The Effect of Syntax-Semantics Mismatch on Referent Predictability: Evidence from Chinese

Wei Cheng (cheng22@email.sc.edu)
Linguistics Program, University of South Carolina
Columbia, SC 29208 USA

Man Yuan (77673817@qq.com)
Yinxing College of Hospitality Management, Chengdu University of Information Technology
Chengdu, 611743 China

Fernanda Ferreira (fernanda@sc.edu)
Psychology Department & Linguistics Program, University of South Carolina
Columbia, SC 29208 USA

Amit Almor (almor@sc.edu)
Psychology Department & Linguistics Program, University of South Carolina
Columbia, SC 29208 USA

Abstract

Previous studies on reference tracking have established the importance of semantic factors in affecting the likelihood of re-mentioning a referent. This paper extends this line of research by investigating the interaction between syntax and semantics in this process. We conducted a Chinese sentence-completion experiment and found that the degree of syntax-semantics mismatch affects a referent's likelihood of re-mention. The results thus support a theory relating a referent's salience in discourse to its likelihood of re-mention.

Keywords: Referent re-mention; accessibility; thematic roles; psychological verbs; Chinese *ba*

Introduction

Recent studies (e.g., Levy, 2008) show that language processing is constrained by linguistic constituents' predictability, i.e., the likelihood of being mentioned subsequently. In the field of reference processing, studies (e.g., Kehler et al., 2008) have established the importance of semantic factors in determining a referent's predictability. However, there are some controversies about the role of syntactic factors in this process (Fukumura & van Gompel, 2010; Kaiser et al., 2011a, b). In this paper, we report a reference production experiment investigating the possible interaction between syntactic and semantic factors that may affect a discourse entity's predictability.

Kehler and Rhode (2013) suggested a division of labor between semantics and syntax in reference tracking. On the one hand, semantic factors such as verb meaning determine the coherence relation between sentences, which affects comprehenders' predictions about the likelihood of mentioning referents. On the other hand, syntactic factors such as grammatical roles mainly affect the form of reference and have little influence on referent predictability.

Kehler et al. (2008) discussed several types of coherence relations which exert different influence on referent predictability. For example, in an Explanation coherence relation as indicated by the connective *because*, stimulus-experiencer (SE) verbs such as *scare* in (1) produce a continuation bias towards *Mary*.

(1) Mary scared Sara because …

Such preference reflects the verb's implicit causality which biases reference towards the cause of an event (namely, the stimulus thematic role) under an Explanation coherence relation. Replacing the connective *because* in (1) with *so* would change the discourse relation from Explanation to Result, therefore changing the referential bias towards the experiencer of the event, i.e., *Sara* in (1).

Although the semantic factors mentioned above present comprehenders a strong cue about the upcoming referent, structural factors are argued to have a small effect on likelihood of mention. For example, Fukumura and van Gompel (2010) found no syntactic effect on a referent's likelihood of re-mention. Instead, they confirmed that re-mention bias is mainly due to semantic factors such as verb meaning and discourse coherence relation. Based on the results, they argued that syntactic factors influence an entity's accessibility in the mental discourse model but semantic factors do not. In other words, they suggested a dissociation between referent accessibility and predictability.

However, research by Kaiser and colleagues (2011a, b) shows that syntax matters in referent predictability. In an aural story-continuation study with agent-patient verbs, they found that in active sentences such as (2a), participants were more likely to continue the sentence with reference to the patient thematic role (i.e., *Lisa*) than agent (i.e., *Mary*). However, when (2a) is passivized as in (2b), the patient preference observed in (2a) disappeared, with the agent and the patient almost equally predictable from the context.

(2) a. Mary slapped Lisa at the zoo. As a result …
   b. Lisa was slapped by Mary at the zoo. As a result …

Kaiser et al. (2011a, b) attributed their findings to a mismatch between syntactic semantic roles. The agent thematic role is more prominent than the patient thematic role. Therefore, in the passive structure, when the agent *Mary* is demoted to the object position of the *by*-phrase, a syntactically less prominent position, a syntax-semantics mismatch happens. According to them, such mismatch would draw comprehenders' attention to the otherwise-less-probable referent (namely, the agent *Mary)* and hence increase its possibility of re-mention. In active sentences like (2a), by contrast, syntactic functions align with thematic roles in terms of prominence. The agent is projected in the subject position, a syntactically prominent position, therefore no mismatch. As a result, referent predictability was not affected.

Since (2a) and (2b) had the same coherence relation, the findings of Kaiser et al. (2011a, b) run against the coherence-based account for referent re-mention as proposed by Kehler and colleagues (2008, 2013). However, they are compatible with Arnold's (2001) Expectancy Hypothesis which argues that accessible entities have a high level of predictability. Antecedent accessibility can be further interpreted as activation in the mental discourse model, the strength of which is modulated by various discourse contexts including syntactic and semantic factors (Arnold, 2010). Under this account, the syntax-semantics mismatch in (2b) makes the agent *Mary* more activated and hence more accessible in comprehenders' discourse model, increasing its probability of re-mention.

This activation-expectancy account (Arnold, 2001, 2010) raises the question of whether the *degree* of syntax-semantics mismatch plays a role in determining referent predictability. According to Arnold (2010), activation is a graded notion. Meanwhile, syntax-semantics mismatch can also be interpreted as a gradient phenomenon, i.e., a continuum from no mismatch to a high degree of mismatch. If a sentence has no mismatch, the referents' activation will not be affected. However, any increase on the mismatch continuum may boost a referent's activation, leading to higher predictability. Crucially, this predication is different from that of Fukumura and van Gompel (2010) who assumed a separation between accessibility and predictability such that any change in referent accessibility will not affect the referent's likelihood of re-mention.

Another unsolved question in previous literature is whether such syntax-semantics mismatch effect in referent predictability works with other types of verbs, such as SE verbs. Kaiser et al. (2011a) included SE verbs in their study but did not find any effect of mismatch. This is expected because they used items such as (3).

(3) a. Mary annoyed Lisa at the zoo. As a result …
　　b. Lisa was annoyed by Mary at the zoo. As a result …

According to Kaiser et al., the experiencer thematic role is more prominent than the stimulus thematic role. Therefore,

in the passive structure (3b), no mismatch occurs in that the experiencer *Lisa* is located in the subject position, a syntactically prominent position. By contrast, there is syntax-semantics mismatch in (3a), as the experiencer is demoted into the object position. Such mismatch did not affect story continuation because the discourse coherence in this sentence also produces a referential bias towards the experiencer. Therefore, the absence of mismatch effect may be due to a ceiling effect of the experiencer bias.

In a written story-continuation task, Rhode and Kehler (2013) used stimuli like (3) but without connectives. They found an effect of mismatch: there were more continuations with the stimulus (i.e., *Mary*) in the passive than in the active (76% vs. 59%). In other words, there was a boost towards reference to the experiencer *Lisa* in the active mismatch condition. However, one potential problem with their study is that they didn't specify the coherence relation between the prompt sentence and the continuation sentence. Therefore, their results were confounded by the different possible coherence relations participants could engage in.

In this paper, we addressed these questions in a sentence-completion experiment by using Chinese SE verbs. A special property of Chinese sentence structure makes it possible to manipulate the degree of syntax-semantics mismatch. Many Chinese active sentences can be expressed in two ways, as shown in (4). In a canonical active structure such as (4a), the experiencer *Fupeng* is projected after the main verb. By contrast, with the use of a dummy word *ba* as in (4b), the experiencer comes before the main verb.

(4) a. Active canonical
　　　Dengxiang jinu-le　　　Fupeng.
　　　Dengxiang anger-ASP Fupeng
　　　'Dengxiang angered Fupeng.'
　　b. Active *ba*
　　　Dengxiang ba Fupeng jinu-le.
　　　Dengxiang ba Fupeng anger-ASP

Following previous proposals (e.g., Li, 2006), we assume that the post-*ba* NP is generated in the postverbal object position and then moves up to its surface position. Psycholinguistic studies (e.g., Bever & McElree, 1988) have found that syntactically moved element is more prominent for comprehenders. Therefore, we can assume that the object NP is syntactically more prominent following *ba* than the main verb. Given this, although in both (4a) and (4b) the experiencer *Fupeng* (object NP) is in a less prominent position than the stimulus *Dengxiang* (subject NP), resulting in a syntax-semantics mismatch, the *ba* structure would involve a lower level of mismatch than the canonical structure. Such difference enables us to directly examine the effect of degree of syntax-semantics mismatch on referent predictability.

Experiment

The experiment focused on SE verbs in Chinese.[1] We manipulated syntax-semantics mismatch through three kinds of structures: active canonical structure, active *ba* structure and passive structure. In the actives such as those in (4) above, the stimulus thematic role is the subject, while the experiencer thematic role is the object. There is a syntax-semantics mismatch in actives, as the experiencer that is more prominent than the stimulus in the thematic hierarchy occupies the object position, a syntactically less prominent position than the subject position (Gordon & Hendrick, 1998). Moreover, as discussed above, the extent to which syntax and semantics are mismatched is different between these two types of active sentences, with the mismatch less serious in the *ba* structure than the canonical structure. However, when the sentence is passivized as in (5), the mismatch disappears as the experiencer is located in a more prominent subject position. Therefore, the three types of structures enabled us to manipulate (the degree of) syntax-semantics mismatch.

(5) Fupeng bei Dengxiang jinu-le.
    Fupeng by Dengxiang anger-ASP
    'Fupeng was angered by Dengxiang.'

We used the connective *yinwei* 'because' to control the coherence relation of discourse. When followed by *because*, SE verbs produce a strong reference bias towards the stimulus rather than the experiencer (Stevenson et al., 1994).

If the likelihood of re-mention is determined by discourse coherence relation (Kehler et al., 2008; Kehler & Rhode, 2013), we would not expect to see a re-mention difference among the three structures because all of them had the same coherence relation pointing to the stimulus as the potential referent.

By contrast, if predictability is determined by accessibility or activation (Arnold, 2001, 2010), we would expect to observe a re-mention difference among the three structures due to different degrees of syntax-semantics mismatch. Specifically, since there is no mismatch in the passive, we would expect to see in this structure a strong preference for the discourse-biased entity (namely, the stimulus). On the contrary, in actives where there is mismatch, the preference for the stimulus may be compromised with a boost of reference to the experiencer which is more activated due to mismatch. Moreover, since mismatch is less serious in the *ba* structure than the canonical structure as argued above, there would be more instances of reference to the stimulus in the *ba* structure than the canonical structure.

Method
Participants

---

[1] Unlike English, Chinese has a limited set of SE verbs (Wen, 2006). Those which can occur with *ba* are usually a resultative compound composed of a morpheme denoting an action and another morpheme denoting the result of this action.

Fifty-one students in a university in China took part in the study for extra credit. All of them were native speakers of Mandarin Chinese.

Materials and Design
There were 18 test items in the experiment. All of them were sentence fragments containing an SE verb with two human characters of the same gender followed by a connective *yinwei* 'because'. In other words, we used a free prompt (i.e., no pronoun) to elicit participants' continuation. The items were counterbalanced in terms of gender such that half items had both female entities and half had both male entities. All items appeared in three types of structures (i.e., active canonical, active *ba*, and passive), making a total of 54 experimental stimuli. Sample stimuli are given below. We also constructed 18 filler items which contained other types of verbs (e.g., *xihuan* 'like', *piping* 'criticize') and connectives (e.g., *ranhou* 'then', *danshi* 'but').

(6) a. Active canonical
    Dengxiang jinu-le   Fupeng, yinwei …
    Dengxiang anger-ASP Fupeng because
    "Dengxiang angered Fupeng because …"
   b. Active *ba*
    Dengxiang ba Fupeng jinu-le,   yinwei …
    Dengxiang ba Fupeng anger-ASP because
    "Dengxiang angered Fupeng because …"
   c. Passive
    Fupeng bei Dengxiang jinu-le,   yinwei …
    Fupeng by Dengxiang anger-ASP because
    'Fupeng was angered by Dengxiang because …'

We used a Latin-Square design to divide the 54 test stimuli into three lists such that every item only appeared in one condition per list. The test stimuli and fillers were pseudo-randomized with one filler between experimental stimuli. To counterbalance the effect of trial order, three reverse lists were also constructed. The 51 participants were randomly assigned to each list, with three lists having eight participants and three lists having nine participants.

Procedure
The experiment was administered in a class by the second author. Participants were given a booklet and asked to produce a meaningful and natural continuation to the sentence fragment.

Coding and Analysis
The second author who is a native speaker of Chinese coded the data first. Afterwards, another trained coder who was a native Chinese speaker and naive to the purpose of this study coded the data independently. The coders noted whether the continuation started with reference to the first or the second NP. The responses which had ambiguous reference or reference to other entities than the two characters in the main clause were coded as unclear. Overall, the two coders achieved a coding agreement rate of 97%. The inconsistently coded responses ($N = 28$) were

excluded from analysis. Thirty-four additional responses, which were coded as unclear, were also excluded. Such data trimming made one participant's valid response rate very low (28%), and this participant was thus eliminated, leaving us the data of 50 participants. Overall, 7.4% of all responses ($N = 909$)[2] were eliminated.

As mentioned above, there is a strong re-mention bias towards the stimulus when a sentence containing SE verbs is followed by *because* (e.g., Stevenson et al., 1994). Since our major concern was whether syntax-semantics mismatch would affect this re-mention bias, we scored participants' reference to either the stimulus or the experiencer.[3] Because the choice between the stimulus and the experiencer is binary, we used mix-effects logit models to analyze our data. These models are suitable for analyzing nominal data (Jaeger, 2008).

Results

We first analyzed whether the list had an effect on the results. We fitted a mixed logit model with structure (active canonical, active *ba*, and passive) and list as fixed effects, and subject and item as random effects with structure and order as random slopes respectively. This model showed no effect of list and no interaction between structure and list (p's > .1).

Since the list did not have an effect on our data, we collapsed the data across the six lists and analyzed the data with a focus on the effect of structure. As Figure 1 shows, participants started the continuation with a strong preference for the stimulus thematic role in all three types of structures (>80%). In other words, the stimulus had a high predictability. However, it can also been seen from Figure 1 that there was a difference among the proportions of reference to the stimulus among the three structures. The passive structure had the highest likelihood of re-mention for the stimulus, the active canonical structure had the lowest likelihood, and the active *ba* structure fell in between.

A mixed logit model with structure as a fixed effect, and subject and item as random effects with no slope, showed a main effect of structure. Specifically, the stimulus was more likely to be referred to in the active *ba* structure (87%) than the active canonical structure (81.2%) ($\beta = -0.54$, SE = 0.27, $p < .05$). However, no significant difference was found between the active *ba* structure (87%) and the passive structure (89.5%) ($\beta = 0.25$, SE = 0.30, $p = .41$), although there was a trend for passives to have a stronger preference for the stimulus compared with active *ba* structures. Thus,

both the passive and active *ba* structures demonstrate a stronger stimulus bias than the active canonical structure.

Alternatively, this pattern can be interpreted from the perspective of the experiencer thematic role. That is, compared with the passive and *ba* structures, there was a boost of reference to the experiencer in the active canonical structure (see Figure 1).
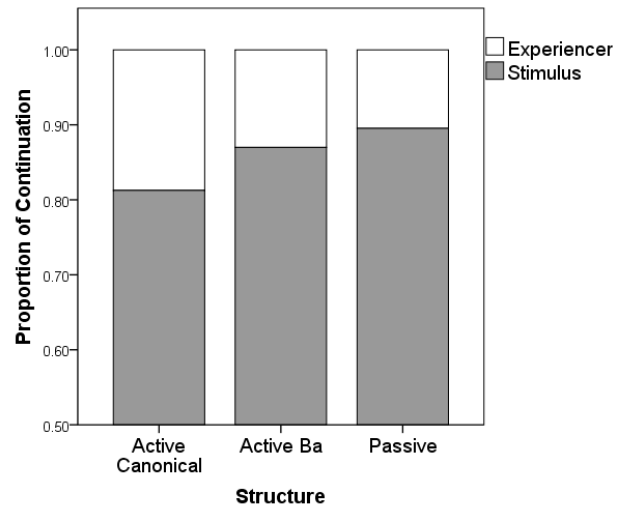

Figure 1: Proportion of continuation reference

Discussion

In this paper, we investigated whether the interaction between syntax and semantics, particularly the mapping between syntactic prominence and thematic prominence, affects the likelihood that a referent will be re-mentioned in subsequent discourse. Moreover, we were interested in whether such effect is modulated by the degree of mismapping between syntax and semantics. By focusing on Chinese SE verbs, we found that although the passive and the active *ba* structure did not differ significantly in terms of the likelihood of re-mentioning the stimulus, both of them had a stronger preference for the stimulus than active canonical structures. Overall, this means that the degree of syntax-semantics mismatch has an effect on referent predictability.

One may wonder whether the discrepancy between passive and active sentences was due to the factor of recency. In passives, the stimulus (the discourse-biased entity) was more recent to the connective *because* in its surface position than the experiencer. Thus, the factors of discourse bias and recency coincided in passives. By contrast, in actives the experiencer was more recent than the stimulus. Although intuitively appealing, this explanation cannot fully account for our results. In our items, the experiencer was more recent than the stimulus in both the active canonical and *ba* structures, and yet we observed a higher probability of reference to stimulus in the *ba* structure than in the canonical structure. Therefore, we suggest that recency is not a crucial reason for our results.

---

[2] The number of responses did not match the number of total trials (18×51=918) because nine trials were not answered by participants.

[3] When coding the data, the coders coded whether participants referred to NP1 or NP2. Note that the surface position of an entity does not align up with its thematic role across conditions. In active canonical and *ba* structures, NP1 is the stimulus and NP2 is the experiencer. In the passive structure, NP1 is the experiencer and NP2 is the stimulus.

Rather, we take the results as a reflection of the effect of syntax-semantics interaction on referent predictability. In the active sentences in our study, syntactic prominence does not match semantic prominence, as the experiencer which is semantically more prominent than the stimulus (Grimshaw, 1990) is located in the object position. In passives, on the contrary, there is no mismatch. Accordingly, our results showed that sentences in passive voice had a stronger re-mention bias towards the stimulus (i.e., the discourse-biased entity) than active sentences with a canonical order. Thus, in the latter structure, there was a boost of reference to the experiencer, an entity that was syntactically demoted but semantically prominent. Such syntax-semantics interaction effect was also noted by Ferreira (1994) who found that when presented with SE English verbs, participants were more likely to produce passives such that semantic prominence aligned with syntactic prominence.

Therefore, our results extend Kaiser et al.'s (2011a, b) findings on agent-patient verbs to SE verbs and lend support to their argument that syntax-semantics mismatch affects a referent's re-mention probability.

In addition, a more important finding of our study is that the degree of syntax-semantics mismatch affects an entity's likelihood of re-mention. We manipulated the degree of mismatch by alternating active items between canonical and *ba* constructions. Although in both structures the experiencer is in a syntactically less prominent object position, resulting in syntax-semantics mismatch, the degree of such mismatch varies between the two structures. As argued in Introduction, the syntactic prominence of the experiencer is higher in the *ba* construction than in the canonical construction, which leads to a lower degree of mismatch in the former. Our results showed that there was a significant difference in a referent's re-mention probability between *ba* and canonical structures. Specifically, we found that compared with *ba* structures, there was a boost of reference to the experiencer in canonical structures. This indicates that the degree of mismatch influences participants' expectations about upcoming entity: the higher level of mismatch, the more references to the otherwise-less-probable entity.

Overall, our results cast doubt on the coherence-driven theory of referent predictability (Kehler et al., 2008; Kehler & Rhode, 2013). According to this theory, whether a referent will be mentioned again in subsequent discourse is determined by the coherence relation of current discourse. Since all the three structures in our materials had the same SE verb and subordinate connective (*because*), they should have the same coherence relation, namely, the Explanation relation in Kehler et al.'s theory. Under this account, therefore, no difference in continuation bias is expected among the three types of sentences. However, as our results showed, there was a significant re-mention difference between the active canonical structure and the two other structures. This discrepancy indicates that discourse coherence relation is not the only factor behind comprehenders' calculation of potential referent.

However, we do not imply that Kehler and colleagues' (2008; 2013) reference processing theory is wrong. In their theory, coherence is involved in the process of predicting upcoming referent, while structural factors are involved in the process of integration. Our results only suggest that the top-down process of prediction involves not only coherence relations as claimed in their original proposal but also some other factors such as syntax-semantics mismatch.

As mentioned in Introduction, the effect of degrees of syntax-semantics mismatch on referent predictability is compatible with Arnold's (2001, 2010) proposal but not with Fukumura and van Gompel's (2010) account. Arnold argued that a referent's predictability is determined by its accessibility or activation in the mental discourse model. Syntax-semantics mismatch is a gradient phenomenon, ranging from no mismatch to a high degree of mismatch. The different degrees of mismatch may make certain entities more or less accessible, resulting in different possibilities of re-mention. By contrast, Fukumura and van Gompel suggested that accessibility and predictability are dissociated, with the former being affected by syntactic factors and the latter by semantic factors. Thus, their theory predicts that the degree of syntax-semantics mismatch has little effect on the referent's likelihood of re-mention.

Our findings confirmed the predictions of Arnold's (2001, 2010) theory but ran against those of Fukumura and van Gomel (2010). In our data, the active canonical structure had the highest level of mismatch, leading to a big increase in referent activation. By contrast, the *ba* structure had a lower level of mismatch, thereby a moderate activation. Accordingly, compared with the *ba* structure, we saw a boost of reference to the experiencer thematic role in the canonical structure. In other words, the experiencer became more predictable in the canonical structure because it was more activated due to a higher level of mismatch in this structure than in the *ba* structure.

One of the results which appeared to be different from the predications of the activation-expectancy account is that the experiencer was not more predictable in the *ba* structure than the passive structure. Given that there was a mismatch in the *ba* construction but no mismatch in the passive, a difference in referent activation and hence predictability was expected. However, the data did not show any significant difference. In the following, we suggest a possible reason for this discrepancy.

On the one hand, due to mismatch, the experiencer was more activated in the *ba* structure than in the passive. On the other hand, entities in the subject position are more accessible than those in lower positions in the syntactic tree (Gordon & Hendrick, 1998). Since the experiencer was in the subject position in the canonical structure but in the post-*ba* position in the *ba* structure (see (7) below), it was more activated in the former than the latter. Therefore, the activation difference between the two structures was cancelled out, making the experiencer equally predictable in both structures. Such activation difference also existed between the canonical structure and the passive. However, it

was not cancelled out because the mismatch in the canonical structure was so big that it greatly boosted the activation of the experiencer in this structure, leading to a higher re-mention bias.

(7) a. Active *ba*
   Dengxiang ba Fupeng jinu-le.
   Dengxiang ba Fupeng anger-ASP
   'Dengxiang angered Fupeng.'
   b. Passive
   Fupeng bei Dengxiang jinu-le.
   Fupeng by Dengxiang anger-ASP
   'Fupeng was angered by Dengxiang.'

One remaining question is why syntax-semantics mismatch affects the experiencer but not the stimulus. After all, the mismatch not only happens to the experiencer but also the stimulus. In the active sentences, the stimulus which is a thematically less prominent entity is located in the subject position, a syntactically prominent position. Thus, if mismatch has any effect, the stimulus should also be affected. That is, like the experiencer, its mental representation should also be more activated, resulting in higher predictability. However, our results did not show any increase in the predictability of the stimulus proportional to the increase in degree of mismatch. Rather, in the active canonical structure which had the highest level of syntax-semantics mismatch, we saw the lowest chance of reference to the stimulus.

We suggest two possible reasons. First, it could be due to the inherent difference between the two thematic roles. That is, the effect of mismatch only applies to a thematically more prominent entity, i.e., the experiencer in our case.

Alternatively, it could be that there is no essential difference between these two thematic roles with regard to whether they will be affected by mismatch. It may be that their asymmetry in our data was due to more general cognitive factors. For example, there may be a threshold for activation such that once an entity is activated to a certain level its representation will not be further boosted. In our materials, the stimulus was the default candidate for re-mention due to discourse bias, and hence might already be activated above the threshold. Therefore, although its syntactic function did not match its thematic role, such mismatch would not lead to a stronger activation. This may explain why the reference to the stimulus did not increase in the active sentences compared with the passives.

Summing up, by looking at Chinese SE verbs we confirmed previous findings on the interaction between syntax and semantics and its effects on referent predictability. Moreover, we found that the degree of mismatch between syntactic and thematic prominence affects the likelihood of re-mention. Such mismatch affects predictability because it modulates a referent's levels of accessibility and activation. Our data highlight the need to look at referent representation and processing from both a linguistic perspective and a more general cognitive perspective.

References

Arnold, J. E. (2001). The effect of thematic roles on pronoun use and frequency of reference continuation. *Discourse Processes, 31*, 137–162.

Arnold, J. E. (2010). How speakers refer: the role of accessibility. *Language and Linguistic Compass, 4*, 187-203.

Bever, T., & McElree, B. (1988). Empty categories access their antecedents during comprehension. *Linguistic Inquiry, 19,* 35-43.

Ferreira, F. (1994). Choice of passive voice is affected by verb type and animacy. *Journal of Memory and Language, 33,* 715-736.

Fukumura, K., & van Gompel, R. P. G. (2010). Choosing anaphoric expressions: Do people take into account likelihood of reference? *Journal of Memory and Language, 62,* 52–66.

Gordon, P. C., & Hendrick, R. (1998). The representation and processing of coreference in discourse. *Cognitive Science, 22,* 389-424.

Jaeger, F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59,* 434–46.

Kaiser, E., Holsinger, E., & Li, D. C.-H. (2011a). Who next? Consequences of syntax-semantics mismatches for likelihood of mention. Poster presented at CUNY, Stanford, CA.

Kaiser, E., Li, D. C.-H., & Holsinger, E. (2011b). Exploring the lexical and acoustic consequences of referential predictability. In I. Hendrickx, A. Branco, S. Lalitha Devi, & R. Mitkov (Eds.), *Anaphora Processing and Applications* (pp. 171-183). Heidelberg: Springer.

Kehler, A., Kertz, L., Rohde, H., & Elman, J. L. (2008). Coherence and coreference revisited. *Journal of Semantics, 25*, 1–44.

Kehler, A., & Rohde, H. (2013). A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation. *Theoretical Linguistics, 39,* 1-37.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106,* 1126-1177.

Li, Y.-H. A. (2006). Chinese *ba*. In M. Everaert & H. C. van Riemsdijk (Ed.), *The Blackwell Companion to Syntax (Vol. 1)*. Malden, MA: Blackwell Publishing.

Rohde, H., & Kehler, A. (2013). Grammatical and information-structural influences on pronoun production. *Language and Cognitive Processes*. Advance online publication. doi: 10.1080/01690965.2013.854918.

Stevenson, R. J., Crawley, R. A., & Kleinman, D. (1994). Thematic roles, focus and the representation of events. *Language and Cognitive Processes, 9*, 519–548.

Wen, Y. (2006). *A Study on Modern Chinese Psychological Verbs*. Doctoral dissertation, Beijing Language University, Beijing.