowledge for a generic reading, and then decide

which interpretation is more appropriate. The difficulty in interpretation and the process of resolving ambiguity may explain why topic sentences with a bare topic NP are judged less acceptable than those with a definite topic NP in a null context. On the other hand, when situated in a relevant context, the association between the context sentence and the bare topic NP is indirect, making the link between the context and the topic sentence less coherent. For example, in our paradigm in Table 1, the context introduced the noun phrase, *butong zhonglei-de xiaoshuo* 'different kinds of novels', denoting a set of different types of novels. This context sentence provided a general background, but did not provide a clear referent for the following bare topic NP, *lishi-xiaoshuo* 'history-novel', to be identified with, and such association requires some kind of world knowledge. The incoherent context thus creates difficulty in interpretation, and it depends largely on the readers' effort to come with a sensible reading. This may explain why bare topics are judged worse than definite topics when they are embedded in a context. The complication in interpreting bare NPs, whether in a null context or in a relevant context, is also supported by Kuo (2008)'s pragmatic approach to the interpretations of bare nouns in Mandarin. To sum up, the significant rating difference between the bare topic NP sentences and the definite topic NP sentences suggests that the ambiguity and context coherence can affect off-line acceptability judgment.

Second, Finding 3, which states that canonical sentences were rated better than all topic sentences, including the island-obeying topic sentences, suggests that ex-grammatical factors like frequency and processing complexity are involved in off-line acceptability judgment. In our experiment, sentences with canonical word order were included as a baseline in order to see how topic sentences are rated in comparison to their canonical counterparts. Our finding confirms the observation that, in a null context (as well as in a relevant context), island-obeying topic sentences are considered worse than their canonical counterparts. At first sight, this may appear strange because both types of sentences are grammatical according to the rules of the language. If grammar itself could not explain the obvious rating difference between the canonical sentences and the island-obeying topic sentences, then some extra-grammatical factors must get involved in affecting the performance

of the acceptability judgment. What would then be the factors to account for the rating difference between grammatical canonical sentences and grammatical topic sentences? Sentence frequency and processing complexity are probably playing a role here. The baseline sentence has a canonical SVO word order and is the unmarked form. The island-obeying topic sentence involves *wh*-movement of the object to a non-canonical position and is of a marked form. Since the marked forms are less frequent than the unmarked forms (Moravcsik, 2006), topic sentences are less frequent than canonical sentences. Past studies show that sentences with less frequent constructions are processed slower than the sentences with more frequent constructions (Gennari & MacDonald, 2008; Real & Christiansen, 2007; Wells, Christiansen, Race, Acheson, & Macdonald, 2009, etc.), and that frequency can positively affect off-line acceptability such that more exposure to the sentence structure would increase its acceptability ratings (Dąbrowska, 2010; Luka and Barsalou, 2005). With the support of these findings, it is clear that frequency should at least play some role here and explain why the grammatical topic sentences are rated as less acceptable than their canonical counterparts.

In addition to the frequency effect, the factor of processing complexity may also account for the difference between the canonical sentences and the grammatical topic sentences. Previous studies show that processing sentences involving *wh*-movement requires the completion of a filler-gap long-distance dependency, and consumes a great amount of computational resources (working memory), not only to hold the filler before the parser identifies the verb/gap later in the sequence but also to integrate the filler and the gap in a long distance to complete the structural dependency (Frazier, 1987; Frazier & d'Arcais, 1989; Gibson, 1998, 2000; Stowe, 1986, etc.). Thus, processing sentences with *wh*-movement is more demanding due to the long-distance dependency involved in the structure. In our case here, the canonical sentences do not involve *wh*-movement, while the topic sentences are derived via *wh*-movement. Thus, the processing of topic sentences is more difficult than the processing of canonical sentences because the former, but not the latter, is a more complex structure that involves a long-distance dependency. And, such kind of processing

complexity will negatively affect the off-line acceptability judgment, and decrease the acceptability of grammatical and ungrammatical sentences (Casasanto, et al, 2010; Sag, et al., 2007). Therefore, it is reasonable to partly attribute the differential acceptability between canonical sentences and topic sentences to the difference in processing complexity associated with these two types of sentences.

4.3 Implications and Conclusion

This study examines the off-line acceptability judgment performance and focuses on whether syntactic knowledge and discourse context may interact with each other in off-line acceptability judgment on topic sentences in Mandarin. Our empirical findings have several implications. First, based on Finding 1, discourse context does not interact with syntactic knowledge in the off-line acceptability judgment performance. This supports the view that grammaticality is categorical. Second, based on Finding 2 and Finding 3, people's acceptability performance, while dominated by their syntactic knowledge, would be affected by extra-grammatical factors like ambiguity, discourse coherence, frequency and processing complexity, and it therefore forms a continuity with a wide range of variations. The broader implication is that linguists' assumption that grammar is categorical is basically correct and using the traditional type of acceptability judgment task actually obtains fairly reliable data. This is supported by Sprouse & Almeida (2012) and Sprouse, Schütze, & Almeida, (submitted)'s findings which show that, for the data from a standard textbook on syntax and from a prestigious linguistics journal, the maximum discrepancy between traditional methods and formal experimental methods is only about 2~5%. Yet, psycholinguists' worry that acceptability judgment behavior displays a wider range of variations which are related to extra-grammatical factors such ambiguity, frequency, processing complexity, and incoherence should also be recognized. It is thus important for syntacticians to be careful when eliciting sentence judgments from their informants. To ensure that the contrast pattern in the judgment data is a result of grammar, linguists should not only pay attention to avoid extra-grammatical factors that may affect acceptability judgments, but

also provide a baseline for comparison. After all, we all want reliable data.

REFERENCES

- Altmann, Gerry T.-M. , Alan Garnham, and Judith-Ann Henstra. 1994. Effects of syntax in human sentence parsing: Evidence against a structure-based proposal mechanism. *Journal of Experimental Psychology: Learning, Memory and Cognition* 20.1: 209-216.
- Altmann, Gerry T. 1989. Parsing and interpretation: An introduction. *Language and Cognition Processes* 4: 1-19.
- Ambridge, Ben, and Adele E. Goldberg. 2008. The island status of clausal complements: Evidence in favor of an information structure explanation. *Cognitive Linguistics* 19.3: 349-381.
- Bader, Markus, and Jana Häussler. 2010. Toward a model of grammaticality judgments. *Journal of Linguistics* 46.2: 273-330.
- Bard, Ellen Gurman, Dan Robertson, and Antonella Sorace. 1996. Magnitude estimation of linguistic acceptability. *Language* 72.1: 32-68.
- Bernstein, Judy B., Wayne Cowart, and Dana McDaniel. 1999. Bare singular effects in genitive constructions. *Linguistic Inquiry* 30.3: 493-502.
- Casasanto, Laura Staum, Philip Hofmeister, and Ivan Sag. 2010. "Understanding acceptability judgments: Additivity and working memory effects." *Proceedings of the 32 Annual Conference of the Cognitive Science Society*, ed. by Richard Catrambone and Stellan Ohlsson, 224-229. Austin, Texas
- Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, Massachusetts: The M.I.T. Press.
- Chomsky, Noam. 1972. *Language and Mind*. New York: Harcourt Brace.
- Chomsky, Noam. 1975. *Reflections on Language*. New York: Pantheon Books.
- Clifton, Charles, Jr., Gisbert Fanselow, and Lyn Frazier. 2006. Amnestying superiority violations: Processing multiple questions. *Linguistic Inquiry* 37: 51-68.
- Cowart, Wayne 1997. *Experimental Syntax: Applying Objective Methods to Sentence Judgments*. London: Sage Publications.

- Crain, Stephen, and Mark Steedman. 1985. On not being led up the garden path: The use of context by the psychological parser. *Natural Language Processing: Psychological, Computational and Theoretical Perspectives*, ed. by David Dowty, Lauri Karttunen and Arnold Zwicky, 320-358. Cambridge: Cambridge University Press.
- Culicover, Peter W., and Ray Jackendoff. 2010. Quantitative methods alone are not enough: Response to Gibson and Fedorenko. *Trends in Cognitive Sciences* 14.6: 234-235.
- Dąbrowska, Ewa 2010. Naïve v. expert intuitions: An empirical study of acceptability judgments." *The Linguistic Review* 27: 1-23.
- Dahl, Östen. 1974. Topic-comment structure revisited. *Topic and Comment, Contextual Boundedness and Focus*, ed. by Östen Dahl, 1-24. Hamburg: Helmut Buske.
- Edelman, Shimon, and Morten H. Christiansen. 2003. How seriously should we take Minimalist syntax? *Trends in Cognitive Sciences* 7.2: 60-61.
- Fanselow, Gisbert, Caroline Féry, Matthias Schlesewsky, and Ralf Vogel. 2004. *Gradience in Grammar*. Oxford: Oxford University Press.
- Featherston, Sam. 2005a. That-trace in German. *Lingua* 115.9: 1277-1302.
- Featherston, Sam. 2005b. Magnitude estimation and what it can do for your syntax: some *wh*-constraints in German. *Lingua* 115.11: 1525-1550.
- Featherston, Sam. 2007. Data in generative grammar: The stick and the carrot. *Theoretical Linguistics* 33: 269-318.
- Ferreira, Fernanda. 2005. Psycholinguistics, formal grammars, and cognitive science. *Linguistic Review* 22: 365-380.
- Frazier, Lyn. 1987. Syntactic Processing: Evidence from Dutch. *Natural Language and Linguistic Theory* 5.4: 519-559.
- Frazier, Lyn, and G. B. Flores d'Arcais. 1989. Filler Driven parsing: A study of gap filling in Dutch. *Journal of Memory and Language* 28: 331-344.
- Gennari, Silvia P., and Maryellen C. MacDonald. 2008. Semantic indeterminacy and relative clause comprehension. *Journal of memory and language* 58: 161-187.
- Gibson, Edward. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition* 68: 1-76.
- Gibson, Edward. 2000. The dependency locality theory: A distance-based theory of linguistic complexity. *Image, language, brain*, ed. by Alec Marantz, Yasushi Miyashita and Wayne O'Neil, 95-126. Cambridge, MA: MIT Press.
- Gibson, Edward, and Evelina Fedorenko. 2010a. The need for quantitative methods in syntax and semantics research. *Language and Cognitive Processes* 28.1-2: 88-124.
- Gibson, Edward, and Evelina Fedorenko. 2010b. Weak quantitative standards in linguistics research. *Trends in Cognitive Sciences* 14.6: 233-234.

- Givón, Talmy. 1983. Topic continuity in discourse: An introduction. *Topic continuity in discourse: A quantitative cross-language study*, ed. by Talmy Givón, 1-42. Philadelphia: John Benjamins.
- Grodner, Daniel, Edward Gibson, and Duane Watson. 2005. The influence of contextual contrast on syntactic processing: Evidence for strong-interaction in sentence comprehension. *Cognition* 95: 275-296.
- Huang, C.-T. James. 1982. *Logical Relations in Chinese and the Theory of Grammar*. Ph.D. dissertation, Massachusetts Institute of Technology.
- Huang, C.-T. James. 1984. On the Distribution and Reference of Empty Pronouns. *Linguistic Inquiry* 15: 531-574.
- Huang, C.-T. James, Yen -Hui Audrey Li, and Yafei Li. 2009. *The Syntax of Chinese*. New York: Cambridge University Press.
- Huang, Ray Rui-heng, and Jen Ting. 2006. Are There Dangling Topics in Mandarin Chinese? *Concentric: Studies in Linguistics* 32.1: 119-146.
- Kaiser, Elsi, and John Trueswell. 2004. The role of discourse context in the processing of a flexible word-order language. *Cognition* 94: 113-147.
- Keller, Frank. 2000. *Gradience in grammar: Experimental and computational aspects of degrees of grammaticality*. Ph.D. dissertation, University of Edinburgh, Edinburgh, UK.
- Kuo, Jenny Yi-chun 2008. A pragmatic approach to the interpretations of Mandarin bare nouns. *Journal of Pragmatics* 40: 1082-1102.
- Li, Charles N., and Sandra A. Thompson. 1976. Subject and Topic: A New Typology of Language. *Subject and Topic*, ed. by Charles N. Li, 457-489. New York: Academic Press.
- Li, Charles N., and Sandra A. Thompson. 1981. *Mandarin Chinese: A Functional Reference Grammar*. Berkeley, CA: University of California Press.
- Li, Yen -Hui Audrey 1990. *Order and Constituency in Mandarin Chinese*. Dordrecht: Kluwer.
- Lin, Chien-Jer Charles. 2008. The processing foundation of head-final relative clauses. *Language and Linguistics* 9: 813-38.
- Luka, Barbara J., and Lawrence W. Barsalou. 2005. Structural facilitation: Mere exposure effects for grammatical acceptability as evidence for syntactic priming in comprehension. *Journal of Memory and Language* 52: 436-459.
- Moravcsik, Edith A. 2006. *An Introduction to Syntax: Fundamentals of Syntactic Analysis*. London; New York: Continuum.
- Myers, James. 2009a. Syntactic judgment experiments. *Language & Linguistics Compass* 3.1: 406-423.
- Myers, James. 2009b. The design and analysis of small-scale syntactic judgment experiments. *Lingua* 119.3: 425-444.

- Myers, James. 2012. Testing adjunct and conjunct island constraints in Chinese. *Language and Linguistics* 13.3: 437-470.
- Ning, Chunyan. 1993. *The Overt Syntax of Relativization and Topicalization in Chinese*. Ph.D. Dissertation, University of Southern California, Los Angeles, CA.
- Pan, Haihua, and Jianhua Hu. 2008. A semantic-pragmatic interface account of (dangling) topics in Mandarin Chinese. *Journal of pragmatics* 40.11: 1966-1981.
- Phillips, Colin. 2010. Should we impeach armchair linguists?. *Japanese-Korean Linguistics* 17, ed. by Shoishi Iwasaki, Hajime Hoji, Patricia M. Clancy and Sung-Ock Sohn, 49-64. Stanford, CA: CSLI Publications.
- Phillips, Colin, and Howard Lasnik. 2003. Linguistics and empirical evidence: Reply to Edelman and Christiansen. *Trends in Cognitive Sciences* 7.2: 61-62.
- Real, Florencia, and M. Christiansen. 2007. Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language* 57: 1-23.
- Ross, J. R. 1982. Pronoun Deletion Processes in German. Paper presented at The Annual Meeting of the Linguistics Society of America. San Diego, CA.
- Sag, Ivan A., Philip Hofmeister, and Neal Snider. 2007. Processing Complexity in Subadjacency Violations: The Complex Noun Phrase Constraint. *Proceedings of the Annual Meeting of the Chicago Linguistic Society* 43.1: 215-229.
- Schütze, Carson T. 1996. *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago, IL: University of Chicago Press.
- Shi, Dingxu. 2000. Topic and topic-comment constructions in Mandarin Chinese. *Language* 76.2: 383-408.
- Shyu, Shu-ing. 1995. *The syntax of focus and topic in Mandarin Chinese*. Ph.D. Dissertation, University of Southern California, CA.
- Sorace, Antonella, and Frank Keller. 2005. Gradience in linguistic data. *Lingua* 115:1497-1524.
- Spivey-Knowlton, Michael J., John C. Trueswell, and Michael K. Tanenhaus. 1993. Context effects in syntactic ambiguity resolution: Discourse and semantic influences in parsing reduced relative clauses. *Canadian Journal of Experimental Psychology* 47.2: 276-309.
- Sprouse, Jon. 2007a. *A program for experimental syntax: Finding the relationship between acceptability and grammatical knowledge*. Ph.D. Dissertation, Department of Linguistics, University of Maryland, College Park, MD.
- Sprouse, Jon. 2007b. Continuous acceptability, categorical grammaticality, and experimental syntax. *Biolinguistics* 1: 117-128.
- Sprouse, Jon. 2011. A test of the cognitive assumptions of magnitude estimation: Commutativity does not hold for acceptability judgments. *Language* 87.2: 274-288.

Chun-Chieh Hsu and Shelley Ching-Yu Hsieh

- Sprouse, Jon, and Diogo Almeida. 2012. Assessing the reliability of textbook data in syntax: Adger's Core Syntax. *Journal of Linguistics* 48: 609-652.
- Sprouse, Jon, and Diogo Almeida. To appear. The empirical status of data in syntax: A reply to Gibson and Fedorenko. *Language and Cognitive Processes*.
- Sprouse, Jon, Carson T. Schütze, and Diogo Almeida.. A comparison of informal and formal acceptability judgments using a random sample from Linguistic Inquiry 2001-2010. (submitted)
- Stowe, Laurie A. 1986. Parsing WH-constructions: evidence for on-line gap location. *Language and Cognitive Processes* 1.3: 227-245.
- Tsao, Feng-fu 1979. *A functional study of topic in Chinese: The first step towards discourse analysis*. Taipei: Student Book Co.
- van Berkum, Jos J. A., Colin M. Brown, and Peter Hagoort. 1999. Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language* 41: 147-182.
- Wells, Justin B., Morten H. Christiansen, David S. Race, Daniel J. Acheson, and Maryellen C. Macdonald. 2009. Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology* 58.2: 250-271.
- Weskott, Thomas, and Gisbert Fanselow. 2008. Variance and informativity in different measures of linguistic acceptability. *Proceedings of the 27th West Coast Conference on Formal Linguistics (WCCFL)*, ed. by Natasha Abner and Jason Bishop, 431-439. Somerville, MA: Cascadilla Press.
- Weskott, Thomas, and Gisbert Fanselow. 2011. On the informativity of different measures of linguistic acceptability. *Language* 87.2: 249-273.

Chun-Chieh Hsu (Corresponding author)
Department of Foreign Languages and Literature
National Tsing Hua University
cchsu@mx.nthu.edu.tw

Shelley Ching-Yu Hsieh
Department of Foreign Languages and Literature
National Cheng Kung University
shelley@mail.ncku.edu.tw

句法知識與文法之外的因素對
句子之合理性判斷的影響：
中文主題句型的實證研究

許淳潔¹ 謝菁玉²

¹國立清華大學

²國立成功大學

在這個研究中，我們探討句法知識和言談語境是否在句子之合理性判斷表現中互相影響。我們檢驗中文主題句型在不同語境中的合理性判斷的結果，並且檢測不同類型的主題名詞如何影響判斷的結果。我們發現不論在無語境或相關語境中，遵守孤島限制的主題句型都比違反孤島限制的主題句型被判斷為較合理。這樣的對比顯示句法知識和言談語境在句子之合理判斷表現中並不互相影響，並支持文法的絕對分類性。此外，我們也發現以限定名詞為主題詞的句子比以無修飾名詞當主題詞的句子被判斷為較合理，而做為基準線的一般句型也比合文法的主題句型被判斷為較合理。這些結果顯示句子合理性的判斷會因句子的歧意、頻率、及處理上的複雜度影響而產生許多的變化，因而表現是為連續性的。總括而言，我們的研究支持傳統上以合理性判斷所得的語料因為句法知識掌控了句子之合理性判斷的表現，但是我們也同提出文法之外的因素會對句子判斷表現造成影響，故而這些因素在收集判斷的語料時須納入考量。

關鍵字：合理性判斷、文法、中文主題句型