

1. Introduction

Traditional *offline*¹-based second language (L2) research in the past decades specifically concentrated on exploring non-native speakers' linguistic knowledge/competence in the L2, aiming to know how their knowledge about the L2 develops over time (White, 2003). However, it is certain that learners' ability to analyze and process input from their L2 is one of the prerequisites for their acquisition of the L2. Thus, investigating into non-native speakers' unconscious online language processing is undoubtedly one other crucial key to the understanding of how they acquire an L2 and develop the L2 knowledge.

In recent years, researchers from various disciplines have begun to explore real-time non-native language processing (L2 processing), using *online*² measures adopted from the fields of experimental psychology and neuroscience, such as eye-movement monitoring, self-paced reading (SPR), or event-related brain potentials (ERPs) (see e.g., Clahsen & Hong, 1995; Duffield & White, 1999; Fernandez, 2000; Frenck-Mestre & Pynte, 1997; Juffs, 1998; Juffs & Harrington, 1996). Among the various issues under investigation, ambiguity resolution (i.e., resolution of structural ambiguities such as those found in centre-embedded or garden-path sentences, in particular) is one of the critical issues in sentence processing research and has received great attention. For research into ambiguity resolution, identifying types of linguistic information that may be involved in the parser's decision-making processes is crucial. Despite the general consensus that successful sentence comprehension involves the rapid application and coordination of multiple cues to interpretation, scholars' views about the points in time at which different information sources come into play and affect the parser's decisions diverge. The investigation into how the parser processes and resolves globally or temporarily ambiguous constructions can help us understand which kinds of information of the initial analysis of ambiguous structures is predominantly based. The investigation into non-native parsers' ambiguity resolution, furthermore, can help reveal whether and how non-native parsers' use of linguistic information in L2 comprehension differs from native parsers' in L1 comprehension.

So far, what we know about learners' L2 ambiguity resolution during non-native language processing has still been little, relative to native speakers' L1 ambiguity resolution. Many previous studies of L2 ambiguity resolution primarily focused on exploring whether proficient L2 learners differ from native speakers in the extent to which their online parsing decisions are affected by sentence-internal/external cues to interpretation, such as verb argument structure,

¹ The term 'offline' here refers to experimental techniques that seek to investigate comprehenders' ultimate interpretations (or judgements) of a particular sentence or grammatical structure, and which are usually untimed.

² The term 'online' here, is used to refer to a type of experimentation that taps into language processing as it occurs, and which might thus provide some insight into comprehenders' unconscious language processing.

lexical-semantic, contextual information and phrase structure-based information (e.g., Frenck-Mestre, 2005; Papadopoulou, 2005; Pan & Felser, 2011). Findings obtained from offline tasks generally show that native speakers and L2 learners are similar in their capability of using various types of information as cues to ultimate interpretation (e.g., Ying 1996, 2004); however, results from studies looking into L2 learners' online processing of ambiguous sentences in the L2 sometimes diverge. For example, as far as the role of lexical-semantic information in L1 versus L2 processing is concerned, many studies indicate no qualitative differences between L1 and L2 ambiguity resolution in that L2 learners, similar to native speakers, are sensitive to the argument structure properties of the verbs and are able to use this knowledge when processing a sentence in an online task (e.g., Frenck-Mestre & Pynte, 1997; Juffs, 1998, 2004); some other studies, however, indicate learners and native speakers are different in their online application of, and response to, verb subcategorization information (that has been considered structural in nature), albeit similar in that they are both sensitive to such type of information during online sentence processing (e.g., Dussias & Cramer, 2008; Guo et al., 2008). On the other hand, as far as the online application of phrase structure-based information is concerned, several recent studies show that online L2 processing is qualitatively different from L1 processing in that L2 processing may involve a less detailed grammatical representation, relative to L1 processing (e.g. Felser et al., 2003; Felser & Roberts, 2007; Marinis et al., 2005; Papadopoulou & Clahsen, 2003). Results from these L2 processing studies suggest that L2 learners may have difficulty in computing fully detailed syntactic structures in native-like ways when processing L2 input possibly due to inadequacies of their L2 grammar; thus, it is further speculated that compared to native speakers, learners might tend to rely more on lexical-semantic and pragmatic information, knowledge of the world, and strong associative meanings or form patterns to compensate for possible inadequacies of the grammatical representations in the L2 grammar. Such speculation is more clearly discussed in the *Shallow Structure Hypothesis (SSH)* proposed by Clahsen and Felser (2006a, 2006b), which claims that L2 processing is 'shallower³' in nature than L1 processing. Moreover, there are also some studies that present evidence showing differences between learners and native speakers in their online use of discourse-level information. For example, Roberts et al. (2008) found that the number of potential antecedents in the context that preceded the

³ The term 'shallow' here refers to 'involvement of non-structural cues to interpretation' during sentence processing. It has been proposed in the psycholinguistic literature that native speakers' interpretation of a sentence involves two different parsing routes, *full parsing* and *shallow parsing*. *Full parsing* involves the establishment of a fully specified syntactic representation for an incoming string of words, while *shallow parsing* involves a less detailed/underspecified syntactic representation, with sentence comprehension achieved based on lexical-semantic information, associative patterns and other surface-level cues to interpretation (Ferreira, Bailey & Ferraro, 2002; Sanford & Graesser, 2006; Sanford & Sturt, 2002).

target sentence affected highly proficient L2 learners' reading of ambiguous pronouns. Also, Pan and Felser (2011) reported that L2 learners' online ambiguity resolution preferences were influenced significantly by the referential context, with the reading that was referentially supported by the preceding context being favoured over the one that was not; native speakers' ambiguity resolution preferences, in contrast, were not found to be modulated by the referential context in the online task. These findings are in line with *the Shallow Structure Hypothesis (SSH)* in that learners may rely more on meaning-based than on syntactic processing strategies in L2 comprehension.

Among the various factors that have been investigated in L2 ambiguity resolution, phrase length is the one that has attracted some attention recently. Different from other factors that are mostly tested to examine the impact of one single type of information, lengthening a phrase causes changes in at least four different aspects of a phrase/sentence – namely, prosodic weight/ heaviness, distributional pattern, degree of modifiability, processing costs that result from memory loads. L2 learners are generally thought to have lower working memory (WM) capacity in the L2, poorer L2 decoding and lexical access ability, and slower processing speed than native speakers, which together reveal learners' general limitations in basic level cognitive processing. Intuitively, when a phrase in a sentence is lengthened, the distance between any two dependents that are separated by the phrase becomes longer, which may thus increase the processing difficulties due to the increased memory loads caused by the longer integration distance between the two dependents. Since learners usually bear heavier WM demands when parsing their L2 due to the increased computational effort required to recognize and retrieve non-native words or phrases, their general WM capacity limitations may thus cause them to be affected more by the change in phrase length. Note, however, besides causing changes in processing loads in relation to increased memory loads, phrase lengthening, in the meantime, may also change properties of other constraints, such as prosodic weight/ heaviness, distributional pattern, and degree of modifiability (which are discussed in details in the following). Thus, investigating into the factor of phrase length in L2 ambiguity resolution not only helps reveal whether and how L2 learners' general capacity limitations affect L2 comprehension in situations of increased processing demands caused by phrase lengthening but also leads us to see how the potential influences from other constraints caused by lengthening a phrase may interact with learners' general capacity limitations.

2. Potential effects of phrase length in parsing

Whilst the role of phrase length in L1 parsing has been explored in some previous L1 studies, effects of phrase length on learners' online ambiguity resolution in the L2 have not yet been systematically investigated. Our study investigates possible effects of phrase lengthening on L2 learners' structural

ambiguity resolution in sentences such as *The policeman watched the spy with binoculars*. The ambiguity of this type is called prepositional phrase (PP) ambiguities. In the example sentence above, the PP *with binocular* can be interpreted as either a VP-modifying (= VP attachment) PP as indicated in (1a) or an NP-modifying (= NP attachment) PP as in (1b).

- (1) a. The policeman [VP watched [NP the spy] [PP with binoculars]]
 b. The policeman [VP watched [NP the spy [PP with binoculars]]]

Many previous monolingual processing studies have shown that native speakers of English tend to favour the VP modification over the NP modification reading in situations when sentences as those in (1) are presented in isolation (e.g., Clifton et al., 1991; Frazier, 1979; Rayner et al., 1983). This usual preference for interpreting ambiguous PP as VP rather than NP modifiers is consistent with the prediction of the phrase structure-based ‘least’ effort’ principles like *Minimal Attachment*⁴, which predicts VP modification is preferred over NP modification as it requires a less complex syntactic representation. Although purely structural models of parsing (that include the principle of *Minimal Attachment*) suggest that an initial analysis is selected on the basis of syntactic information, with other information (e.g. lexical-semantic or pragmatic information) coming into place at a later stage of parsing, other factors have been found to affect comprehenders’ disambiguation preferences for structurally ambiguous structures, including the definiteness of the postverbal NP (Spivey-Knowlton and Sedivy, 1995), the choice of preposition (Katsika, 2009), the verb type selected (compare, e.g., Britt, 1994; Schütze and Gibson, 1999; Spivey-Knowlton and Sedivy, 1995), and the presence of discourse context that provides biasing referential information (e.g., Altmann and Steedman, 1988; Pan & Felser, 2011; Papadopoulou and Clahsen, 2006; van Berkum et al., 1999). In addition to the above ones, phrase length is also one of the potential factors that has recently attracted researchers’ attention and has been investigated in some monolingual processing studies. For example, Ferreira & Henderson (1991) conducted a study investigating the effects of the length of the ambiguous region on the processing of garden-path sentences containing subject/object ambiguities. Thornton et al. (2000) examined how the length of potential attachment sites for an ambiguous phrase affects L1 ambiguity resolution. Both studies indicate that

⁴ The *Minimal Attachment* principle dictates that the parser should commit to the simplest structural analysis possible; that is, pursue an analysis that requires the fewest syntactic nodes in the grammatical tree diagram and avoid the postulation of any unnecessary nodes (Frazier, 1979, 1987). Concerning PP attachment ambiguities of the kind under investigation, it predicts that the reading of the PPs interpreted as VP modifiers should be preferable to that of the PPs interpreted as NP modifiers because VP attachment requires the computation of fewer nodes when compared to NP attachment, as illustrated in (a) and (b) below.

(a) [VP V NP PP] (b) [VP V [NP NP PP]]

phrase length can be linked to processing difficulty to a certain degree.

More specifically, Ferreira & Henderson (1991) used a grammaticality judgment task to conduct a series of experiments to test how the *length* factor affects readers' reanalysis of garden-path sentences containing subject/object ambiguities, such as *Because Bill drinks wine is never kept in the house*. Under the presupposition of incremental parsing, this sentence is syntactically ambiguous in that the word *wine* could initially be analyzed as the direct object of the verb *drinks* in the first/embedded clause, but then has to be reanalysed as the subject of the second/matrix clause as soon as the disambiguating word *is*, which serves as the sentential matrix verb, is reached. The first two experiments of Ferreira & Henderson's (1991) manipulated the length of the ambiguous region, with the number of words in the ambiguous phrase being varied, such as in the following examples (2a) and (2b).

(2) a. Early closure (the ambiguous phrase as the subject of the matrix clause):

After the Martians invaded THE TOWN / THE TOWN THAT THE CITY BORDERED was evacuated.

b. Late closure (the ambiguous phrase as the direct object of the embedded clause):

After the Martians invaded THE TOWN / THE TOWN THAT THE CITY BORDERED the people were evacuated.

It was found from the two experiments that the length of the ambiguous phrase affected the ease with which participants successfully parsed the garden-path sentences. Participants' mean grammaticality judgments accuracy was significantly lower for the long- than for the short- ambiguous phrase condition in both the early and late closure sentences, but the effect of length appeared stronger for the early closure sentences than for the late closure ones. The follow-up experiments further confirm the increase of difficulty in the reanalysis caused by increasing the length of the ambiguous phrase was attributable to an increase in the lengthened phrase's syntactic complexity, but was not due to the increase in the number of words. Taken together, Ferreira & Henderson (1991) found that length has a significant impact on the ease of syntactic reanalysis and that this length effect is mainly due to the increased processing difficulty caused by the increased distance between the disambiguating word in the sentence and the initially misanalyzed phrase.

Besides affecting the reanalysis of garden-path sentences, phrase length has also been found to play a crucial role in the resolution of modification ambiguities in Thornton & MacDonald's (1999) study. They carried out two self-paced reading (SPR) experiments to examine how length affects the resolution of verb phrase modification ambiguities in sentences like *She taught (VP1) the kids to dive (VP2) in a single afternoon / into the deep end*. This sentence contains a temporarily ambiguous prepositional phrase (PP) that can potentially be construed as either the

modifier of the distant VP, namely VP1 (= *in a single afternoon*) or the modifier of the local VP, namely VP2 (= *into the deep end*). The experimental sentences were manipulated with the post-verbal noun phrase being lengthened through the addition of prenominal adjectives, such as *the kids* versus *the six and seventh grade kids*. The results from Thornton & MacDonald's (1999) study show that whilst no particular preference was observed towards either distant (=VP1 attachment) or local modification (=VP2 attachment) in the situations of short post-verbal NPs, local modification was found to be favored over distance modification when the NPs were lengthened. This finding indicates that the degree to which local modification is preferred is strengthened when the linear distance between the potential attachment sites increases, which thus provides evidence in support of the idea that any recency effects in modifier ambiguity resolution should be modulated by distance change resulting from lengthening phrases. Following up to Thornton & MacDonald's (1999) findings, which are indicative of phrase length effects in verb phrase modification/attachment ambiguity resolution, Thornton, MacDonald & Arnold (2000) subsequently tested how the phrase length factor might affect native English speakers' processing of sentences containing PP attachment ambiguities in VP – NP – PP configurations, where the PP could potentially modify either the VP (the distant site) or the NP (the local site), such as in *the police watched the spy with binoculars*. Instead of manipulating the length factor by adding material between the potential attachment sites as Thornton & MacDonald (1999) did, Thornton et al. (2000) did this by adding material to one of the potential attachment sites. Specifically, they manipulated length by adding prenominal modifiers to the post-verbal NP that was one of the potential attachment sites for the PP modifier, as indicated in brackets in the following – 'short' versus 'long' NP: *The salesman glanced at the (amazingly young rude) customer with suspicion/ripped jeans and then walked away*. In this case, the distance between the PP modifier and the verb is greater in the long NP than in the short NP condition, but the distance between the PP and the post-verbal NP remains the same across the two length conditions. Having a group of native English speakers tested with sentences of this type, Thornton et al. (2000) found that the native English speakers' VP attachment preference observed in the short NP condition was eliminated in the long NP condition. Their finding that processing difficulty for VP modification (reflected in increased reading times) was greater when the potential VP attachment site was further separated from the modifier compared to when it was closer to it, is generally consistent with *recency*-based proposals. It is suggested in their proposals that the relative strength of *recency* should be modulated by distance, with the power of locality strengthened when the distance between two dependent elements is increased (i.e. in this case, between the verb and the PP). Thornton et al.'s results demonstrated that phrase length plays a significant role in L1 PP attachment ambiguity resolution.

Together, the results from Thornton & MacDonald's (1999) and Thornton,

MacDonald & Arnold's (2000) studies indicate that any recency-based effect is modulated by the distance between two dependents, with the strength of locality reinforced when the two dependents are separated further, compared to when they are close together. Even the findings by Ferreira & Henderson's (1991) study of the reanalysis of garden-path sentences, which showed that the parser found it harder to recover from an initially misanalyzed analysis when the distance between the head of the misanalyzed phrase and the disambiguating word became greater, appear to be in line with the assumption that processing tends to be more difficult when grammatical dependencies are separated further.

Summarizing from the above empirical findings, we can see that there seems to be an increase in processing difficulty in the situations of increased distance between two dependent elements during sentence processing. Some researchers such as Fodor (1998), Gibson (1998) and MacDonald (1999) have all discussed the potential effects of phrase length in their studies and indicated an increase in processing difficulty when the distance of dependencies is increased, although their proposals differ in details. Specifically, there are at least three possible factors that can explain this length effect that have been proposed⁵:

1. Integration resources: Gibson (1998) proposed the *Syntactic Prediction Locality Theory* (SPLT), which specifies the relationship between sentence processing mechanisms and access to available computational resources. The SPLT rests on two major components, *integration cost* and *memory cost*, both of which are hypothesised to be constrained by locality. That processing difficulty increases when the distance between dependents is made longer (or more specifically for the case regarding attachment preferences that we discuss above, as in Thornton et al., (2000) – that the relative strength of locality is strengthened when two dependent elements are separated further due to phrase lengthening) mainly fits the prediction of the *integration cost* component hypothesised in the SPLT, according to which the longer distance that the integration of two dependents needs to cross, the heavier processing difficulties it will manifest⁶.

⁵ Note that these different factors may not necessarily be independent of each other. More than one may possibly exert influence simultaneously.

⁶ Take Gibson's (1998) study for example. When processing sentences such as (i) and (ii) below, from an integration cost perspective, when the parser encounters the preposition *about* in (i) and attempts to associate it with the verb *talked*, the integration process between the two requires the parser to cross a longer phrase, while in (ii) the integration of the preposition *to* and the verb *talked* simply needs to cross a shorter one. Thus the processing of sentences such as (i), which involve longer-distance integration, are expected to require more integration costs, thus being more difficult to process than sentences such as (ii) that require shorter-distance integration.

(i) The landlord talked [PP to the visitor who had just arrived and was staying for the summer] [PP about the problem].
(ii) The landlord talked [PP about the problem] [PP to the visitor who had just arrived and was staying for the summer].

2. Prosodic weight: Fodor (1998, 2002) proposed the *Implicit Prosody Hypothesis (IPH)*, which suggests that the default prosodic contour will still be projected even in a silent reading, during which the prosodic analysis is implicit rather than explicit. Fodor demonstrated that there is a natural prosodic tendency, dubbed ‘the prosodic same-size-sister constraint’, according to which a constituent should be more preferable to attach to a prosodic unit that is equal to it in prosodic weight/heaviness. In other words, a constituent likes to have a sister of its own size for prosodic weight balance⁷. Applying this prosody-based, ‘same-size-sister’ hypothesis to ambiguous PP attachment constructions as investigated in Thornton et al.’s (2000) study, the observed stronger tendency for the ambiguous PP being attached low to the local NP in the long NP condition could be explained by the lengthened NP forming an independent prosodic unit that is equal in size vis-à-vis the attaching PP modifier.
3. Distribution-based patterns: MacDonald (1999), on the other hand, raised another proposal to account for a locality effect associated phrase length. MacDonald suggested that phrase ordering preferences in speech production can create specific distributional patterns in language, to which comprehenders are also sensitive in resolving syntactic ambiguities. Specifically, the general speech production constraints (e.g. Hawkins, 1994; Wasow, 1997) require short phrases/constituents to be uttered prior to longer ones. Based on this general short-before-long phrase ordering preference in speech production, MacDonald argued that a long-before-short phrase ordering that violates the short-before-long principle tends to be used to convey a certain intended meaning. Take the following sentences containing verb modification ambiguities for example.

- (i) Bill said that John had left yesterday. (Long-short order)
- (ii) Bill said yesterday that John had left. (Short-long order)
- (iii) Yesterday, Bill said that John had left. (Short-long order)

In (i), the adverb *yesterday* can either non-locally modify the distant verb *said* or locally modify the verb *left* in the embedded sentential complement *that John had left*. In (i), the final adverbial *yesterday* is shorter than the

⁷ This proposed prosodic tendency is based on Fodor’s observation that for an RC attachment construction with complex NP antecedents (*NP1 of NP2 + RC*), short RCs are more inclined to attach low and long RCs to attach high. Fodor suggests that if an attaching constituent is prosodically light in weight, low attachment (guided by a recency-based parsing strategy like the universal principle of *Late Closure*) is preferred. The attachment of prosodically heavier constituents, however, is argued to be determined by both the relative prosodic weight of the attaching constituent vis-à-vis its possible hosts and the prosodic patterns of a language.

preceding phrase *that John had left*, which violates the short-before-long constraint. To convey the intended meaning in which *yesterday* modifies *left* locally, expressing at what time John had left, however, the long-before-short phrase order in (i) is the only option, since the short-before-long order as in (ii) and (iii), *yesterday* undoubtedly has to modify the main verb *said*. Based on these characteristics of distributional information regarding phrase ordering, MacDonald argued that the long-before-short phrase order may tend to be associated with local sites, since it is the only phrase ordering pattern that can convey the meaning of local-site modification. This distributional constraint proposed by MacDonald (1999) provides another explanation for Thornton et al.'s (2000) observation of local modification preferences in the long NP condition, where the post-verbal NP appearing before the PP modifier was greatly lengthened, resulting in a long-before-short pattern.

This paper aims to explore possible effects of phrase length in PP ambiguity resolution by replicating Thornton et al.'s (2000) study of phrase length effects in L1 PP ambiguity resolution and to extend the investigation to non-native speakers' PP ambiguity resolution preferences. The phrase length manipulation in this study is carried out by adding prenominal adjectives to the post-verbal noun phrase in sentences of the V – NP – PP structure. In terms of the structurally ambiguous PP attachment constructions (V – NP – PP) under investigation, the three possible factors that are discussed earlier (namely, integration resources, prosodic weights and distributional patterns) may all result in a preference for local modification (= NP attachment), although the theoretical bases behind each are quite different. However, it should be noted that phrase lengthening, via adding extra prenominal modifiers to the NP also involves changes in the NP's information content and might thus affect modification preferences. For example, Thornton, MacDonald & Gil (1999) manipulated the degree of specificity of the local NP by using either a less specific noun phrase with an indefinite determiner or a more specific noun phrase qualified by a possessive pronoun, as in *The plumber by an enamel sink / our kitchen sink with....* According to the authors, the noun phrase *an enamel sink* was unspecific because there are many enamel sinks in the world. Compared to *an enamel sink*, the noun phrase *our kitchen sink* seems more specific, because one house is usually equipped with a single kitchen sink. Their results show that RTs were significantly longer for the ambiguous modifier starting from *with* to attach to the local NP when the NP was a more specific noun (i.e., *our kitchen sink*), compared to when it was an indefinite one (i.e., *an enamel sink*). This finding was therefore argued to reflect the infelicity of a heavily modified/specified noun being further modified by extra information. That is, adding information to a noun may strengthen its specificity, hence decreasing modifiability.

Returning to this current study investigating phrase length effects in PP ambiguity resolution, let us now take the findings of Thornton et al.'s (1999) study

into further consideration: when the length factor is manipulated via the addition of information, in addition to the possible length effects that may be caused by the three factors discussed earlier (which would direct the ambiguity resolution to local attachment), possible concomitant pragmatic effects may also arise. Concerning the PP attachment construction, whilst the three distance-concerned (or length-based) proposals predict that the likelihood of the PP modifier being associated with the verb (= VP modification) should decrease when the distance between the two dependents increases, the pragmatics-based accounts of modifiability predicts the opposite. That is, an already modified post-verbal NP in the condition when the NP is lengthened via information addition (termed ‘the long NP’ condition here) should be less felicitous to take the PP as its additional modifier, compared to a simple NP in the ‘short NP’ condition (where no prenominal modifiers are added to the post-verbal noun). The infelicity of further modification on the NP site in the long condition may thus strengthen the degree to which VP modification disambiguation is preferred.

This current study does not side-step the problem of possible concomitant pragmatic effects that may simultaneously arise from length manipulations via the addition of information – the same as Thornton et al. (2000) did. By doing this on purpose, we can additionally explore the possible pragmatic consequences (e.g. of modifiability) that the phrase length manipulations might result in and how this pragmatic factor interacts with the other distance-based ones (i.e. integration resources, prosodic weights and distributional patterns). Another important reason for not avoiding the possible concomitant pragmatic effects is that this study mainly aims to look into L2 learners’ processing of structurally ambiguous sentences in comparison to native speakers’. There has been some evidence showing that L2 processing is different from L1 processing in that L2 learners rely more on non-structural cues to interpretation, compared to native speakers (e.g., Pan & Felser, 2011; Roberts et al., 2008 – see also the *Shallow structure Hypothesis* in Clahsen & Felser, 2006a, 2006b). Therefore, this study provides good opportunities not only to probe how structural and pragmatic constraints triggered by the same length manipulation might simultaneously affect processing but also to examine if non-native speakers who have been found to rely more on meaning-based processing strategies in L2 comprehension do behave differently from native speakers in the extent to which their online processing is affected by structure/distance-based information vis-à-vis pragmatics-based modifiability cues that are non-structural in nature.

3. Experiment

To explore possible effects of phrase length on native and non-native speakers’ initial disambiguation preferences for sentences containing PP attachment ambiguities, we carried out an online self-paced reading experiment and adopted materials similar to those used by Thornton et al. (2000) with the factor of *phrase length* being manipulated via addition of prenominal modifiers to the post-verbal

noun. In the following, we dub the condition when the post-verbal noun is lengthened 'the long NP condition', and the condition when the noun is not lengthened 'the short-NP condition'.

For sentence containing PP attachment ambiguities, general parsing principles such as *Minimal Attachment* predict a general VP over NP modification preference. However, participants' disambiguation choices/preferences may also be affected by phrase length, as discussed in some length-based proposals stated above (e.g., Fodor, 1998; Gibson, 1998; MacDonald, 1999). These proposals, though different in their detailed accounts of the role of phrase length, all predict the strength of recency to be modulated by linear structural distance, such that establishing a syntactic or semantic link between two elements in the discourse becomes more difficult the further the two are separated from each other. Here, we dub the proposal of this kind a (structural) recency-based hypothesis. With regard to the type of PP attachment ambiguities under investigation and with the method of manipulating the *phrase length* factor modelled on Thornton et al.'s (2000) study, on the view of the recency-based hypothesis, we would expect VP modification to be less felicitous in our long-NP condition, where the addition of prenominal modifiers attached to the post-verbal NP further separates the PP modifier and the potential VP modification site, compared to the short-NP condition. Besides considering the change in linear distance caused by lengthening the NP, we also need to take into account possible changes to pragmatic properties that could result from the same length manipulation. By adding additional adjectives/adverbials to the post-verbal NP, we also add extra information to the NP, which may thus result in a change to its modifiability property. The modifiability proposal, in contrast, suggests that an NP that is already heavily modified by prenominal modifiers, as in our long-NP condition, should instead be less felicitous with another (PP) modifier added to it than in the case of an unmodified NP in the short-NP condition. Thus, on the view of the modifiability hypothesis, we would predict NP modification to be less preferred in the long than in the short NP condition, which is opposite to what the recency-based hypothesis leads us to expect.

Furthermore, if L2 learners generally rely more on meaning-based processing strategies than native speakers do, and if these include the rapid integration of pragmatics-based cues of modifiability provided by phrase lengthening, we expect to see them disfavour NP modification in the long NP condition where the modifiability of the already modified post-verbal NP is reduced, compared to the short NP condition. However, if L2 learners are more sensitive to the recency-based constraint (triggered by the factor of integration resources, prosodic weights, or distributional patterns⁸), we should see learners exhibit a processing disadvantage

⁸ Note that, expect for the prosody-based constraint, *prosodic weight*, which is considered non-structural, the constraints of *integration resources* and *distributional patterns* are both factors that are structurally-concerned.

for VP modification in the long-NP condition, where the linear structural distance between the V and the PP modifier is increased, compared to the case in the short-NP condition. Particularly, L2 learners usually bear heavier WM demands when parsing their L2 due to the increased computational effort required to recognize and retrieve non-native words or phrases. Learners' general WM capacity limitations may lead us to expect that they should suffer more from the increased processing load caused by NP lengthening, hence showing more difficulty with VP modification in our long-NP condition than the native speakers do, because they may find it more difficult to keep the verb active in working memory until the PP modifier has been processed. This trend predicted based on learners' presumed WM shortage is in agreement with the expected consequences caused by L2 learners' limited integration resources/ or heavier integration costs and is also in line with the general prediction of the recency-based constraint, such that processing difficulty (for VP modification) is increased with the increased distance between the two dependent elements (i.e. between the verb and the PP).

3.1 Method

3.1.1 Participants

Participants consist of 36 Chinese-speaking learners of English (17 male and 19 female, mean age = 22.11, range: 17 – 37; SD = 12.48) and 36 native English speaker controls (11 male and 25 female, mean age = 28.64, range: 18 – 60; SD = 5.19) who all participated in the experiment voluntarily. All of the participants, including the native English and the Chinese L2 participants, had normal or corrected-to-normal vision and had never been diagnosed as dyslexic. Of the 36 Chinese-speaking learners, 17 resided in the U.K and had been immersed in English for 15.47 months on average (range = 1 – 103 months, SD = 23.28), whilst the other 19 Chinese participants resided in Taiwan at the time they participated in the experiment and had never been immersed in an English-speaking country. The learners' average age of first exposure to English was aged 11.08 years (range = 5 – 13; SD = 1.75). All of the participants were naïve in regard to the ultimate purpose of this experiment.

Besides the main experiment, the Chinese-speaking participants also completed a vocabulary test and the grammar part of the Oxford Placement test (OPT; Allan, 2004)⁹. The vocabulary test aimed to help ensure that learners were familiar with the verbs, nouns and adjectives that were used in the critical target sentences. The OPT test, on the other hand, served two purposes: one was to serve as a criterion for the L2 English proficiency measurement, providing us with the L2 learners' general English proficiency at the time of testing; the other was to allow

⁹ The OPT test we used was a paper-and-pencil task in which participants were asked to complete a gap or gaps in a sentence or paragraph by choosing the grammatically correct choice from three given answer options.

us to explore the potential effects of L2 proficiency on ambiguity resolution in non-native language processing in the further analyses. The group mean for the OPT scores of the Chinese-speaking participants was 69.78 (out of 100), which corresponds to the upper intermediate level according to the OPT scale (range = 51 – 89, SD = 9.55). Moreover, to allow us to examine potential working memory effects on participants' processing performance, all participants were additionally required to complete a reading span test, which provided a measure of their working memory capacity (the test for L1 English group was based on Daneman & Carpenter, (1980)¹⁰; the test for Chinese L2 English group was based on Harrington & Sawyer, (1992)¹¹). The purpose of the reading span tests was to examine the relationship between the individual differences of working memory capacity and participants' performances, which would then allow us to assess the possible role of working memory capacity in online ambiguity resolution. When performing these reading span tests, participants were required to read increasingly larger sets of sentences and then recall word of each sentence at the end of each set. The group mean reading span score was 3 (out of 6) for the native English controls (range: 2 – 4.5, SD = 0.68) and 25 (out of 42) for the Chinese participants (range: 15 – 39, SD = 5.88)¹².

3.1.2 Materials

The materials were adapted from Spivey-Knowlton & Sedivy's (1995) and comprised of 16 experimental items and 32 fillers. The experimental sentences were sentences containing PP attachment ambiguity of the general form V – NP – PP. The matrix verbs used were either psych or perception verbs, and the post-verbal noun phrase was always definite. The average string length of the entire PP modifier (with + NP) was matched across the two modification conditions as

¹⁰ Daneman and Carpenter's (1980) carried out a study exploring the relationship between native English speakers' performances in several reading comprehension measures and in a reading span test with heavy processing and storage demands that involved in memorization and recall of each final word of a series of sentences. They found that the participants' reading spans as measured by the devised reading span test highly correlated with their reading comprehension performances.

¹¹ Harrington & Sawyer (1992) investigated the relationship between second language working memory capacity (measured by the reading span test devised by Harrington & Sawyer, 1992) and the reading comprehension performances of a group of advanced L2 learners. The reading span test by Harrington & Sawyer was derived from the original test by Daneman & Carpenter (1980), but its design was changed to make the test applicable to non-native speakers, with the syntactic structure of the test sentences simplified and the lengths shortened. Their results also showed a strong statistical correlation between participants' reading span and reading comprehension skill with those with larger working memory capacities scored higher on measures of reading skill.

¹² Note that Harrington and Sawyer's (1992) reading span test was a variant of Daneman and Carpenter's (1980) and its scoring procedure was slightly different from Daneman and Carpenter's. According to Whitney et al. (2001), however, the scores obtained by either measure were found to be highly correlated in the statistics.

much as possible, with VP-modifying PPs consisting of 17.19 and the NP-modifying ones of 17.06 characters on average. Besides, the word frequency of the disambiguation nouns in the PPs (e.g., '*suspicion*' from *with strong suspicion*; '*jeans*' from *with ripped jeans*) were also matched as closely as possible (VP modification: 35.56; NP modification: 36, according to the CELEC database).

A 2 x 2 design was applied with Phrase Length (*short-NP or long-NP*) and Attachment (*VP modification or NP modification*) as within-subjects factors. The experimental sentences contained either a short or a long post-verbal noun phrase (e.g., *the customer* versus *the amazingly rude young customer*) that was followed by a PP modifier that was semantically disambiguated towards either VP or NP modification interpretation, such as in *John glanced at the (amazingly rude young) customer with strong suspicion / with ripped jeans*. The manipulation of the NP length was through the insertion of three prenominal modifiers (three adjectives or two adjectives plus one intensifying adverbial) to the definite post-verbal noun.

Some pilot testing revealed that Chinese-speaking participants at the proficiency range to be tested showed difficulty in comprehending the experimental stimuli when the presentation of the sentences was one word at a time, suggesting that a word-by-word presentation type might have exceeded these participants' processing capacity and resulted in an excessive number of comprehension errors. Therefore, segment-by-segment presentation was chosen to avoid unacceptably large proportion of unusable data. Each experimental item was divided into five segments as indicated by the slashes: [John/glanced at/the (amazingly rude young) customer/with strong suspicion (with ripped jeans)/and then walked away.]

The 16 experimental items were distributed across four counterbalanced presentation lists using a Latin Square design and then randomized with 32 filler items. To help ensure that participants read the texts carefully for meaning during the testing, all the experimental items and half of the fillers were followed by a yes/no comprehension question. The end-of-trial comprehension questions asked the participants about the content of the texts they just read but did not explicitly probe how the PP was disambiguated. The accuracy rates of the individual participants' answers to the comprehension questions were calculated afterwards.

3.1.3 Procedures

All participants were tested individually in a quiet setting. The English participants were tested in a single session which included the main experiment and the reading span test. The Chinese participants were tested in two separate sessions around one week apart. Session one included the main experiment and the reading span test, which together took the participants around 40 minutes to complete. Session two involved the testing of the paper-and-pencil proficiency and vocabulary tests, which took about 40 – 45 minutes overall.

The main experiment was a segment-by-segment self-paced reading (SPR) experiment, using the non-cumulative moving-window technique (Just, Carpenter

& Woolley, 1982). The SPR task was programmed and administered using the DMDX experimental software package (Forster & Forster, 2003), which controlled the presentation of the stimuli and the recording of participants' reading times and end-of-trial responses. In the SPR paradigm, participants controlled the presentation of the stimuli by pressing the 'Continue' button on the computer keyboard to receive the next segment that would then replace the previous one on the screen. They pressed the designated button to read through each phrasal segment of a sentence one after another as quickly as they could, while reading for meaning. Participants' response time to each segment will be calculated individually, with the assumption that processing difficulties will be reflected in an increase in reading times. The uninterrupted ongoing reading process of the readers with an added time pressure in the SPR paradigm enables us to see the participants' unreflected/unconscious responses to either of the modification alternatives (PP modifier disambiguated towards either VP or NP modification) for the different length conditions. At the end of all experimental trials and half of the filler items, there was a yes/no comprehension question. The participants were instructed to press a designated 'Yes' or 'No' button to answer the question accurately based on the content of the texts they had just read. The main testing started with oral instructions and a practice session that included three practice items to familiarize the participants with the task. Once the participants felt confident about the procedure, they proceeded to the main part of the test. In the middle of the test, they were given a short break for a few minutes in order to ease any potential tiredness resulting from gazing at the screen for a long time.

3.2 Results

The overall accuracy of the end-of-experimental trial comprehension questions was 93.15% (range: 87.5% - 100%) for the native English speakers, and 86.11% (range: 87.5% - 100%) for the Chinese-speaking L2 English learners. Despite that the Chinese-speaking participants descriptively manifested lower accuracy than the native English controls, the overall accuracy rates were high across the two language groups, suggesting that all participants paid attention to the task and read the experimental items carefully for meaning. Statistical analyses of the reading times were carried out for correctly answered trials only. We also removed individual trials from the L2 dataset that contained any unknown vocabulary items according to the results of the vocabulary test, which affected 14% of the L2 correctly-answered trials. Furthermore, to avoid the results being distorted by extreme data points, reading times of 2.5 SDs more or less than the mean reading times (RTs) per group and condition in each segment were defined as outliers and excluded from the analyses. The removal of the outliers affected 4.4% of the remaining L1 data and 6% of the remaining L2 data for the critical region; 4.8% of the L1 and 6.3% of the L2 for the post-critical region.

We statistically analyzed participants' reading times for both the critical PP

region (the disambiguating region at which the experimental conditions started to diverge) and the post-critical region (the segment that immediately followed the critical one). To know whether the two participant groups' reading time patterns differed statistically, preliminary repeated-measures between-Groups ANOVAs were firstly performed on the data for each of the two regions of interest, with Length (*Short-NP, Long-NP*) and Attachment (*VP modification, NP modification*) as two-levelled within-subjects factors and with Language Group (*L1 English group, Chinese L2 group*) as a two-levelled between-subjects factor. For the critical region, we found a significant Length by Attachment interaction in the subjects analysis [$F(1, 67) = 6.863, p = .011$; $F(1, 28) = 1.259, p = .271$] and two interactions with Group: a two-way Length by Group interaction [$F(1, 67) = 9.229, p = .003$; $F(1, 28) = 9.116, p = .005$] that was significant by both subjects and items, and a three-way Length by Attachment by Group interaction [$F(1, 67) = 8.407, p = .005$; $F(1, 28) = 3.909, p = .058$] significant by subjects and marginally significant by items. There was also a main effect of Group that reached significance for both the subjects and items analyses [$F(1, 67) = 231.584, p = .000$; $F(1, 28) = 359.218, p = .000$], indicating that the Chinese L2 learners generally read the texts more slowly than the native English controls.

Given that the results of the preliminary analyses indicated different processing patterns in the two language groups (that were revealed by the two interactions with Group observed in the preliminary analyses), we subsequently separated the reading time data of the two language groups and examined each group individually, using repeated measures ANOVAs (within-group) with Length and Attachment as within-subjects factors. For the native English controls, these showed only a significant main effect of Length [$F(1, 35) = 22.032, p = .000$; $F(1, 14) = 8.484, p = .011$] at the critical region and a marginal main effect of Length by items [$F(1, 34) = 2.055, p = .161$; $F(1, 14) = 3.298, p = .091$] at the post-critical region that both indicated that processing speed was generally slower in the long NP condition than in the short NP condition, irrespective of modification types. The analysis of the Chinese group's reading times at the critical region, on the other hand, showed a marginal main effect of Length [$F(1, 32) = 2.939, p = .096$; $F(1, 14) = 2.991, p = .106$], which went in the opposite direction to what was seen in the native English controls, with the PPs being processed more slowly in the short than in the long NP condition. Importantly, there was a Length by Attachment interaction [$F(1, 32) = 7.797, p = .009$; $F(1, 14) = 4.168, p = .061$] that was significant in the subjects analysis and marginally significant in the items analysis. This interaction was indicative of the pattern that the PPs disambiguated towards NP modification were read faster than those disambiguated towards VP modification in the short-NP condition, while the PPs disambiguated towards VP modification were read faster than those disambiguated towards NP modification in the long-NP condition. Subsequent paired *t*-tests (two-tailed) confirmed that this

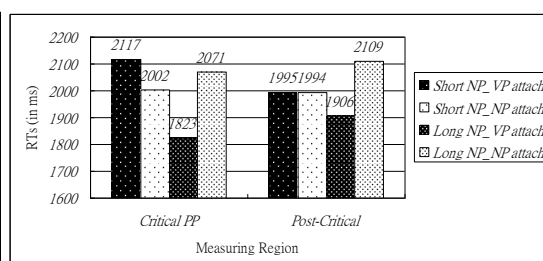
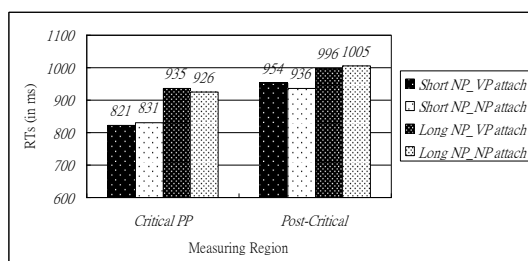
observed interaction between Length and Attachment was mainly carried by the PPs disambiguated towards VP modification being processed faster than those disambiguated towards NP modification in the long-NP condition ($t_1 = -3.634$, $p = .001$; $t_2 = -2.130$, $p = .051$). The within-group ANOVA analysis was also carried out for the post-critical segment but failed to show any statistically significant effects here. Figures 1 and 2 clearly present the RT patterns of each group per segment/ condition.

3.3 Further analyses

Given that the differences in sensitivity/and response to phrase length observed

Figure 1: The native English group' RTs

Figure 2: The Chinese L2 group' RTs



between the native speakers and the learners in the SPR task might have been influenced by

individual differences in working memory capacity or learners' general L2 English proficiency, we further examined the relationship between the reading time patterns and individual differences in WM capacity (as measured by the reading span tests) and L2 proficiency (as measured by the OPT) by running some additional statistical analyses for the critical segment at which we found clear L1/L2 differences. To examine whether the observed L1/L2 differences might be in connection to differences in participants' WM capacity, repeated-measures ANOVAs with Reading Span as a covariate were carried out for both the native and the non-native speakers' reading times of the critical PP segment. The results of the WM analysis failed to show any statistically reliable interactions with Reading Span in either the native English group or the Chinese L2 group, but only showed a main effect of Reading Span in the Chinese group [$F(1, 31)$, $p = .030$], indicating the trend that low-span Chinese learners' reading speed was faster than high-span ones' in general.

In addition to considering the possible effects of learners' general working memory limitations, we also take into account the possibility that the observed L1/L2 differences could have also been affected by the learners' L2 proficiency.

Learners at the higher end of the range may potentially pattern more like the native speakers, when compared to learners at the lower end of the range. We performed an additional analysis for the critical segment by running repeated-measures ANOVAs with the Chinese L2 participants' OPT scores as a covariate. The results of this analysis did not show any significant main effect of, or interactions, with L2 proficiency as reflected by the learners' OPT scores.

In short, the additional subject-specific reading span and L2 proficiency factors did not have any measurable influence either on native and non-native participants' online ambiguity resolution preferences, or on their sensitivity to our phrase length manipulation. In particular, the absence of a three-way Length by Attachment by Reading Span (or OPT scores) interaction failed to support the hypothesis that the observed L1/L2 differences in the extent to which ambiguity resolution preferences were affected by phrase length manipulation might be accounted for by either general working memory shortage or low proficiency in the L2.

4. Discussion

In summary, the main results from the SPR experiment are as what follows:

- Compared to the native English group, the Chinese L2 group showed higher error rates in answering the comprehension questions and longer reading times overall.
- The native speakers' online ambiguity resolution preferences were not modulated by phrase length in either the critical or post-critical regions, although the *phrase length* factor was found to influence their processing in that processing speed tended to be slower after the long NPs than the short NPs in general.
- The Chinese L2 learners' ambiguity resolution preferences, in contrast, were found to be modulated by phrase length, in that their processing of PPs disambiguated towards VP modification was significantly faster than it was for disambiguated towards NP modification in the long-NP condition. In addition, the Chinese learners showed a general length effect, with the experimental items being processed more quickly in the long-NP than in the short-NP condition overall.
- Individual differences in WM capacity or L2 proficiency did not have a significant impact on participants' parsing performance, including their online ambiguity resolution preferences and their sensitivity to phrase length manipulation.

In the following, these findings are discussed based on the predictions outlined earlier.

4.1 Effects of phrase length in L1 and L2 processing

The first glance of the results from the SPR experiment showed that the Chinese L2 learners generally had slower processing speed and exhibited higher error rates in answering the end-of-trial comprehension questions than the native English controls. These findings are not surprising because learners who additionally need to decode a non-native script and retrieve non-native lexical items during L2 processing are usually thought to be slower and less accurate readers when compared to native speakers. However, as regards to the investigation into phrase length effects on L1 versus L2 PP ambiguity resolution, the results showed striking contrasts between native and non-native speakers in the extent to which their online processing of ambiguous sentences are affected by phrase length.

For the native English speakers, although we found that phrase length manipulation had an impact on general processing speed in that their reading times were generally higher for the long than for the short NP condition, irrespective of modification type (as indicated by the main effect of Length on the critical and post-critical regions), it did not seem to modulate the native speakers' online ambiguity resolution statistically. The general reading-time disadvantage for the long NP condition could be indicative of processing being slowed down after encountering more complex phrases, namely, a lengthened NP in the long-NP condition. More specifically, the longer reading times required to process a more complex NP (i.e. the lengthened post-verbal NP) in the long NP condition might have been prolonged and carried over to later segments, resulting in the increase of the RTs in the long-NP condition. Based on the hypotheses and predictions outlined earlier, we expected that the length manipulation might modulate participants' ambiguity resolution preferences, following either the pattern suggested by the recency constraint or the pattern suggested by the modifiability constraint. However, besides the main effect of Phrase Length discussed above, we did not find the online ambiguity resolution preferences of the native speakers to be significantly affected or modulated by length manipulation, neither in the critical PP nor in the post-critical regions¹³.

¹³ Note, however, that the absence of the length effects modulating native English speakers' ambiguity resolution preferences appeared inconsistent with the findings by Thornton et al. (2000) who found native English speakers' degree to which VP modification disambiguation was preferred over NP modification disambiguation was significantly lessened or even eliminated in the long-NP condition. While Thornton et al. used word-by-word presentation, we changed this to segment-by-segment presentation, in consideration of non-native speakers' presumed computational resource limitations. The absence of effects in native speakers might thus be due to the fact that the presentation type that we deemed suitable for non-native speakers might have made processing too easy for the native speakers to show any effects. Also, we should perhaps not dwell on this discrepancy too much and probe any further into this, because that the main interest of our study is to examine the degree of sensitivity to length manipulation in non-native language processing in comparison to native language processing, rather than re-examining the role of phrase length in

The Chinese L2 learners, on the other hand, showed a clear trend of their modification preferences being modulated by phrase length, such that the reading-time advantage for NP modification in the short-NP condition was shifted to an advantage for VP modification in the long-NP condition. Specifically, the Chinese learners' slight descriptive preference for NP-disambiguated items in the short-NP condition was turned around in the long-NP condition, resulting in a statistical VP over NP modification preference. The pattern seen in the learners obviously goes against the expectation of any (linear distance-based) recency constraint, which would predict that the degree to which VP modification is preferred should decrease with an increase in the linear structural distance between the verb and the PP modifier (i.e. in the long-NP condition where the potential VP modification site was further separated from the PP modifier). However, what we instead observed was a VP over NP modification preference in the long-NP condition, which seems compatible with the prediction of the pragmatics-based modifiability constraint, according to which a noun that has already been heavily modified (i.e. *by the prenominal modifiers in the long-NP condition*) should less readily take an additional PP modifier. That NP modification disambiguation was not favoured in the long-NP condition in the online SPR task could thus be suggestive of the initial sensitivity to the modifiability constraint outlined in our earlier predictions. Taken together, our finding that learners are guided more than native speakers by pragmatic cues supports the hypothesis that nonnative speakers rely more on non-structural cues to interpretation than on phrase-structure based parsing principles during online comprehension, compared to native speakers.

In addition, we also notice that the Chinese learners showed a general effect of phrase length showing shorter RTs for the long than for the short NP condition in general, which went in the opposite direction when compared to the phrase length effect seen in the native speakers'. That the learners generally processed the stimulus items more quickly after reading long NPs than after short NPs seems somewhat counter-intuitive. We speculate that learners' shorter reading times for the long-NP condition overall might reflect their expectation of the appearance of more phrases/expressions after a long post-verbal NP in the long-NP condition, perhaps due to an attempt to balance the excessive 'weight' or 'heaviness' (e.g. of prosody, pragmatics, or simply non-linguistic visual perception) of the long NP.

4.2 The absence of effects of individual differences in WM capacity and L2 proficiency

Besides taking into account the possibility that the observed L1/L2 differences in sensitivity to phrase length were due to qualitative differences between L1 and L2 processing, we should also consider the possible role of participants' individual differences in WM capacity and L2 proficiency during processing and to what

extent these factors might give rise to the observed L1/L2 differences. According to the results from the analyses of WM capacity and L2 proficiency, except for a main effect of Reading span observed in the learners' data of the critical segment, indicating that learners' general reading speed varied according to the size of their working memory capacity, no other statistically reliable main effects of, or interactions with, either Reading Span/WM capacity or L2 proficiency were found. This indicates that these additional factors did not have a significant impact on participants' online ambiguity resolution, and that the observed L1/L2 differences in sensitivity to phrase length manipulation could not simply be accounted for by individual differences in WM capacity or in the learners' L2 English proficiency.

5. Implications and conclusion

In conclusion, we found that whilst the native English controls' ambiguity resolution preferences were not measurably affected by length manipulation in the online SPR task, our Chinese L2 English participants' preferences were significantly modulated by phrase length. The results indicate that learners' ambiguity resolution preferences may be affected more by pragmatic constraints (i.e. the modifiability of the post-verbal NP) than by parsing principles based on hierarchical phrase structures such as *Minimal Attachment* (which are unable to account for the observed interactions with phrase length) or other distance/length-based recency constraints (such as factors of *integration costs*, *distributional information*, and *prosodic heaviness*). Our findings are generally in line with Pan and Felser's (2011) in that learners make greater use of pragmatic cues for interpretation than native speakers, which is an ability that may perhaps help them compensate for their reduced ability to build detailed syntactic representations and apply phrase-structure based parsing principles during online processing. These findings thus appear to support the 'second clause' of *Shallow Structure Hypothesis*, according to which non-native comprehenders are more sensitive to non-structural factors than native comprehenders. In English language classrooms, ESL learners are often found to have difficulty in comprehending English texts and even misinterpret them while reading. Although this could simply be due to lack of enough L2 grammatical knowledge that is necessary for successful comprehension in the L2, the findings of this study provide an implication that learners' failure or difference in interpreting L2 texts might not necessarily involve lack of L2 knowledge – instead, it may possibly be caused by their characteristic shallower processing strategy that makes them nonnative-like.

References

- Allan, D. (2004). *The Oxford placement test [2nd Edition]*. Oxford: Oxford University Press.
- Altmann, G. & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, 30, 191-238.
- Britt, M. (1994). The interaction of referential ambiguity and argument structure in the parsing of prepositional phrases. *Journal of Memory and Language*, 33, 251-283.
- Clahsen, H. & Felser, C. (2006a). Grammatical processing in language learners. *Applied Psycholinguistics*, 27, 3-42.
- Clahsen, H. & Felser, C. (2006b) Continuity and shallow structures in language processing: A reply to our commentators. *Applied Psycholinguistics*, 27, 107-126.
- Clahsen, H. & Hong, U. (1995). Agreement and null subjects in German L2 development: New evidence from reaction-time experiments. *Second Language Research*, 11, 57-87.
- Clifton, C., Speer, S., Abney, S., 1991. Parsing arguments: phrase structure and argument structure as determinants of initial parsing decisions. *Journal of Memory and Language*, 30, 251-271.
- Daneman, M. & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450-466.
- Duffield, N., & White, L. (1999). Assessing L2 knowledge of Spanish clitic placement: Converging methodologies. *Second Language Research*, 15, 133-160.
- Dussias, P. E. & T. R. Cramer (2008). Spanish-English L2 speakers' use of subcategorization bias information in the resolution of temporary ambiguity during second language reading. *Acta Psychologica* 128, 501-513.
- Felser, C. & Roberts, L. (2007). Processing wh-dependencies in a second language: A cross-modal priming study. *Second Language Research*, 23, 9-36.
- Felser, C., Roberts, L., Gross, R. & Marinis, T. (2003). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics*, 24, 453-489.
- Fernández, E. (2000). Bilingual sentence processing: relative clause attachment in English and Spanish. Doctoral dissertation, CUNY Graduate Center, NY.
- Ferreira, F., Bailey, K.G.D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, 11, 11-15.
- Ferreira, F. & Henderson, J. M. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language*, 30, 725-745.
- Fodor, J. D. (1998). Learning to parse? *Journal of Psycholinguistic Research*, 27,

285-319.

- Fodor, J. D. (2002). Prosodic disambiguation in silent readings. *Proceedings of the North Eastern Linguistic Society*, 32, 113–132.
- Forster, K. & Forster, J. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments & Computers*, 35, 116-124.
- Frazier, L. (1979). *On Comprehending Sentences: Syntactic Parsing Strategies*. Bloomington, IN: Indiana University Linguistics Club.
- Frazier, L. (1987). Theories of sentence processing. In Garfield, J. (ed.), *Modularity in knowledge representation and natural language understanding*. Cambridge, MA: MIT Press. 291-307.
- Frenck-Mestre, C. (2005). Ambiguities and anomalies: What can eye-movements and event-related potentials reveal about second language sentence processing? In Kroll, J. and De Groot, A., editors, *Handbook of bilingualism*. Oxford: Oxford University Press.
- Frenck-Mestre, C. & Pynte, J. (1997). Syntactic ambiguity resolution while reading in second and native languages. *Quarterly Journal of Experimental Psychology*, 50A, 119-148.
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68, 1-76.
- Guo, J., Guo, T., Yan, Y., Jiang, N. & Peng, D. (2008). ERP evidence for different strategies employed by native speakers and L2 learners in sentence processing. *Journal of Neurolinguistics*, 22, 123-134.
- Harrington, M. & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition*, 14, 25–38.
- Hawkins, J. A. (1994). *A performance theory of order and constituency*. Cambridge, England: Cambridge University Press.
- Juffs, A. (1998). Main verb versus reduced relative clause ambiguity resolution in second language sentence processing. *Language Learning*, 48, 107-147.
- Juffs, A. (2004). Representation, processing, and working memory in a second language. *Transactions of the Philological Society*, 102, 199-225.
- Juffs, A. & Harrington, M. (1996). Garden path sentences and error data in second language processing. *Language Learning*, 46, 283-323.
- Just, M., Carpenter, P. & Woolley, J. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, 111, 228-238.
- Katsika, K., 2009. Exploring the minimal structure in prepositional phrase attachment ambiguities: evidence from Greek. *Lingua*, 119, 1482–1500.
- MacDonald, M. C. (1999). Distributional information in language comprehension, production, and acquisition: Three puzzles and a moral. In B. MacWhinney

- (Ed.), *The emergence of language* (pp. 177-196). Mahwah, NJ: Lawrence Erlbaum Associates.
- Marinis, T., Roberts, L., Felser, C. & Clahsen, H. (2005). Gaps in second language sentence processing. *Studies in Second Language Acquisition*, 27, 53–78.
- Pan, H.-Y. & C. Felser (2011). Referential context effects in L2 ambiguity resolution: Evidence from self-paced reading. *Lingua*, 121, 221-236
- Papadopoulou, D. (2005). Reading-time studies of second language ambiguity resolution. *Second Language Research*, 21, 98-120.
- Papadopoulou, D. & Clahsen, H. (2003). Parsing strategies in L1 and L2 sentence processing: A study of relative clause attachment in Greek. *Studies in Second Language Acquisition*, 25, 501-528.
- Papadopoulou, D. & Clahsen, H. (2006). Ambiguity resolution in sentence processing: the role of lexical and contextual information. *Journal of Linguistics*, 42, 109–138.
- Rayner, K., Carlson, M. & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior*, 22, 358–374.
- Roberts, L., Gullberg, M. & Indefrey, P. (2008). On-line pronoun resolution in L2 discourse: L1 influence and general learner effects. *Studies in Second Language Acquisition*, 30, 333-357.
- Sanford, A., Graesser, A. (2006) Introduction: Shallow processing and underspecification. *Discourse Processes*, 42, 99 – 108.
- Sanford, A. J. & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends in Cognitive Sciences*, 6, 382-386.
- Spivey-Knowlton, M. & Sedivy, J. C. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition*, 55, 227-267.
- Thornton, R., & MacDonald, M. C. (1999). *The role of phrase length in modification ambiguities*. Manuscript submitted.
- Thornton, R., MacDonald, M. C., & Arnold, J. E. (2000). The concomitant effects of phrase length and informational content in sentence comprehension. *Journal of Psycholinguistic Research*, 29, 195-203.
- Thornton, R., MacDonald, M. C. & Gil, M. (1999). Pragmatic constraint on the interpretation complex noun phrases in Spanish and English. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 25, 1347-1365.
- van Berkum, J., Brown, C. & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language*, 41, 147-182.
- Wasow, T. (1997). Remarks on grammatical weight. *Language Variation and*

Change, 9, 81-105.

- White, L. (2003). *Second language acquisition and Universal Grammar*. New York: Cambridge University Press.
- Whitney, P., Arnett, P., Driver, A., Budd, D., 2001. Measuring central executive functioning: what's in a reading span? *Brain and Cognition*, 45, 1-14.
- Ying, H. G. (1996). Multiple constraints on processing ambiguous sentences: Evidence from adult L2 learners. *Language Learning*, 46, 681-711.
- Ying, H. 2004. Relevance mapping: a study of second language learners' processing of syntactically ambiguous sentences in English. *Second Language Research*, 20, 232-255.

Appendix –List of the experimental items used:

1. Bill/glanced at/the (amazingly rude young) customer/with strong suspicion (ripped jeans)/ and then walked away.
2. Mary/hoped for/the (really great karaoke) prize/with eager anticipation (a lot of cash)/ but got nothing finally.
3. Peter/sought/the (highly competitive corporate) promotion/with firm determination (cash benefits)/but was fired instead.
4. Jenny/looked to/the (highly respected local) priest/with anxious hope (grey hair)/and he helped her.
5. Jack/listened to/the (famous old German) opera/with new earphones (English subtitles)/ and liked it very much.
6. Annie/expected/the (newly upgraded downtown) bus/with much anticipation (air conditioning)/but waited at the wrong stop.
7. Julia/ saw/the (poorly disguised undercover) policeman/with great panic (a machine-gun)/but he didn't see her.
8. William/looked for/the (seriously damaged fishing) boat/with his telescope (colorful sails) /but saw nothing all day.
9. John/ knew/the (highly complex secret) security code/with absolute certainty (thirteen digits)/but never used it.
10. Jane/despised/the (incredibly lazy chemistry) student/with obvious hatred (dirty clothes)/but always treated him fairly.
11. Richard/heard/the (widely advertised student) performance/ with much pleasure (thirty violins)/and loved it all.
12. Andy/strove for/the (internationally competitive academic) award/with great effort (considerable prestige)/ and finally won it.
13. Claire/demanded/the (fully detailed midterm) report/with firm insistence (sales figures)/so we all worked on it.
14. Betty/tasted/the (carefully hidden Christmas) chocolate/with (great delight) whisky flavour/and quickly ate it all.
15. Emma/ loved/the (temporarily unemployed older) man/with all her heart (amazing looks)/ but she feared marriage.
16. Helen/smelled/ the (very expensive French) perfume/with much satisfaction (secret ingredients)/and said it was perfect.