How native-Chinese speakers establish discourse-level mental representations of English-language academic texts: A mixed method study

PhD. Education

Institute of Education: University of Reading

James Patrick Wagstaffe

September 2021

Table of contents

Declaration of authorship

I hereby confirm that this is my own work and that the use of all materials from other sources has been properly and fully acknowledged.

James Patrick Wagstaffe

(14/09/2021)

Acknowledgements

As with all large projects whose achievement is accredited to only one name, behind this PhD thesis there is at least a small platoon of people without whom it could not have happened. Firstly, and most formally, it is important to state that this project was funded, and thus made possible by, the SeNSS/ESRC Doctoral Training Partnership (award no. ES/P00072X/1). Asides from meeting the tuition fees, as well as providing me with a income, they have at every time of asking been fast and generous in acceding to all other request for funding I made along the way to meet the challenges or grasp the opportunities of PhD life. To the whole team at SeNSS, as well as those at the University of Reading who played a role in helping win the scholarship in the first place, I say an enormous thank you! This literally would not have happened without you. As the research conducted was publicly funded through this grant, the data has been made publicly available at the following location: https://doi.org/10.17864/1947.000383

Secondly, I must once again (and again and again) take a moment to thank my wonderful, loving supportive family – and not only for all the last minute proofreading! They have been, as they always have been, the rock upon which I and all my endeavours are built.

Finally, and most importantly, I would like to thank my amazing, fantastic supervisory team, Professor Suzanne Graham and Dr Holly Joseph. I have taken an unusual road into academia. My PhD journey started in a small room in China, where having just completed an MA at distance, and consequently more or less alone, I decided I wanted to take the next step, and get seriously into research. Shrouded from head to toe in doubts about how wise this course of action was, or whether I was up to the task, I wrote up a research proposal, and began sending it out to pretty much every university in the UK or Hong Kong with a research interest in second-language reading comprehension. After what felt like an age of silence, broken only by the occasional 'sorry, we don't feel your project is quite suitable for us', I received an email back from Suzanne. And ever since, I have felt myself to be in the safest of hands. From my initial application to Reading, through the scholarship application, my first steps as researcher, up until the very last day as I sit here just a few hours from submission she has been generous with her time, wise in her advice and unfailingly supportive every step along the way. I really cannot thank her enough.

No less, I must also thank Holly for introducing me to, and then guiding me through, the wacky world of eye-tracking. I cannot say the process of eye-tracking hasn't occasionally driven me completely nuts. It is a technology and approach to research that has the infuriating quality of being just useful enough to make it worthwhile taking another breath and trying to find out why, yet again, a £20, 000 pound machine, whose only function is to track eyes, *still* can't find the damn

pupil. Still, we got there in the end, and I will be back for more. I thank you all from the bottom of my heart.

Abstract

This thesis reports on a mixed methods study which investigated the ability of native-Chinese speaking readers of academic English to identify incongruent words in coreferential noun phrases, as well as the extent to which such an ability may vary as a function of reading proficiency. Additionally, the study also explored the extent to which the presence or absence of a grammatical determiner (e.g. this/these) may contribute towards such an ability. The study was composed of two major elements: an eye-tracking experiment and follow up think-aloud interviews. Reading proficiency was measured according to the participants' IELTS scores. In addition, individual differences, in terms of vocabulary knowledge, grammar knowledge and working memory capacity were explored in relation to the participants' ability to detect the incongruencies. The results of the study suggest that the ability to detect an incongruent word in a text depends primarily on the reader being able to establish a situation model mental representation of that text. Furthermore, the results also suggest that this ability is chiefly, although not entirely, confined to upper-intermediate or advanced-level readers. While the measures of individual differences used in this study did not significantly predict for the ability to identify the incongruent words in the text, analysis of the transcripts from the think-aloud interviews found that the main impediment to the formation of situation model representations of academic text was difficulty with linguistic processing. The implications of these findings are discussed in relation to both the practice and theory of second language reading pedagogy. In terms of practice, the implications include (*i*) the need to develop metacognitive strategy training designed to help L2 readers establish situation model representations of the texts they read; (*ii*) the need for specific instruction on the grammatical construction and cohesive function of complex coreferential noun phrases in academic writing; (*iii*) the potential for using anomaly detection as a tool for measuring L2 reading comprehension; (*iv*) the need for university managers to provide teaching staff with training on how to help L2 readers establish situation model representations of the texts they read. With regards to theory, the results

of the study lend weight to Walter's (2007) suggestion that L2 reading should be understood more in terms of accessing fundamental cognitive processes through the L2, than of a transference of pre-existing reading skills from the L1 to the L2.

Chapter 1: Introduction

The internationalisation of education is a growing global phenomenon, which in recent years has resulted in the intake of large numbers of non-native English speaking students into UK universities, as well as the development of English Medium Instruction (EMI) courses hosted by UK universities operating overseas. As a result of this development, there has been a significant increase in the number and range of university courses being offered to non-native students at both undergraduate and postgraduate level (Dearden, 2014). In the UK context, the largest single group taking up these opportunities to pursue higher education in English are the Chinese, who now account for almost one third of all non-EU overseas students (ukcisa.org.uk, 2018). Given these large numbers of Chinese students, and the correspondingly large contribution they now make to the UK university sector, there is arguably a need, and perhaps even a duty, for specific attention to be given to meeting their needs.

However, despite the opportunities this widening access to UK university education offers, there is evidence to suggest that where students have insufficient language skills to fully participate in their courses, educational outcomes may suffer (Knagg, 2013; Trenkic & Warmington, 2018). Given the importance of reading in tertiary level education, the development of strong English literacy skills is a key component of academic success. As a corollary, in view of the inherently linguistically challenging nature of academic texts, (Halliday, 1993a), the question of how to help second-language students (L2) develop adequate reading skills is one which deserves concerted attention by both researchers and educators. This need to rapidly develop L2 reading skills may be of particular concern in respect of the Chinese student community, owing to the considerable linguistic distance between the English and Chinese languages. The current project seeks to address one facet of this issue, by investigating how Chinese readers establish discourse-level mental representations of academic English texts.

1.1 Discourse comprehension and the linguistic creation of text

While much reading research has concentrated on either word or sentence processing (Rayner, 2009), the focus of this investigation is on the comprehension of extended discourse, of the type students typically encounter during their university education. This level of discourse comprehension requires readers to "discern the author's communicative intent" (Koda, 1993, p. 123) in such a way as to understand the complex messages encoded in academic discourse. Written discourse analysts such as Tadros (1994) and Hoey (2001) argue that written texts function as a site of communicative interaction between reader and writer. To participate fully in this interaction, both parties are required to cooperate on the basis of a shared knowledge and understanding of the linguistic and cultural norms which shaped the construction of the text. Within this view, it is important to note that text does not simply accumulate from the juxtaposition of words, phrases and sentences on a page, but rather is created through the use of a complex array of local lexical and grammatical markers (Halliday & Matthiessen, 2014), which serve first to establish cohesive semantic relationships between the messages in adjacent clauses (Winter, 1994) and then between those in extended sections of discourse, ranging from a single paragraph to many pages of text (Hoey, 2001). As such, it seems clear that developing "an understanding of text coherence and the way it is conveyed is critical for successful text comprehension" (Koda, 2005, p. 128). Although there are many types of cohesive device, one of the most important for establishing coherence in text are *coreferential noun phrases* (Kintsch & Rawson, 2005). These are noun phrases within a text which refer to the same entity or concept, such that a writer may repeat and develop ideas over extended sections of discourse. Perhaps the most common, although by no means the sole, type of coreferential noun phrases used in texts are synonyms, whereby a writer repeats an idea in a text using two different words which share a common meaning. As will be seen in Section 2.2.1 below, within the genre of academic writing, much of the work of argumentation, explanation and exemplification is done through the use of highly complex coreferential noun phrases. However, as will also be explained, there are substantial grammatical differences between the construction of such noun phrases in the English and Chinese languages. As such, if native Chinese speakers are to be empowered to

engage fully with the academic texts they read as part of their university courses, it is important to know how they process coreferential noun phrases in English so that problems, and solutions, can be identified.

1.2 What do we talk about when we talk about reading?

Throughout this thesis, reference will be made to both reading and readers. Given the broad array of issues which fall within the purview of L2 reading research, including such aspects as word access processing (e.g. Fukkink et al., 2005), strategy use (e.g. Brevik, 2019) vocabulary and grammar knowledge (e.g. Zhang, 2012) or cross-linguistic transfer (e.g. Melby-Lervåg & Lervåg, 2011), to name but a few, it is important to define exactly what is meant by the term *reading* in any given study. In the current thesis, following Kintsch (1998), the word *reading* refers to the ability to extract information from a written text with sufficient efficacy and accuracy as to allow the reader to establish a mental representation of the meanings encoded in that text. That is, reading is defined chiefly in terms of *comprehension*. Within this view, reading comprehension is understood to result from the integrated operations of a multi-componential system, composed of three major elements: (*i*) *an automated cognitive processing system*; (*ii*) *a memory-based linguistic and socio-cultural knowledge system*; (*iii*) *a behavioural system*.

The first element in this overarching system, *the automated cognitive processing system*, encompasses two major components: (*a*) word access processing, and (*b*) discourse integration processing. These two process work in a dual system, whereby the reader accesses the meaning of individual words in each sentence, establishes higher order representations of the meaning of the text by integrating the words together to establish sentence level propositional representations, before integrating these individual propositions to establish discourse level representations (see Sections 2.4 and 3.7.4 for a fuller discussion of integration in reading comprehension). The second element in the overall system, *the memory-based linguistic and socio-cultural knowledge system*, includes the knowledge the reader holds in long term memory which must be accessed in order for them to comprehend the text, including *lexico-grammatical* knowledge, *topic* knowledge, and *socio-cultural* knowledge of how the text was created (e.g. text structure,

text type). Finally, the third element, *the behavioural system*, is also composed of two major sub-elements: (*a*) reading behaviours by which the reader (more or less) consciously interacts with the text, primarily in terms of eye movements, and (*b*) metacognitive problem solving strategy use, whereby the reader (more or less) consciously monitors their comprehension of the text, and initiates regulatory procedures where comprehension difficulties occur. Importantly, as will be explained in more detail in the ensuing sections, the view of reading advanced in this thesis posits that individually none of these three major systems constitutes reading *per se*– rather, reading results from the *combined* product of these various operations in terms of the mental representations they allow the reader to establish of the meanings encoded in a text.

1.3 The influence of cross-linguistic transfer

Part of the research in this thesis will address the issue of cross-linguistic transfer, which is to say the influence of a reader's first language on how they process their second (Grabe & Stoller, 2015). While crosslinguistic transfer is an issue that must be recognised regardless of a reader's linguistic background (Grabe, 2009), when considering languages as distant from each other as English and Chinese, it is even more important (Koda, 2005). Perhaps the most immediately apparent difference between Chinese and English is in the orthography. However, while learning to read Chinese characters undoubtedly presents significant challenges to readers whose first language is written in an alphabetic script, this problem is arguably not as severe for native-Chinese speakers who have grown up in the People's Republic of China (PRC) who are learning to read English. As part of the process of learning to read Chinese characters, children in China are taught *pinyin*, a Romanised phonetic script used to represent how each character is pronounced. Research has shown that when native-Chinese speaking children are compared with non-native Chinese speaking children, where both groups speak both English and Chinese, they perform very similarly on their ability to pronounce words written in both English and *pinyin*, even if native-Chinese speaking children show much greater awareness of the lexical tones which characterise the Chinese language (Zhou et al., 2018; Zhou & McBride, 2018). Although anecdotal, this is consistent with the experience I built up

teaching English to L2 students in China over a period of twelve years. During that time, it became apparent that the ability of Chinese high-school and university-level students to process the orthography of English language texts as they read aloud was often deceptively good. When asked to do so, my students could usually read passages up to a page or two in length clearly and fluently, appearing to stumble only when it came to unusually long words. However, when asked to the explain what those texts meant, they often encountered substantially greater difficulty. Although this is undoubtedly an anecdotal observation, and one that, as far as I am aware, has yet to be tested under more rigorous conditions, it was of sufficient concern that when new teachers arrived from the UK to teach in the school in China where I worked, part of their orientation training included a warning not to assume that the students understood a text they had read aloud simply because they had read it fluently.

However, although experience using *pinyin* may provide at least some advantage to native-Chinese speakers from the PRC in terms of processing the orthography of English, they must still contend with a number of other issues. In addition to the marked differences which exist between these two languages in terms of their lexico-grammatical construction, down to the most basic levels of syntax and morphology, more subtle differences also emerge at the level of discourse in terms of the linguistic devices used to establish cohesion in written texts. As Koda (2005, p. 111) points out, "cohesive signalling… is by no means universal across languages. When L2 [cohesive] devices are not prevalent in the L1 they … can create processing difficulties". This non-universality of cohesive signalling between English and Chinese has been demonstrated by a number of studies. For example, Halliday (1993b) conducted a comparison of the different ways that Chinese and English texts construct scientific, and by extension, academic knowledge more broadly. His study indicated that both languages make use of a variety of similar lexico-grammatical techniques, such as the nominalisation of processes (e.g. transform – transformation), or the use of hyponymy and meronymy[1] to create taxonomies, all of which can play a role in the establishment of textual

[1] Hyponymy establishes superordinate categories of the type: ***a*** is a type of ***b*** (e.g. a chair is a type of seat). Meronymy established part-whole relationships of the type: ***x*** is a part of ***y*** (e.g. a leg is part of a chair).

cohesion (Halliday & Matthiessen, 2014). However, he also found that in general Chinese makes less use of explicit grammatical markers to indicate semantic links within and between clauses, preferring instead to leave the reader to infer such relationships for themselves. Similarly, Wu (2014, p. 1659) explains that Chinese "relies more on semantic meaning rather than cohesive devices to achieve coherence", while Yang (2014) argues that although cohesion exists in both Chinese and English as a textual resource, the linguistic distance between these two languages means there are considerable differences in the way this is achieved, going on to explain that even where similarities do exist, for example in the use of linguistic markers of coreference (e.g. *this*, *these*), their frequency of use may be very different. Consequently, it is not clear how well native-Chinese readers of English language academic texts may be able to benefit from the various explicit cohesive device which guide proficient native-English readers as they seek to comprehend complicated academic writing.

1.4 A cross-disciplinary approach to studying reading comprehension

Given the considerable differences which exist between English and Chinese, when investigating how native-Chinese speaking readers process complex academic texts written in English, it is arguably necessary to take a cross-disciplinary approach. As will be explained below, a considerable amount of research has been conducted in a wide a range of academic areas, encompassing Second-Language Education, Applied Linguistics and Psycholinguistics, all of which provide insight into the issue of how second-language readers process academic text. In the field of Second-Language Education, much work has been done investigating the cognitive and metacognitive strategies that second-language readers use, and how these contribute towards reading proficiency (Abbott, 2006; Anderson, 1991; Block, 1986; Ghavamnia et al., 2013; Tsai et al., 2010; Yapp et al., 2021). Similarly, second-language researchers have investigated the role such diverse aspects as linguistic knowledge and working memory capacity play in predicting difference between higher and lower proficiency readers (Brunfaut et al., 2021; Jeon & Yamashita, 2014; Joh & Plakans, 2017; Peng et al., 2018; Yamashita & Shiotsu, 2017; Zhang, 2012). In the domain of Applied Linguistics and Written

Discourse Analysis, researchers have turned their attention to the various systems of cohesion which exist in English, and which allow writers to bind disparate words and sentences together to create integrated texts capable of conveying subtle and complex ideas over almost infinitely extendable sections of discourse (Francis, 1994; Halliday & Matthiessen, 2014; Halliday & Hassan, 1976; Hoey, 2001; Tadros, 1994). In the field of psycholinguistics, investigators of both first and second-language reading comprehension have developed models which seek to explain how readers establish mental representations of the text they encounter, on a word-by-word, sentence-by-sentence, or even paragraph-by-paragraph basis (Bernhardt, 2010; Gernsbacher, 1997; Kintsch, 1998; Perfetti & Stafura, 2014). However, while each of these fields of scholarship have made considerable strides in advancing our understanding of how academic texts are both written and read, a need now exists for the findings from these diverse disciplines to be brought together so that a more holistic three dimensional picture of second-language reading comprehension may start to emerge (Ariel, 2001a; T. J. Sanders & Gernsbacher, 2004).

1.5 The structure of this thesis

In order to address the various issues raised above, the study reported on in this thesis investigated how native-Chinese speaking students enrolled in a UK university processed coreferential noun phrases in academic texts. As will be described below, the design for the study drew on research from a wide array of disciplines, in order to shed light on the specific issues this group of students may face as they attempt not only to read, but also to *learn* from the texts the encounter over the course of their university studies. **Chapter 2** of this thesis presents a review of the literature which informed and contextualised the study, finishing with a statement of the four research questions addressed. **Chapter 3** describes the methodology that was used for the study, including a discussion of the purpose, strengths and limitations of the research design. **Chapter 4** presents the results from a quantitative eye-tracking experiment that was conducted, while **Chapter 5** presents the results from primarily qualitative analysis of follow up think-aloud interviews which were conducted with a sub-sample of the participants who took part in the eye-tracking experiment.

Chapter 6 entails a discussion of the combined findings from the eye-tracking experiment and think-aloud interviews, situating them within some of the issues and controversies which surround the scholarship on second-language reading comprehension. **Chapter 7** sets out the conclusions of the study, as well as implications that the findings may have for the theory and practice of second-language reading pedagogy. Finally, **Chapter 8** sets out the limitations of the study which must be borne in mind when considering its conclusions.

Chapter 2: Literature review

2.1 Introduction

There is a growing view within the field of discourse processing that in order to understand how readers build coherent mental representations of the messages in a text, it is necessary to approach the problem from two distinct, yet inextricably entwined perspectives; namely: the *linguistic*, and the *psycholinguistic* (Ariel, 2001; Sanders & Gernsbacher, 2004). The linguistic aspect of text production has been studied in considerable depth both by discourse analysts (e.g. Francis, 1994; Hoey, 2001; Tadros, 1994) and grammarians (e.g. Halliday & Matthiessen, 2014; Halliday & Hassan, 1976). While this field of inquiry has covered the nature of text and language from a multitude of angles, the one most pertinent to the current study is that of textual cohesion in general, and, more specifically, of *coreference* as a particular form of cohesive device. The psycholinguistic aspect of reading is concerned with the many reader-internal cognitive processes that are involved in the act of extracting and reconstructing the messages that writers encode in their texts, such as to allow the reader to form a coherent mental representation of the meanings therein. While this dualistic view of reading could be applied to all types of written discourse, the discussion here will focus primarily on academic texts. Similarly, while much of the discussion could apply equally to all readers, regardless of their language background, the current study focuses on native Chinese speaking readers of English, as part of the argument is based on the claim that second language [L2] reading is inevitably affected in a multitude of ways by issues of cross-linguistic transfer from the reader's first language [L1] (Koda, 2005). As such, it was decided to focus on a single L1 language group.

In this discussion, I will first consider reading from the linguistic perspective, by conducting a detailed examination of how academic texts are lexico-grammatically created in both English and Chinese, as well as considering some of the problems this may cause Chinese readers as a result of the considerable differences which exist between these two languages. Secondly, I will discuss reading from the psycholinguistic perspective, by reviewing a number of studies which investigate how second-language readers construct

mental representations of the texts they read, with a particular focus on the issue of coreference as an anaphoric device.

2.2 The linguistic construction of academic texts

Discourse analysts have long argued that written texts function as a locus of social interaction between a reader and a writer, each of whom must work cooperatively, and *knowledgably*, if the meaning potential of a text is to be fully realised (e.g. Francis, 1994; Hoey, 2001; Tadros, 1994). Within this account, as Hoey (2001, p. 13) explains, it is assumed that "when writers compose their texts, they draw upon models that have become normal within their cultures; when readers process these texts, they do the same". This view thus assumes that in order to successfully understand a text, a reader must first understand how it was linguistically and discursively constructed.

2.2.1 The linguistically complex nature of academic texts and the reading difficulties this may present for novice readers

This need to come to grips with the linguistic and discursive norms which guide and constrain text production is perhaps nowhere more evident than in the field of academic writing. In an analysis of the grammatical features which characterise scientific, and by extension, other academic texts, Halliday (1993a, 1993b) identifies several areas which have the potential to cause difficulties for novice readers. Of these, the most relevant to the current discussion are those of *lexical density* and *complex nominalisation* as these are most related to the problems faced by native-Chinese readers, an issue which is focused on specifically in Section 2.2.2.

2.2.1.1 Lexical density

Lexical density refers to the number of words in a clause which carry its semantic content, and generally results in a relatively high number of adjectives and nouns compared to either spoken or narrative texts. Halliday (1993a) explains that in unplanned spoken English, lexical density is quite low, with utterances typically only containing between two to four lexical words. In written discourse, by contrast, due to the

higher degree of planning to which it is often subjected, the lexical density rises to around four to six content words per clause. However, in scientific writing, lexical density is frequently of another order of magnitude altogether, rising to as high as ten to thirteen lexical words per clause unit. For example, Halliday (1993a, p. 76) cites the following sentence in illustration of this point:

> Griffith's **energy balance approach** to **strength** and **fracture** also **suggested** the **importance** of **surface chemistry** in the **mechanical behaviour** of **brittle materials**.

Inspection of this sentence reveals two distinct issues. Firstly, from a semantic perspective, the degree of lexical density means that is conceptually extremely rich, and thus potentially hard to understand. Secondly, from a more grammatical perspective, it can be seen that twelve of the thirteen lexical words it contains are packed into just two noun phrases, meaning that it requires the reader to engage in a considerable degree of parsing if they are to identify the sentence's syntactic structure sufficiently well as identify its subject and object, and thereby extract its propositional meaning.

2.2.1.2 Complex nominalisation

This tendency in academic writing to stack nouns together to form lexically and conceptually dense noun phrases extends to another feature of scientific texts which Halliday (1993b) identifies as being potentially problematic for novice readers; namely, that of *complex nominalisation*. Nominalisation is the grammatical process whereby verbs (or more importantly the processes they represent), or adjectives, are changed into nouns, such that their syntactic status may be altered. In the case of *complex* nominalisation, this process is taken one step further, and extra information, in the form of either the subject of the verb, or the circumstances or manner in which it occurs, may also be collapsed into the resulting noun phrases. For example, the clause '*electrons move in orbits'* may be rephrased in nominal form as 'the *orbital movement of electrons'* (example taken from Halliday, 1993b, p. 128). The purpose of such nominalisation in academic writing is frequently to create what Thompson (2004, p. 179) describes as "texture", which is to say the "the quality [that makes writing] recognizable as a text rather than a collection of unconnected words or clauses".

In explaining how texts are cohesively created, Halliday and Matthiessen (2014) illustrate how a clause may usefully be analysed from a number of different perspectives. The most common analysis divides the clause structure into a subject and a predicate, with the subject typically appearing at the head of the clause, for example, *The sun* (subject) *shone brightly* (predicate). However, from a more semantic perspective, clauses may also be divided into what Halliday and Matthiessen (2014, pp. 119-121) call *theme* and *rheme* units. The *theme* portion of the structure, which, like the subject, appears at the head of the clause, carries *given*, i.e. previously established information. The *rheme* portion of the structure carries *new* information, which allows the message of the text to be driven forward. The interweaving of successive clause structures, with information first being introduced as *new* in the rheme section of one clause, before being represented as *given* in the theme section of subsequent clauses, is one of the facets of grammar which allows texts to be experienced as cohesively integrated wholes (Hoey, 2001). According to Halliday (1993a, p. 81), this type of theme-rheme cohesive structure has come to be particularly prized within the language of scientific discourse, as it allows writers "to proceed step by step, with a constant movement from 'this is what we have established so far', to 'this is what follows from it next' [such that] these two parts, both 'the taken from granted' and the new information' [may be] presented in a way that makes its status in the argument clear" . He goes on to explain how "the most effective way to do this, in English grammar, is to construct the whole step as a single clause, with the two parts turned into nouns, one at the beginning and one at the end, and a verb in between saying how the second follows from the first" (Halliday, 1993a, p. 81).

Such nominalised theme-rheme patterning not only serves to create local coherence by linking adjacent clauses, it can also be used to create more global discourse coherence, by establishing clear points of transition between paragraphs or even larger sections of text. In a lexico-grammatical process Francis (1994, p. 83) calls "retrospective labelling", noun phrases of the type described above, although in this case almost always containing a demonstrative pronoun (e.g. *this*, *these*) to signal their anaphoric status, can be used to provide summative, and often even evaluative, encapsulations of a previous section of discourse. Global

coherence can thus be achieved by placing such a discourse encapsulating noun phrase in thematic position of a paragraph initial sentence, thereby establishing its information content as "the frame of reference within which subsequent argument is developed" (p. 85). Importantly, these retrospective labels typically have "a clear topic-shifting and topic-linking function: they introduce changes of topic, or shifts within a topic, while preserving continuity by placing new information within a given framework" (Francis, 1994p. 86), thereby contributing to the overall structural coherence of a text.

From a pedagogical perspective, it is important to note that there are a variety of ways in which noun phrases may be used to create cohesion within a text, each of which may be problematic for novice readers. For example, in the case of Francis' (1994) retrospective labels, the noun phrases frequently serve to characterise the nature of the preceding discourse, by explicitly indicating how it ought to be evaluated, for example, by naming it as a *problem*, a *solution* or an *opportunity.* Similarly, writers may choose to sum up a previous passage of text by using what Halliday and Hassan (1976) term 'general nouns', but which are now perhaps more commonly known as *shell nouns*. These are a class of nouns (e.g. *issue*, *question* or *reduction*) that function almost as deictic pro-forms, that is to say words in a text which depend for a substantial portion of their meaning on the semantic context in which they appear. Finally, these cohesively motivated noun phrases may result from the kind of complex nominalisations described above. Each of these types of noun phrases may cause comprehension problems to novice readers. In the case of evaluative nouns, such as *problem* or *opportunity*, it is incumbent upon the reader to infer the connection between the previous section of text and the subsequent evaluative encapsulation of that text. Similarly, in the case of Halliday and Hassan's (1976) shell nouns, there is a conceptual leap to be made between connecting a single, underspecified nominal label with a preceding passage of text. Finally, in respect of the kind of lexically dense, grammatically complex noun phrases already highlighted, it can be imagined that the sheer amount of information packed into these constructions, coupled with the frequent need to interpret the cohesive

links they establish between both local and global sections of discourse, means they have the potential to cause comprehension difficulties for any novice readers of scientific or other academic texts.

2.2.2 Particular problems for Chinese readers

Given the many complexities which characterise academic discourse, it is not hard to imagine how much more demanding the task of reading such texts may be for non-native speakers, especially those from a language as lexically, grammatically and orthographically divergent from English as is Chinese (Koda, 2005). Firstly, for any non-native speaker, the issue of lexical density described above is likely to cause difficulties, simply in terms of the amount of often rather technical, low frequency vocabulary they need to know. However, the tendency in academic writing to compact ideas down into lexically dense noun phrases is likely to complicate matters still further. The issue of phraseology in academic discourse has come under increasing scrutiny as an area of study by researchers in the fields of discourse analysis, second-language vocabulary acquisition and second-language academic writing (Coxhead, 2008). However, its impact on second-language academic reading appears to have attracted less attention. This is perhaps unfortunate, not least in respect of native Chinese speakers. In a comparative analysis of the construction of Chinese and English language scientific texts, Halliday (1993b) notes that there are a number of similarities between texts produced in these two languages, in terms of both the rhetorical structures and general linguistic devices used. He goes on to argue that these similarities result in part from the fundamental discursive demands of science as a discipline (e.g. the need to create taxonomies). However, he also points out that the considerable lexico-grammatical differences which exist between Chinese and English mean these ends are achieved in very different ways. Of principal relevance to the current discussion is the issue of complex nominalisation as a form of cohesive device, which Halliday (1993b) makes clear both Chinese and English scientific texts use frequently, for the reasons discussed above. However, as will be seen below, the non-inflected nature of Chinese, and its consequent lack of derivational morphology, means the ways in which such nominalisation is achieved in Chinese are very different than in English. Specifically, while in English

verbs or adjectives can be changed into nouns by altering the suffix, for example *rotate* becoming *rotation*, in Chinese, the transformation of a verb or an adjective into a noun can be achieved either by simply altering its position in the syntax, or by adding the character '*xing*' (性) after the original word (Halliday, 1993b). Thus, for example, the Chinese adjective equivalent to the English word *unstable*, '*bu wen'* (不稳), may be converted into a noun, equivalent to the English word *instability*, by creating the phrase '*bu wen xing'* (不稳性) (example from Halliday, 1993b, p. 12). Similarly, the Chinese adjective equivalent to the English word *elastic*, '*tan*' (弹) (which is typically used as a verb equivalent to the English word *bounce*), may be nominalised by adding '*xing*' (性) to form the Chinese noun equivalent to the English word *elasticity*, namely '*tan xing'* (弹性). Examination of these examples makes clear that the way nominalisation works in Chinese is subject to a much lower degree of both manipulation and variation than is the case in English. As such, while the complexities which characterise English derivational morphology may pass unnoticed by native speakers, and may at least have some familiarity to speakers of other Indo-European languages, to Chinese speakers they have the potential to add a considerable burden in terms of efficient text processing (Armstrong et al., 2018; Cameron & Lee, 1999).

Moreover, the problems do not end here. Chinese differs substantially from English not only in terms of the morphological formation of individual words, but also in terms of how it forms multi-part noun phrases of the type which frequently characterise academic writing (Halliday, 1993a, 1993b). Specifically, while in English complex nominalisation makes use of both *pre-positional* and *post-positional* markers, for example by using adjectives, preposition phrases or relative clauses (and not infrequently a combination of all three), Chinese grammar allows only for the use of pre-positional marking, with all modifiers thus appearing before the head noun in the phrase (Halliday, 1993b; Hsiao & Gibson, 2003). For example, in the noun phrase below (adapted and translated from Sielen, 2013), the head noun *soup* is premodified by an indefinite article *a*, the adjective (adj.) *microbial*, and postmodified (*i*) by the prepositional phrase *of algae and bacteria*, and (*ii*) by the relative clause *that needed little oxygen to survive*. Translated into Chinese, the head noun 汤 (*soup*) is

now at the end of the sentence, preceded directly by the adjective *microbial*, with the modifying phrases (***i***) and (***ii***) placed at the beginning rather than the end of the sentence, as was the case in the English equivalent.

English: A (***adj***.) *microbial* **soup** (***i***) of algae and bacteria (***ii***) that needed little oxygen to survive

Chinese: 一种(***i***)由藻类和细菌而(***ii***)不需要氧气为生存 (***adj***.) *的微生物* ***(head noun)*** *汤*

Hence, given when attempting to process complex English noun phrases, in addition to dealing with morphology, Chinese readers are also confronted with substantially more variegated grammatical structures than exist in their native language.

Consequently, given the important cohesive function complex nominalisations of the type described above often play in academic writing, coupled with the reading difficulties they may present to native Chinese speakers, it might be the case that Chinese students are most likely to experience comprehension difficulties at some of the most vital points in the texts they read. Specifically, where ideas are being developed more fully or the discussion moves from one topic to the next. In order to understand this concern in more depth, it is necessary now to look at discourse comprehension from the psycholinguistic perspective.

2.3 The psycholinguistic reconstruction of text – reading in a second language

Having provided an overview of how academic texts are linguistically constructed, the following section will review the existing literature on how texts are psycholinguistically reconstructed during the process of second-language reading comprehension. First, a brief survey of some of the early research on L2 reading is provided, as many of the ideas and terminology generated in these seminal studies continue to shape the debate around the field. Secondly, studies which investigate the various components believed to underpin reading comprehension, such as vocabulary and grammar knowledge or working memory are reviewed, Thirdly, an overview is presented of a new wave of studies which investigate the types of discourse-level

mental representations second-language readers are able to establish as they process a text. Finally, a review is given of a number of studies which have investigated the specific issue of how second-language readers are able to process the explicit markers of cohesion that writers used to bring coherence to their texts.

2.3.1 A brief survey of the early literature on second-language reading comprehension

For nearly three decades, much of the research in second-language reading comprehension has been framed by a question first asked by Alderson (1984); namely, are difficulties in second-language reading chiefly explainable in terms of deficits in a learner's first language literacy ability, or rather weaknesses in their level of second-language linguistic proficiency (Bernhardt, 2010)? When first posing this question, Alderson (1984) noted conflicts in the literature which investigated variance in L2 reading performance, stating that:

> On the one hand, subjects have difficulty understanding texts despite knowing the words and structures, and on the other hand the interpretation of words and syntactic structures – that is, grammar and vocabulary – seems to be the main factor in poorer reading performance in the second-language than in the first language" (p. 2)

In attempting to resolve this conflict, Alderson drew upon two competing schools of thought. One school, which came to be known as the Linguistic Interdependence Hypothesis, held that difficulties in second-language reading stemmed primarily from weaknesses in the learner's first language reading ability (Bernhardt & Kamil, 1995). This view drew directly upon the work of researchers like Coady (1979), who claimed that:

> Many students have very poor reading habits to transfer from their first language, and thus, in many cases, we must teach reading skills which should have been learned in first language instruction (p.12).

Coady' view of reading was based heavily on Goodman's (1973) Psycholinguistic Guessing Game model of reading, and its attendant claims of 'reading universals', i.e. the notion that the process of reading remains

fundamentally the same regardless of the language being read. Goodman's view of reading comprehension assumed that fluent reading is a process whereby a reader first generates hypotheses of what information a text is likely to contain, such as to allow them to make predictions as to what linguistic forms they are likely to encounter while they read it (Grabe, 2009). Thereafter, the reader uses their knowledge of the language being read to 'sample' the text in search of linguistic cues which can either confirm or confound the predictions they have made (Alderson, 1984; Stanovich, 1980a). It was assumed within this model that 'good' readers make strong use of context to generate and test their hypotheses regarding how the text is likely to develop, and test these hypotheses mainly by reference to 'semantic cues' within the text. In contrast, it was argued, weaker readers pay more attention to surface, syntactic features of the text, and are thus hampered in their attempts to build strong semantic representations of its message, as a result of getting bogged down in the many linguistic redundancies that are said to characterise authentic native writing (Goodman, 1973). Consequently, the difference between good and bad reading came to be seen as resulting from the adoption of either 'good' or 'bad' reading strategies. Adherents of the Linguistic Interdependence Hypothesis argued that the key to successful second-language reading was simply to *transfer* the reading strategies developed in L1 reading to the new L2 context (Cummins, 1981, 1998). As a corollary, they believed that problems in L2 reading resulted primarily from a failure to adequately transfer those strategies, for whatever reason.

However, concerns were soon raised regarding the validity of this view of the second-language reading process, perhaps most notably by Clarke's (1980) introduction of the Short Circuit Hypothesis of ESL Reading. This hypothesis, whilst accepting the fundamental tenets of the Psycholinguistic Guessing Game model of reading, with its emphasis on good and bad reading strategies, argued that before these strategies could be successfully transferred from the L1 to the L2, the reader would need to reach a certain *linguistic threshold* in their L2 knowledge (Clarke, 1980). Owing to this assertion, the Short Circuit Hypothesis also came to be known as the Linguistic Threshold Hypothesis (Bernhardt & Kamil, 1995). Clarke (1980) reported two

experiments both of which worked with adult native-Spanish speaking readers of English. In the first, 21 participants were asked to complete cloze tests in both Spanish and English, and the results compared to see what differences could be found between the performance of 'good' and 'bad' first language readers. Their first language reading proficiency was assessed according to their performance on the Spanish language cloze, and answers considered unacceptable were analysed for "syntactic and semantic acceptability" (p. 204). The analysis rested on three assumptions. Firstly, Clarke assumed that:

> in a cloze test, a response which is totally syntactically and semantically acceptable indicates that the subject has understood what s/he has read. Responses which are not acceptable provide evidence about the processes used by the reader in responding to a mutilated text (p. 204)

Secondly, it was assumed that good readers would focus on larger sections of the text than weaker readers when trying to complete the cloze tests, and also "will rely on semantic cues rather than syntactic cues" (p. 204). Finally, the assumption was made that "that, given equivalent English proficiency, the difference between good and poor readers would be basically the same in both language" (p. 204). As a result of this analysis, Clarke (pp. 204-205) claimed that:

> in their native language, the good readers seemed to rely on semantic rather than syntactic cues. Forty-one percent of the unacceptable responses of the good readers were judged to be semantically acceptable with minor syntactic adjustments, compared to 25% for the poor readers [but that] the poor readers relied on syntactic cues more than did the good readers (50% versus 35% syntactically acceptable).

However, when results from the cloze tests taken in the participants' second language were compared, Clarke reported that the situation changed, so that although the good readers produced more semantically acceptable responses, their advantage reduced from 16 to four percentage points. From this he concluded that "The difficulties of reading in a second language seem to have reduced the distinction between good readers and poor readers" (p. 205)

The second of the studies in Clarke (1980) investigated what he called the 'oral reading' ability in Spanish and English of two readers, who had either been judged to be strong or weak readers in their native

Spanish, although in this case it was not reported how this designation was made. The study used a miscue research procedure, which:

> requires subjects to read a passage in its entirety and to relate as much as they can remember after they have finished. The errors produced during the reading are analysed on a number of linguistic levels and their comprehension score is calculated from their retelling (p. 205)

The results of this second study were reported to be similar to the first, in that:

> The good reader produced fewer miscues than did the poor reader both in Spanish and in English [and was] was evaluated by independent judges to have understood more of readings than did the poor reader. However [when reading in English] the good L1 reader appeared less able to focus on semantic cues in the target language than in the native language… causing him/her to revert to poor reader strategies when confronted with a difficult or confusing task in the second language (pp. 205-206)

Clarke attributed this reversion from so called good to bad reading strategies to the fact that "limited control over the [second] language 'short circuits' the good reader's system" (p.206),

In reviewing these two schools of thought, with reference to a large body of studies on L2 reading that had been carried out to date Alderson (1984) concluded that while only moderate to low correlation had been demonstrated to exist between first and second-language reading ability, a considerable amount of evidence existed which supported the notion of a minimum linguistic threshold which must be attained before pre-existing reading comprehension abilities can be transferred from the first to the second language. However, he went on to caution that the exact nature of this threshold remained unclear, and needed to be considered in the light of a great many unanswered questions, including:

> What the nature of this threshold is: to what extent is it syntactic, semantic, conceptual, discoursal? Does the level of the threshold vary for different learners, and for different tasks? Is it conceivable that good first-language readers will require a lower threshold before being in a position to utilize their good reading strategies? Will the attainment of a higher level of competence compensate a poor first-language reader? (Alderson, 1984, p. 20-21)

Following Alderson's (1984) discussion, many of the questions he raised were subsequently investigated by L2 reading researchers throughout the 1990s (Bernhardt, 2010). For instance, Taillefer (1996) investigated the question of whether task complexity might influence the linguistic threshold which must be attained before previously acquired reading strategies can be transferred from the L1 to the L2. Taillefer recruited 55 native-French speaking second-year university students, all of whom had studied English for seven years in secondary school, and for a further year in university. However, despite having studied for the same number of years, the participants represented a range of English language proficiency levels. The participants selected for the study had all performed equally on a French language reading test in a previous study, but were divided into a high L2 linguistic proficiency group, and a low L2 linguistic proficiency group. The study used two tasks to measure L2 reading proficiency. The first task, *scanning*, she defined as cognitively easy, the second, *receptive reading*, she defined as cognitively hard. In the scanning task, the participants were asked to scan a long and a short document and highlight each occurrence of two key words and two key figures which they had been told to look out for. They were given a score for the number of instances they spotted. In the receptive reading task, the participants were asked to pretend they were preparing for a debate by reading a text to understand the author's point of view – that is, they needed to do more than simply remember the text, they were encouraged to really *understand* it. In order to measure their comprehension of the text, the participants conducted two further tasks. Firstly, they were asked to respond to a series of statements by saying that the author of the text read either would or would not agree with them. Secondly, they were presented with a summary of the text which had five sentences underlined, which they were told were potentially problematic, as well as with two possible corrections for each one. The participants had to judge if (*i*) the sentences were in fact problematic, and (*ii*) if they thought they were, which would be the best correction. Thus, for each of the five sentences there were three possible versions, only one of which was an accurate reflection of what was said in the text. When performing these tasks, the participants were not allowed to look at the text they had read, as Taillefer wanted them to " reflect on what [they had] understood instead of merely identifying points in an

available text" (p. 465). The participants' second-language proficiency was measured using vocabulary and grammar tests taken from the TOELF exam and a cloze test, and their French language reading proficiency was also measured using a cloze test. The results of multiple regression tests performed as part of the analysis showed that both first language reading proficiency and second-language linguistic proficiency were significant predictors of second-language reading proficiency. However, this effect was mediated by the type of reading tasks the participants were engaged in. In the cognitively 'easy' scanning task, L1 reading proficiency accounted for 10.6% of the variance, while L2 linguistic proficiency did not make a significant contribution at all. By contrast, in the more difficult receptive reading task, second-language linguistic proficiency "weighed in three times stronger than L1 receptive reading [as measured by] *t* values." (p. 469), suggesting that as the cognitive difficulty of the reading task increases, L2 linguistic knowledge becomes a more important factor. From these results, Taillefer (1996, p. 470) concluded in respect of the Linguistic Threshold Hypothesis that "the language threshold is not an issue for a lower order cognitive task in L2, whereas on a higher order task, limited L2 proficiency short circuits the transfer of L1 reading abilities to the L2 contexts". Thus, her findings provided support for the existence of a linguistic threshold that second-language readers must reach if they are to understand a text in detail, such as to make judgements about the opinions expressed by the author, rather than simply engage in relatively easy tasks such as scanning.

This position of limited synthesis between the variables of L1 reading ability and L2 linguistic proficiency, whereby each is argued to make a distinct, albeit unequal contribution to second-language reading, has been supported by other studies. For example, Bernhardt and Kamil (1995) conducted research on native English speaking readers of Spanish at three different levels of development, *beginner*, *intermediate* and *advanced*, with the specific aim of investigating the Linguistic Independence Hypothesis versus the Linguistic Threshold Hypothesis. With this end in mind, Bernhardt and Kamil measured both the participants' level of English reading comprehension, using a standardised reading comprehension test, as well as their level of L2

Spanish proficiency. By conducting regression analyses on the data collected, they were able to claim that "[L1] reading variables account for between 10 and 16 per cent in second-language reading, [L2] language proficiency accounts for 30 to 38 per cent" (p. 25). These findings led them to conclude that:

> these data seem to indicate that first language reading ability is a very important variable in second-language reading achievement, however, in those studies that are able to account for language proficiency, this construct seems to be a substantially more powerful predictor of L2 reading ability (p. 30).

Bernhardt and Kamil's (1995) findings were replicated in a host of other research projects throughout the 1990s (Bernhardt, 2010) to such an extent that by the early 2000s Bernhardt felt able to claim that:

> Considering that cognate and non-cognate languages were studied; both children and adults were considered; and different measurement schemas employed, the studies produced remarkably consistent findings: They all estimated the contribution of first-language reading to second-language reading to be between 14% and 21% and the contribution of language knowledge to second-language reading performance to be around 30%. (2005, p. 137)

2.3.2 The components of the second-language reading comprehension: Linguistic knowledge and the Threshold Hypothesis

Building on this early work, much of the contemporary literature on second-language reading comprehension has focused on using statistical analysis to identify the various components which underpin second-language reading comprehension, as well as the relevant contribution each of them make to the reading process (Jeon & Yamashita, 2014). For example, Zhang (2012) used structural equation modelling to investigate the relative contribution of vocabulary and grammar knowledge to the reading comprehension ability of 190 advanced native-Chinese speaking learners of English. In this study, vocabulary knowledge was measured separately in terms of both breadth and depth of knowledge, while grammar knowledge was measured separately in terms of both explicit and implicit knowledge. The results suggested that together vocabulary and grammar knowledge accounted for 81.1% of the variance in reading comprehension scores, thereby lending more support to the

Threshold Hypothesis. However, they also indicated that only *implicit* grammar knowledge made a unique significant contribution after vocabulary knowledge had been accounted for.

A related study by Jeon and Yamashita (2014), who conducted a meta-analysis of 58 published studies drawn from 29 peer-reviewed journals focused on applied linguistics, first and second-language reading and literacy studies as well as education arrived at similar conclusions. In their study, the authors identified four correlates of reading comprehension which have been widely researched; namely, *L2 decoding*, *L2 vocabulary knowledge*, *L2 grammar knowledge*, and *L1 reading comprehension*. In addition, they identified a further six which have received less attention, but which nonetheless have been demonstrated to be correlated with successful L2 reading comprehension; *L2 phonological awareness, L2 orthographic knowledge, L2 morphological knowledge, L2 listening comprehension, working memory,* and *metacognition*. Similarly to Zhang (2012), Jeon and Yamashita found evidence to support the existence of a linguistic threshold, identifying second-language linguistic knowledge as accounting for the largest share of the variance in L2 reading comprehension, although in this instance grammar accounted for a larger proportion than vocabulary, at 72% and 62% respectively.

A number of more recent studies have also supported these findings (Cheng & Matthews, 2018; Sparks, 2015; Yamashita & Shiotsu, 2017). For example, Yamashita and Shiotsu (2017), who grounded their research in Gough and Tunmer's (1986) Simple View of Reading, investigated the relative contributions of L1 reading ability, L2 listening ability and L2 linguistic knowledge to second-language reading comprehension. The Simple View of Reading is a framework which posits that reading comprehension is essentially the combined product of word level decoding skills and oral language, i.e. listening comprehension skills. Yamashita and Shiotsu gathered data from 325 university level Japanese speaking learners of English. Using structural equation modelling, their results showed that L2 linguistic knowledge was highly correlated (r= .86) with L2 reading comprehension skill, regardless

of the participants' level of L2 proficiency. L2 listening comprehension also correlated highly with L2 reading comprehension, although there was more variation between L2 proficiency levels, with correlations raging from *r*= .90 when the entire sample was analysed to *r*= .71 when a subsample of high proficiency participants were analysed separately. Finally, L1 reading comprehension was found only to correlate moderately (*r*= .45) with L2 reading comprehension ability. From this Yamashita and Shiotsu (2017, p. 56) concluded that their results "mirror the finding reported in many previous studies that have supported the Threshold Hypothesis".

2.3.3 The components of the second-language reading comprehension: working memory capacity and the role of prior knowledge

Another issue which has received considerable attention in both the first and second-language literature on reading comprehension is working memory capacity (Grabe & Stoller, 2015). In second-language acquisition studies, working memory capacity has been found to be associated with both proficiency and processing outcomes (Linck et al., 2014), although variation has been identified in the effects of working memory capacity arising from such issues as the age of the learner or the type of task they are engaged in (Juffs & Harrington, 2011). Within the L2 reading comprehension literature, while the exact role of working memory remains the subject of some debate (Joh & Plakans, 2017), a growing number of studies have emphasised the relationship between working knowledge and prior knowledge, in terms of both L2 linguistic as well as relevant topic knowledge. For example, working with a sample of 94 adult learners of Spanish as a second language, Leeser (2007) investigated the relationship between what he referred to as 'topic familiarity' and working memory capacity on the one hand, and reading comprehension and processing of future tense morphology on the other. The participants read texts on both familiar and unfamiliar topics, their comprehension of which was measured using recall protocols, and also completed "form recognition and tense identification test to determine processing of future tense morphology" (p. 229). In addition, Leeser used reading span tests to measure their working memory, and used the results to categorise the participants as having either

low, medium or high working memory capacity. The results of the analysis suggested that "topic familiarity was the most consistent variable, yielding significant results on all dependent measures (i.e., comprehension recall, form recognition, and tense identification)." (p. 252). However, although Leeser found a significant relationship between working memory and the participants' comprehension recalls, "post hoc analyses revealed that learners benefited from higher working memory only if they were familiar with passage topics." (p. 252). From this, he concluded that "the results lend further support to previous research that possessing relevant background knowledge promotes better L2 reading comprehension" (p. 253).

Leeser's results were supported by a more recent study by Joh and Plakans (2017), who investigated the relationship between working memory, L2 linguistic knowledge and topic knowledge of the text being read. Working with 80 university Korean students of ESL, they defined topic knowledge as "knowledge of the words/phrases appearing in a given text that directly pertain to the topic of the text" (p. 111), and measured linguistic knowledge using a 60 item test, half of which measured grammar knowledge and the other half general vocabulary knowledge. Working memory was measured using a reading span test, similar to the one used by Leeser (2007). The participants' reading comprehension proficiency was measured by having them read eight passages between 80 and 400 words long, on a variety of topics, and then completing both multiple choice questions and questions that required a short written answer. The results suggested that:

> working memory significantly predicted L2 reading comprehension only when the readers had sufficient knowledge, especially knowledge of the topic in the given text. Otherwise, their comprehension performance was mostly determined by their L2 linguistic knowledge, even when they had considerable working memory capacity. (p. 107)

From this, Joh and Plakans (2017) concluded that working memory capacity only plays a facilitating role in L2 reading comprehension when the reader has sufficient "knowledge of topic-related vocabulary in a given text" (p. 115).

Finally, a related series of studies by Alptekin and Erçetin (2010, 2011) investigated the relationship between working memory capacity, topic familiarity and L2 reading comprehension. However, in approaching this issue, they sought to introduce a greater degree of specificity to what is meant by the term reading comprehension, building on Kintsch's (1998) Construction-Integration (CI) model of reading comprehension (see Section 2.4 for a description of the CI model), by dividing it into two types of comprehension; namely, *literal* and *inferential* comprehension. Literal comprehension they defined as referring to the reader's understanding of the propositions stated directly in the text. Inferential comprehension they defined as that part of the reader's understanding of information which was not directly stated in the text, and thus needed to be inferred. Firstly, Alptekin and Erçetin (2010) explored the role of first and second-language working memory capacity in both literal and inferential L2 reading comprehension. Working with a sample of 43 Turkish university students, Alptekin and Erçetin had them read a short story in English and then complete a twenty-item multiple choice comprehension test. Half of the questions were designed to measure literal comprehension by testing knowledge of information that was explicitly mentioned in the test. The other half were designed to measure inferential comprehension by testing knowledge of content for which the reader needed to make either connective or elaborative inferences to understand. In addition, the participants completed a reading span task in both English and Turkish. The results of their analysis only found significant correlations between L2 working memory capacity and the students' scores from the questions designed to measure inferential comprehension. From this, Alptekin and Erçetin concluded that L1 working memory plays little to no role in L2 reading comprehension, and that L2 working memory plays a role in inferential comprehension, but not in literal comprehension.

Subsequently, Alptekin and Erçetin (2011) examined the relationship between working memory capacity and topic familiarity on the one hand, and literal and inferential comprehension on the other. Working with 62 advanced readers of English from a Turkish university, they measured the participants' L2 working memory again using reading span tests. In addition, in order to test the effect of topic familiarity, they had

the students read an autobiographical short story, written by an American writer, about immigrants in New York City in the 1900s who were struggling to adapt to their lives in the New World. Half of the participants read the original version of the text, while the other half read a version which had been *nativized*, a process which "involves the sociological, semantic, and pragmatic adaptation of the textual and contextual cues of the text into the reader's own culture-specific mental framework, while keeping its linguistic and rhetorical content essentially intact" (p. 246). For example, the setting of the story was adapted so that *New York* was changed to *Istanbul*, and *Brooklyn* was changed to *Taksim*. Similarly, cultural locations such as *churches* were changed to *mosques*, and the names of famous contemporary politicians were changed so that US president Taft was changed to Turkish Prime Minister İnönü. In addition, adaptions were made to make the nativized version of the text more culturally appropriate for the period in Turkish history; for example, a couple who were dating in the original version were changed to a couple who were engaged to be married. After they had read the texts, the participants' comprehension was measured using multiple choice questions which were again written to test both literal and inferential comprehension. The results of the analysis revealed that that, in general, the participants performed significantly better when answering literal than inferential comprehension questions. However, they also indicated that while the participants with both high and low working memory capacity performed similarly on the literal comprehension questions, the participants with high working memory capacity performed significantly better on the inferential questions, regardless of which version of the text they read. Similarly, it was found that readers who read the nativized version of the text performed significantly better on the questions measuring inferential questions than those who read the original text, but that no such difference was found to exist in relation to the literal questions; no significant interactions were found however. From these results, Alptekin and Erçetin (2011, p. 258) concluded that:

> The findings suggest that content familiarity, working independently from working memory, improves inferential comprehension by providing more opportunities for higher-level operations to cater to the situation model of interpretation, yet content familiarity does not seem to affect

> lower-level operations characteristic of literal understanding. As such, the findings confirm the generally held view that content familiarity has a positive effect on readers' performance but the findings delimit this effect to inferential comprehension only

Taken together then, the results of these various studies suggest that although working memory is strongly associated with L2 reading comprehension, it is a relationship which depends to at least some degree on the reader's prior knowledge of the topic of the text. Furthermore, it may also be that the role of working memory is confined to inferential comprehension, rather than comprehension as a general construct.

2.3.4 Comprehension monitoring and the Error (Anomaly) Detection Paradigm

In addition to linguistic knowledge and working memory capacity, another area which has received a considerable amount of attention in the literature on second-language reading comprehension is the use of metacognitive strategies (Grabe, 2009). Metacognition includes a wide range of issues, including goal setting, generating and maintaining motivation and progress monitoring (Hu & Gao, 2017). However, as will be explained in more depth in Section 3.3.3, one of the central research design features of the current project was the use of an Anomaly Detection Paradigm; consequently, the area of metacognition most relevant to this thesis is **comprehension monitoring**. For reasons of space, this discussion will only focus on that issue. Comprehension monitoring is the mechanism whereby skilled readers evaluate their comprehension of a text, and is argued to happen because readers "constantly ask themselves if what they are reading makes sense" (van der Schoot et al., 2012, p. 1666). Although comprehension monitoring was once described as a "neglected essential" in ESL reading (Casanave, 1988, p. 283), its importance is now widely recognised in both the first and second-language reading comprehension as central to reading success (Hessel et al., 2021; Perfetti & Stafura, 2014; Raudszus et al., 2019). One of the most commonly used methods to research comprehension monitoring is an Anomaly Detection Paradigm, sometimes known as an Error Detection Paradigm (Hessel et al., 2021; Morrison, 2004). The design and conclusions of such studies rest on a simple premise. Errors, of one sort or another, are embedded in a text. If the reader notices the

error, they are judged as having been able to monitor their comprehension, if they do not notice it, they are judged as having been unable to monitor their comprehension (Morrison, 2004). However, the question of what exactly allows comprehension monitoring to occur, at least in so far as the ability to detect errors or anomaly in a text is concerned, remains the focus of much research attention. One particular issue that has been investigated is the extent to which a reader's ability to monitor their comprehension is affected by whether or not they are reading in their first or second language. A number of studies have been conducted to research anomaly detection in second-language reading, focusing on such issues as incongruencies between long-distance grammatical dependencies such as relative clause or clitics (Stepanov et al., 2020) or conflicting grammatical gender (Bultena et al., 2020). However, of most relevance to the current thesis are Morrison (2004) and Hessel et al. (2021) as they focused primarily on the issue of semantic rather than grammatical integration. For example, Morrison (2004) worked with a sample of 52 native-English speaking undergraduate students enrolled in a Canadian university. The participants did two reading comprehension tests, one in English (L1) and one in French (L2). In addition, they also did a cloze test to measure their linguistic proficiency of French in terms of lexical and morpho-syntactic knowledge. Finally, they also completed an error detection task in both French and English. The texts they read had two types of errors embedded in them. The first type, which Morrison called *micro-level* errors involved inserting graphemic errors in homonymous words, for example changing the word *exp**e**nse* to *exp**a**nse* in the phrase "promoting one candidate at the expense of another" (p. 89). The second type of error, which Morrison called macro-level, or discourse-level, errors, she defined as "words that contradict information presented in the preceding or following sentence(s)."(p. 89). As they read the texts, the participants underlined any inconsistencies which they noticed, and were also asked to offer a credible solution to fix the anomaly. The results of the study found that, on average, the participants performed weakly at detecting the errors in the text in both their first and second language. In English, on average they were found to have detected only 54.21% of the incongruencies, although the scores ranged from a minimum detection rate of 20% to a maximum rate of 100%. When reading in French, the average detection rate was lower, with only 38.75%,

although again there was considerable variation, ranging from a minimum of 0% detected up to a maximum of 72.22%. In terms of the types of errors detected, the results showed that in English, micro-level errors were detected more frequently than macro-level ones, at a rate of 66.36% and 45.22% respectively. However, when in French, this trend was reversed, with participants identifying 41.98% of macro-level anomalies and 35.16% of micro-level ones. Furthermore, using Pearson's Product Moment found significant positive corelations between first and second-language reading comprehension in terms of the participants' overall performance as well as their ability to detect both micro and macro level error. However, paired-sample *t*-tests indicated that while the difference between the average L1 and L2 detection rate of micro-levels errors (66.36% and 35.16% respectively) was significant, the difference between the average detection rate of macro-level errors (45.22 and 41.98% respectively) was not. Finally, the analysis also discovered significant positive correlations between reading proficiency and the rate of error detection in both the first and second language. From these results, Morrison (2004) concluded that since the participants' performance when detecting macro-level errors was similar in both their first and second language, it may be that "readers tackle L2 higher-level tasks in the same manner as equivalent L1 tasks (p. 95)... suggesting there is a transfer of abilities from the L1 to the L2" (p. 97).

A more recent study by (Hessel et al., 2021) looked at comprehension monitoring, as defined as the ability to detect anomalies in a text. In this study, Hessel et al. used eye-tracking to compare the comprehension monitoring abilities of 63 Year 5 students in a UK school who were either native speakers of English (*n* = 40) or spoke English as an additional language (*n* = 24). The participants read 20 two sentence stories. The first sentence of the stories consisted of a simple event, for example an unspecified animal barking. The second sentence continued the story in such a way as to be either congruent or incongruent with the first sentence, in this case, for example, by featuring a dog (*congruent*), or a cat (*incongruent*). The participants read ten congruent stories, and ten incongruent stories, in randomised order, and eye movement data was gathered on the target word (i.e. the word which was either congruent or incongruent)

in each story. The results of the study showed that “Go-past duration, rereading, and total times were longer for target words read in the inconsistent condition and they received more rereading and regressions in” (p. 165), although there was no main effect for congruency on the gaze duration measure (see Section 3.3.1 for a full description of the various measures used in eye-tracking). In other words, the results suggested that although the presence of an incongruent word did not effect initial processing of that word, the incongruency did affect integration processing, suggesting that, on average, the participants were able to notice them. When the issue of language background was taken into account, although it was found that the participants who spoke English as an additional language read all of the target words for longer than their native-English speaking counterparts, “language group did not interact with inconsistency on any eye movement measure, confirming that the two groups of children did not differ in their online monitoring of inconsistent target words.” (p. 165). In terms of the effect of vocabulary size, the results suggested when it came to reading the target words in the congruent condition, participants with larger vocabularies read the target words faster than those with smaller vocabularies. Furthermore, participants with larger vocabularies were more likely to regress to, that is to say reread, the target words in the incongruent conditions that their counterparts with weaker vocabulary knowledge. From these results, Hessel et al. (2021, p. 170) concluded in respect of language background that:

> our data are clear in showing no group difference in children’s processing of inconsistencies... This further supports the notion that any qualitative differences in online reading between bilingual and monolingual samples are better understood as differences in vocabulary knowledge rather than a reflection of categorical differences between first and second-language speakers.

Consequently, while both Morrison (2004) and Hessel et al. (2021) concluded that readers are able to monitor the performance similarly in both their first and second language, Morrison framed the issue as being one of a strategy transfer from the first to the second language, whereas Hessel et al. noted the centrality of vocabulary knowledge as the operative factor, thereby at least tacitly lending weight to the Linguistic Threshold Hypothesis.

2.4 Reading comprehension as the establishment of mental representations

The growing interest in the role of prior knowledge in relation to working memory has coincided with a trend in both first and second-language reading studies to investigate not only the components which underpin reading comprehension, but also the extent to which L2 readers are able to establish coherent discourse-level representations of the texts they read (Raudszus et al., 2019), as well as the knowledge resources they depend upon to do so. For example, Walter (2004) investigated the ability of L2 readers to transfer reading comprehension skills to their second language. Walter defined comprehension skill in terms of Gernsbacher's (1997) Structure Building Model, having observed that previous attempts to investigate how such transfer might take place had neglected "to define the nature of the comprehension skills that initially fail to transfer and later do transfer" (p. 315). The Structure Building Framework argues that reading comprehension is, at its heart, a process whereby readers establish new memory structures based on the information they extract from written texts. Gernsbacher (1997) explained that there are three main stages involved in this process: *laying a foundation*, *mapping* and *shifting*. In the first stage, as the reader starts processing the text, relevant memory nodes are activated to form the bedrock foundation of the structure being built. Next, as the reader processes more of the text, assuming it is on a connected topic, closely related memory nodes are activated and mapped onto the foundation. The more coherent the new information is with that already active in memory, the more likely it is to be mapped onto the foundation of the structure. Finally, once a reader encounters information that is no longer closely coherent with that already active in text, for example because of a change in topic at a paragraph break, they shift and begin a new information structure.

In explaining the view of reading comprehension Walter had thus adopted, she stated that "in skilled comprehension, readers… build large cohesive, hierarchical structures, in which each new [semantic] element, as it is integrated [into the reader's mental model of the text], activates preceding related elements" (2004, p. 316). Based on this definition, she argued that "one way to gauge the skill of the reader in building mental structures is to measure their success in recalling earlier information from the text" (2004,

p. 317), and that one possible measurement of successful recall was anaphoric resolution, that is the ability to identify the referent of anaphoric marker such as a pronoun or repeated noun. This is because in order to resolve the anaphor, the reader must be able to integrate the idea or entity it represents with an idea previously read. In the study, Walter tested a group of 41 native-French speaking high school students of English, divided into a lower-intermediate group and an upper-intermediate group (containing 19 and 22 participants respectively), in order to investigate the relationship between their level of reading comprehension and their ability for anaphoric resolution. This was taken as a proxy for their ability to build coherent mental structures of the texts read. In the trial, the students were first required to read a number of stories in English and French, and then complete a summary exercise for each one in order to test their understanding. Texts in both languages were used to ascertain whether they were 'skilled comprehenders' in French, as well as to establish their ability to decode at the sentence level in the English, so that this could be ruled out as the source of any later L2 comprehension difficulties. Secondly, the participants were given an anaphor resolution task in which for each anaphor participants were asked to explain orally 'what does this word mean' as a prompt to identify its antecedent. The test contained two types of anaphoric referent: (*i*) immediate, defined as cases where the anaphor was in either the same clause or the previous clause from its antecedent; and (*ii*) remote, defined as cases where the anaphor and its antecedent were more than two clauses apart. In addition, the participants completed a word span test of verbal working memory.

Perhaps unsurprisingly, Walter (2004) reported that both the upper-intermediate group and the lower-intermediate group scored better on the summary task for the stories read in French than for those read in English. However, it was also noted that the difference between the lower-intermediate group's scores for the French and English language stories was significantly larger than was the case for the upper-intermediate group. Given that the texts read had been written at a level that purported to ensure that the lower-intermediate level group could process them at the sentence level, Walter concluded that the difference between the two group's ability to produce summaries for the French and English language text resulted from the participants in the upper-intermediate group having been more successful than the lower-

intermediate group at transferring their reading comprehension skills from the L1 to the L2. Furthermore, analysis of the anaphor resolution test revealed a number of interesting findings. Firstly, it was found that, overall, the upper-intermediate students resolved the anaphors more successfully than their lower-intermediate counterparts. Secondly, for the upper-intermediate group, it was found that in both the French and English texts immediate and remote anaphors were resolved similarly well. However, while the lower-intermediate students were generally able to resolve both immediate and remote anaphors in the L1 texts, in L2 they were able to resolve the immediate references, but not the remote ones. Given that immediate anaphoric references were defined as being those in adjacent clauses, Walter (2004) concluded that both the upper and lower-intermediate groups had proved adept at sentence-level processing in both the L1 and L2. By the same token, she further concluded that as the remote anaphors were defined as being those more than two clauses as apart, the ability to process these could be taken as indicative of the participants' ability to establish coherently structured discourse-level representations of the texts. As such, since the upper-intermediate students were able to resolve these remote anaphors in both the L1 and the L2, while the lower-intermediate students could resolve them in their L1, but not in their L2, this failure suggested that the lower-level students' ability to establish structured discourse-level representations of the texts had been negated by their relatively low level of proficiency in the L2. With regards to the role of working memory, Walter reported that:

> having higher verbal working memory corresponded to being better at reading comprehension for both groups [but that] this was significantly more the case for the lower-intermediate group, [among whom] small advantages in working memory give significant advantages in comprehension (p. 332)

From this, she concluded that:

> a crucial element in transferring reading comprehension skills from L1 to L2 appears to be the transfer of structure building ability; and this ability appears to be linked to the development or working memory in L2 (p. 333)

That Walter (2004) defined reading comprehension in terms of Gernsbacher's (1997) Structure Building Model had a significant impact on the conclusions she later drew from these results (Walter, 2007). The Structure Building Model proposes that when constructing mental representations of a narrative, readers use the same domain-general cognitive architecture regardless of whether the stimulus input for the story is in the form of oral language, written language or in picture form (Gernsbacher, 1997). Building on this model, and the results of her 2004 study, Walter (2007) concluded that it could be mistaken to consider reading comprehension skills as being 'transferred' from the L1 to the L2. Rather, she hypothesised that if reading comprehension, understood as the ability to construct coherently structured discourse-level representations of a text, is domain general, as opposed to language specific, then rather than comprehension skill being transferred from the first language to the second, it may be the case that L2 readers become able to *access* fundamental cognitive processes through the second language. She thus proposed that the term access might be a more appropriate metaphor than the more commonly used *transfer* for understanding how established L1 reading comprehension skills may become available for use in the L2.

In order to test this theory, Walter (2007) recruited 20 lower-intermediate, defined as middle school students, and 22 upper-intermediate, defined as upper-school native-French speaking readers of English. All of the participants were asked to read a total of 19 narrative texts, in both English and French, which contained semantic anomalies that had been purposefully embedded in the text. The anomalies were of two types, main point anomalies and subsidiary point anomalies. This was done as the Structure Building Framework predicts that main point anomalies should be easier to detect than subsidiary point anomalies, as main points in the text should have higher levels of activation as a result of being read more frequently. Consequently, information in the text which contradicts one of these main points should be more salient to the reader. Walter reasoned that if her participants were indeed able to detect main point anomalies more readily than subsidiary point ones, this would indicate that they had been able to access their fundamental comprehension skills, through their second language. The participants were told that there were a number

of contradictions in the text, and asked to press a button on the computer they were reading on if they detected one, and then to tell the researcher what they thought the anomaly was. The results of the study indicated that "both groups performed well on L1 anomaly detection, with no significant differences in their scores" (p. 28). From this, Walter concluded that "the lower-intermediate readers were competent comprehenders in their L1 [so that] any differences in L2-based comprehension cannot be attributed to maturity of intellectual development" (p. 29). Furthermore, while both the upper and lower-intermediate readers were better able to detect main point than subsidiary point anomalies in the L2, their ability to do so was weaker in the L2 than in the L1. However, the results also showed that the lower-intermediate readers' performance declined to much greater extent than the upper-intermediate readers when they were reading in their second language. In explaining this finding, Walter (2007, p. 29) concluded that:

> the results are consistent with the hypothesis that lower-intermediate learners, when reading texts in their L2, are unable to access the structure building skill that they deploy well in identical circumstances in their L1. It is a problem of access.

In seeking to explain why lower-intermediate readers may be unable to access structure building skills, Walter focused on the issue of working memory (WM), arguing that (p.30):

> differences in verbal L2-based WM scores, and differences in correlations between verbal L2-based WM and anomaly detection performance, correspond to the hypothesis that the capacity of verbal L2-based WM is inordinately taken up by lower-level processing for Lower-intermediate learners (p.30).

Another series of studies which investigated L2 reading comprehension in terms of the kind of mental representations readers are able to establish were conducted by Ushiro et al (2017, 2020). However, rather than referring to the Structure Building Framework, they grounded their view of reading comprehension in Kintsch's (1998) Construction-Integration model. In seeking to provide a psycholinguistic account of discourse processing, Kintsch (1998) explained that there are two major levels at which readers create a mental representation of a text's message, the *textbase* and the *situation model*. In addition to this primary

division, a second key distinction is made between three types of representation that comprise the textbase; namely *propositional*, *microstructure* and *macrostructure* representations (Kintsch, 1998). Propositional representations are those which are formed as the reader processes the linguistic surface structure of individual sentences. Microstructure representations are local discourse-level representations that result from the integration of adjacent propositions into a higher order representation of their combined meaning. Macrostructure representations result as microstructure representations are subsumed into a representation of the main themes and topics of the text read. While in reality these various representations are not developed in a unidirectional, linear fashion, for ease of understanding it is helpful and sufficient to first consider them in this way (Kintsch & Rawson, 2005). Psychologically speaking, the microstructure of the text results from the combination of individual propositions encoded at the clause level, "according to the words of the text and their syntactic relationships and by analysing the coherence relations among these propositions, which are often, but not always, signalled by cohesion markers at the linguistic level" (Kintsch & Rawson, 2005, p. 314). The second level of representation, the macrostructure, results from the identification and organisation in memory of the global topics the text contains (Kintsch, 1998), and it is usually only these more global themes that are subsequently available for recall after the reading of a text has been completed (Kintsch & van Dijk, 1978). For example, when reading a newspaper article, while details such as the names of participants or specific locations might be forgotten after a few days, the reader is likely to remember at least the main events which were described, which country they occurred in, as well as their personal feelings about those events. The final level of representation, the situation model, occurs when the reader integrates the messages in the textbase with any relevant prior knowledge they may hold of the topic, thereby creating "a mental model of the situation described by the text." (Kintsch & Rawson, 2005, p. 315). Importantly, it is not until the situation model has been established that readers can be said to have fully 'understood' a text (Kintsch, 1998). Thus it is important that all three types of representation are successfully achieved for reading comprehension to occur. Since its inception, Kintsch's (1998) Construction-Integration model has been used to frame research into reading comprehension from a variety

of angles, including the role of translation in second-language reading (Kim & Clariana, 2015), anaphor resolution (Varma & Janssen, 2019), the knowledge representations L1 children establish of digital texts (Fesel et al., 2015), as well as both first and second-language discourse processing (Li & Clariana, 2019; Zwaan, 2016).

Ushiro et al (2017), explored the potential of a reading intervention aimed at Japanese EFL university students to improve their capacity to establish situation model representations of narrative texts. Similarly to Walter (2007), Ushiro et al used an anomaly detection paradigm to investigate the ability to establish discourse-level representations. Specifically, in the experiment, the participants read twelve passages, in random order, each of which were composed of six sections, described as follows (p. 95):

i. an *introduction* (two sentences) introducing a main character of the story,
ii. an *elaboration* (three sentences) describing the character traits that were either consistent or inconsistent with target sentences,
iii. a *filler* (four sentences) intervening between elaborations and target sentences,
iv. a *target* sentence stating a character's action relevant to the prior traits,
v. a *post-target* sentence describing another action of the character unrelated to the consistency with elaborations,
vi. a *closing* (two sentences) concluding the story.

The target sentence (*iv*) contained information that was either consistent or inconsistent with a character trait given about one of the characters in the story in the elaboration section (*ii*). The target and antecedent sentences were placed several lines apart so that the researchers could explore the readers' ability to detect inconsistencies at the global rather than the local level. Six of the passages were provided in what Ushiro et al. referred to as the *standard instruction condition*, in which the participants were told simply to read the texts for understanding. The other six were presented in what they called the *situational-instruction condition*, in which the "participants were instructed to mentally visualize the texts while reading in order to comprehend the situation described there." (p. 96). The researchers looked for signs of processing disruption at two time points; during initial reading and during any subsequent rereading of the target

sentence (see Section 3.3.1 for a detailed discussion of eye-tracking measures). The results of the experiment indicated that “participants detected global inconsistencies more in the situational condition than in the standard condition” (p. 99); that is, they argued, when the readers had been able to establish a situation model representation of the text.

Following on from this, Ushiro et al (2020) conducted a similar eye-tracking experiment, in which they investigated whether Japanese university students could more readily detect inconsistencies in a narrative text which related to protagonists, temporal aspects of the story, or spatial ones, arguing that these represented different aspects of the situation model. Following a procedure similar to the one used in their 2017 study, the researchers found that the participants were able to identify inconsistencies related to the protagonists, but not those related to either temporal or spatial aspects of the story. However, the results differed somewhat between the two studies in terms of at what point in the reading process disruption occurred. While in the first study, participants showed evidence of disruption during both initial and rereading of the target sentences, in the second experiment, there was only evidence of disruption during rereading, as indicated by a higher number of regressive eye-movements when the target sentence was inconsistent with the antecedent. From this, they concluded that “EFL readers understand intersentential protagonist links via late conscious or strategic process”(p. 107). Given that regressive eye movements are argued to reflect comprehension processing rather than initial word processing (Staub & Rayner, 2007), Ushiro et al concluded that EFL readers are able to establish situation models of narrative texts well enough to understand details of the protagonist. As a corollary, from the fact that the participants did not show evidence of having detected inconsistencies relating to the temporal and spatial aspects of the stories, they concluded that the “situation models constructed by EFL readers are more limited than those of L1 readers in terms of (a) the number of dimensions understood, and (b) they ways they process inconsistent information” (p. 108).

Support for the view that the establishment of a situation model is necessary for complete comprehension of a text was also provided by Raudszus et al. (2019), who compared how L1 and L2 fourth-grade children establish discourse-level representations of expository texts. Working with a sample of 76 monolingual and 102 bilingual children, they investigated the relationship between reading comprehension on the one hand, and textbase memory as well as situation model formation on the other, to see how these two factors might interact with a number of other cognitive and linguistic variables, including such issues as linguistic knowledge, working memory and decoding skill, that have been shown to affect reading comprehension. Unlike Walter (2007) and Ushiro et al (2017, 2020) who used an anomaly detection paradigm to investigate the extent to which L2 readers are able to establish discourse-level mental representations, Raudszus et al used a pathfinder network approach. Pathfinder analyses:

> build on the assumption that the mental representation of a text can be approximated by a network of concepts and their relations, properties of a reader's representation of the text can be examined... The similarity of a child's text representation to the literal text can be seen as an indicator of textbase memory, while similarity to an expert model can be seen as a proxy for situation model quality (Raudszus et al.,2019, p. 107)

In order to assess the participants' ability to establish textbase and situation model representations, they were asked to read a 500 word text on malaria, a subject which was known to be unfamiliar to them. After reading the text, the children were shown a computer screen with words taken from the article, and instructed to drag related words closer together, and words they felt to be unrelated further apart. In this way, it was argued, it would be possible to develop an picture of the kind of mental representations they had been able to establish of the relationships between the concepts in the text. The conceptual maps that the participants produced in this way were then compared with those of an expert, to see how accurate the children's situation model representations were. In addition, the participants' maps were also compared with the sequence the information was displayed in the original text to see how closely there textbase memory representations aligned with the original.

The results of the analysis revealed that for both L1 and L2 readers, "general reading comprehension was positively correlated with vocabulary, grammar, textbase memory, and situation model building ability... [and for the L2 readers] a positive correlation between general reading comprehension and decoding was found" (Raudszus et al.,2019, p. 113). In addition, in line with the other studies reviewed so far, vocabulary (8.6%) and grammar knowledge (7%) were found to be the strongest predictors of general reading comprehension, as measured by a standardised reading test, followed by situation model building (1.8%). However, once shared variance was taken into account using a commonality assessment, it was found that grammar knowledge explained 26% and vocabulary 22.6% of the variance in combination with other predictors. Situation model building accounted for 14% of the variance, once combined with vocabulary, grammar, textbase memory and nonverbal reasoning. Textbase memory, which did not act as a unique predictor of general reading comprehension, accounted for a small proportion of the shared variance at 8%, in concert with vocabulary, grammar, nonverbal reasoning and situation model formation. Given that the ability to establish situation model representations served as both a unique and shared predictor of general reading comprehension, and that both the L1 and L2 readers performed similarly in terms of their ability to establish situation model representations, despite the L2 readers exhibiting deficit in terms of both vocabulary and grammar knowledge, Raudszus et al. concluded that situation model building is an important predictor of both first and second-language comprehension in addition to linguistic factors.

2.5 The role of explicit cohesive devices in second-language reading comprehension

Another issue which has received considerable attention in the literature regarding both L1 and L2 reading comprehension is the processing of various cohesive devices (Grabe & Stoller, 2015). The primary goal of discourse processing is to allow the reader to establish a coherent mental representations of the text (Gernsbacher, 1997; Kintsch & Rawson, 2005). One of the key lexico-grammatical features of the texts which allow this to occur are cohesive devices, including such items as conjunctions (e.g. and, but, because) and pronouns on the grammatical side, as are direct and indirect repetitions of content words (e.g. synonyms)

on the lexical side (Halliday & Hassan, 1976). The following section will review some of the literature relating to how second-language readers process conjunctions as well as markers of coreference, as these are the cohesive devices most frequently used for establish passage level coherence (Halliday & Matthiessen, 2014).

2.5.1 Processing conjunctions in second-language reading comprehension

Conjunctions, sometimes referred to as connectives, are a wide category of cohesive devices used to explicitly indicate to the reader the logical relationships which exist between ideas in the text. In this regard, they are semantic in nature (Halliday & Matthiessen, 2014). Connectives can be used at any discourse level, binding clauses together at the intrasentential level, sentences together within paragraphs, and paragraphs together to produce extended sections of text (Thompson, 2004). Perhaps unsurprisingly, given their ubiquity in both narrative and expository texts, conjunctions have been widely studied in both the first and second-language literatures on reading comprehension. For example, Van Silfhout et al. (2015) investigated the extent to which first language Dutch eighth grade students (11 – 12 years old) are able to derive benefit from the presence of conjunctions in expository and narrative texts. Using eye-tracking, they worked with a sample of 141 eighth grade children taken from three academic tracks: preuniversity, high-level prevocational, and low-level prevocational. The participants' level of reading proficiency was first measured using a standardised Dutch reading test. After that, they were asked to read four stimulus texts, two expository and two narrative, from which data on their eye movements were gathered. The narrative texts were chosen so that they included typical features of this text type, including dialogue and emotional reaction. The expository texts included one on biology, and the other on economics. For both text types, the researchers manipulated one text so that it contained conjunctions of various types, and the other so that it did not. After reading each text, the participants completed tasks designed to test their comprehension at both the local and the global level. To measure their comprehension of local coherence, they answered what Van Silfhout et al. (2015, p. 55) described as "open-ended bridging inference questions to capture higher-order processes such as inferences and activating prior knowledge". In order to measure their

comprehension of global coherence, the participants were asked to reorder sets of sentences which had been taken out of the texts, so that "to order the sentences correctly, students had to use both their world knowledge and the knowledge they obtained from the text about the temporal/linear sequence of text events" (p. 56). Data was gathered on three regions of the sentence (see Figure 2.1)

Figure 2. 1: Regions of interest from Van Silfhout et al. (2015, p. 58)

(7) A plant makes its own food.

[That's why]$_{\mathbf{0}}$ [we call]$_{\mathbf{1}}$ [a plant]$_{\mathbf{2}}$ [a producer.]$_{\mathbf{3}}$

(8) A plant makes its own food.

[We call]$_{\mathbf{1}}$ [a plant]$_{\mathbf{2}}$ [a producer.]$_{\mathbf{3}}$

The result indicated that the participants did receive processing benefit from the presence of conjunctions in the text, regardless of their level of proficiency level or the text type being read. Analysis of the eye-tracking data suggested that the participants processed the text in region 1 (in this case '*we call*') directly after the conjunction (in this case '*that's why*') in region 0, in the first pass reading time, i.e. from when their eye first entered region 1 until it first departed. Furthermore, in region 2, analysis of the regression path duration, i.e. the time during which the eye moved left again after first entering a region (presumably to reread previous text), they exhibited shorter processing times when there was a conjunction in the sentences versus where there was not. From this, the researchers inferred that even when they needed to reread a previous part of the text to clarify their understanding, they were able to do so faster when there was a conjunction to help them understand the logical relations being established. From these results, Van Silfhout et al. (2015, p. 64) concluded that:

> Connectives seem to assist readers in establishing coherence relations between clauses. Based on our online data, we propose that connectives function as '**processing instructions**'... by signalling how two clauses should be integrated at positions at which students do not have enough information to establish the coherence relation.

However, although conjunctions may act as what Van Silfhout et al. describe as processing instructions in second-language learners, the conditions under which L2 readers are to take advantage of them has been the subject of debate for several years. For example, Geva (1992) investigated the extent to which L2 readers of English, at a variety of levels of proficiency are able to comprehend the meaning of conjunctions to establish discourse relations at a variety of levels. Working with a sample of 100 second-language English speakers, from a variety of different language backgrounds, she tested their knowledge of conjunction at three discourse levels; intrasentential, intersentential and in paragraphs. Intrasentential comprehension she measured using a 30-item fill-in-the-blank task, in which the participants had to choose from four options to complete the second half clause of a two clause-sentence, depending on the conjunction shown at the end of the first clause. Intersentential comprehension she measured using a sentence follow on task, in which the participants had to choose between four sentences to decide which most logically followed a two-clause sentence, depending on the logical relationship that pertained between the first and second sentence, as indicated by the conjunction in the first. Knowledge of conjunctions at the paragraph level Geva tested by means of a cloze test, in which the conjunctions were deleted from a paragraph, requiring the participants to choose from a range of options to choose the conjunction that best signalled the logical relations implied by the text. In addition, the participants were also asked to read three one-page expository text, each of which came in three versions. The *explicit* version contained conjunctions, the *implicit* version had the conjunctions removed, and the *highlighted* version had the conjunctions typed in bold face text. The texts were counterbalanced so that each student read a text in each of three conditions. After reading each text they answered four comprehension questions focused on the logical relations in the text. The scores were then added together so that each participant had a score totalled out of twelve – as such, the exact relevance of the three versions of the text is somewhat unclear. This final score was then used as a measure of the participants' ability to comprehend academic text. Finally, a measure of the participants' level of oral English proficiency was obtained by asking their English language teachers to grade their proficiency on scale of 1 (*low*) to 7 (*high*) according to a number of predetermined criteria. This produced a range of scores from 2 to

5. The results of this study revealed that participants varied in their knowledge of conjunctions at all three discourse levels according to their proficiency level. As their oral proficiency grew, so too did their performance on all of the tasks undertaken. It was also found that the lower proficiency speakers did significantly worse on the academic reading task than those at a higher proficiency. Finally, Geva (1992) performed hierarchical regression analysis in which the participants' scores from the academic reading test were taken as a dependent variable, and the scores from three test of knowledge of conjunctions at different discourses levels as three independent variables. The findings revealed that the results from the paragraph-level cloze test were the strongest predictors, accounting for 21% of the variance. The second strongest predictor was the scores from the intersentential test, accounting for 9.5% of the variance. However, the scores from the lowest discourse level, the intrasentential level, did not act as a significant predictor at all. From these results, Geva (1992) concluded firstly that the ability to understand conjunctions at the intrasentential level does not, in and of itself, allow the readers to comprehend the logical relations that pertain at a more global discourse level. Secondly, given that there was a positive relationship between real proficiency and performance on all of the tasks performed in the study, she further concluded that:

> L2 students with better oral language proficiency in the context of an academic environment demonstrate a better ability to deal with the logical implications of conjunctions in academic reading tasks at all levels of discourse [so that] those L2 learners who are more proficient in the L2, in terms of their lexicon and various aspects of syntactic knowledge, are better able to process and integrate information at more global levels in reading tasks (p. 743).

While it must be noted that details were not provided as to the exact nature of the scale used to determine the participants' level of oral proficiency, meaning it is unclear what exactly is meant when Geva referred to "lexicon and various aspects of syntactic knowledge", the results of the study do suggest that the Threshold Hypothesis discussed above (Section 2.3.2) may hold true in respect of how L2 readers process conjunctions as markers of the logical relations pertaining between the ideas in a text.

That this may indeed be the case was further supported by a more recent study by Ozono & Ito (2003). In their study, they divided 60 second-year students of English in a Japanese university into proficiency groups according to their results on a specially devised cloze task which tested knowledge of grammar and vocabulary. The participants then completed tasks which tested their knowledge of conjunctions expressing three types of logical relationship; *for example* (illustrative), *therefore* (causative) and *however* (adversative). They read six passages, two for each type of logical relation. In the fourth sentence of each passage the target conjunction was blanked out, and the participants were asked to choose either *for example*, *therefore* or *however* to fill the gap depending on the logical requirements of the sentence. In each case, they were also given the option of *in addition*, which was used as a distractor. The participants were awarded one point for each correct answer. The results of this tasks revealed a number of differences between readers at different levels of English proficiency. Firstly, while the participants in the high proficiency group were more successful at using, in running order of success, *however*, *therefore* and *for example,* the lower proficiency students were more successful at using *for example*, *therefore* and *however*. Furthermore, although the higher proficiency selected from between the three types of conjunctions evenly, and did not select the distractor, the lower proficiency participants chose *for example* much more frequently than the other two options, and frequently also chose *in addition*, which was never the right answer. What is more, the results suggested that when the lower proficiency students did use *however*, they generally did so incorrectly, and were only slightly more successful in their use of *therefore*. From these results, Ozono and Ito concluded that, for L2 readers with lower language proficiency, illustrative logical relations are easier to understand than causal ones, and that adversative relations are the hardest of all. Positing a potential explanation for this finding, the authors suggested the difference between higher and lower proficiency readers in this regard may be due to the amount of cognitive load they have to bear as they read. While Ozono and Ito (2003) did not frame their conclusion in this way, it is also arguable that the relationship between language proficiency and the ability to process conjunctions they uncovered may also lend weight to the existence of a

linguistic threshold second-language readers must achieve before they can process increasingly complex logical relations.

However, this need to achieve a certain linguistic proficiency before being able to process conjunctions notwithstanding, there is also clear evidence from a number of studies which have directly compared L1 and L2 readers to suggest that under the right circumstances both adult and children second-language readers can derive benefit from the presence of conjunctions as processing instructions. For example, Bosch et al., (2018) used eye-tracking to compare how 27 first language Dutch speaking children and 19 first language Turkish speaking children studying immersively in a Dutch speaking school processed conjunctions as they read. Tests carried out prior to the eye tacking experiment indicated that both sets of participants performed equally in respect of cognitive features such as working memory. However, the native-Turkish speakers performed significantly worse than the L1 Dutch speakers on vocabulary and comprehension tests. During the eye-tracking, the participants read 16 two clause sentences in Dutch, each of which included a causal relationship between the two clauses. Half of the sentences included an explicit Dutch conjunction, and in the other did not include a conjunction, so that the causal relation needed to be inferred implicitly. After reading each sentence, the participants were asked a comprehension question to test if they had accurately comprehended the causal relationship established. In addition to this eye-tracking task, the participants also completed a grammaticality test to measure their level of syntactic knowledge. The data thus gathered were subjected to linear mixed effects modelling, in which the dependent variable was the average processing time for each of the words in the sentence. This was calculated by establishing the total reading time for each sentence, and then dividing that by the number of words in the sentence. In this way, Bosch et al, (2018, p. 62) hoped to "correct for differences in sentence length (due to the manipulation of connective presence)". The independent variables were the presence or absence of a connective, the participants' language back ground (L1 or L2) and their level of syntactic knowledge. The results of this

analysis did not find a main effect of any of these variables on the average processing time. However, once the sample was divided by language background, the results showed that:

> L2 readers with lower syntactic knowledge had longer processing times for sentences without a connective than for sentences with a connective [suggesting that] individual differences in children's syntactic knowledge accounted for the effect of coherence marking on online sentence processing for L2 readers, but not for L1 readers (p. 64).

Hence, the study showed that although there may be a linguistic threshold that second-language readers need to achieve to make use of conjunctions, it may also be the case that under certain circumstances the presence of explicit cohesive markers may also facilitate reading comprehension in L2 readers with weaker syntactic knowledge.

Finally, a study by Degand and Sanders (2002) provided evidence that this benefit from explicit cohesive markers is not only limited to second-language children, but can apply to adults as well. Working with a sample of 31 university level native French speakers who were studying Dutch, and 23 native Dutch speakers who were studying French, Degand and Sanders investigated the impact of two types of connectives, conjunctions and signalling phrases, on both first and second-language expository text comprehension. Each participant read a total of 18 expository articles, of around 250 words each, nine in French and nine in Dutch. Each text could be seen in one of three possible versions. One version had only conjunctions, one used only longer signalling phrases which directly stated the logical relation (e.g. *this happened because*...), and one contained no connectives at all. After each article, the participants answered four short answer questions, two of which directly tested knowledge of the target logical relation, and two tested knowledge of some other idea stated in the text. In addition, the participants' L2 proficiency was measured using standardised tests. The results suggested that in every case the participants performed better in their first language than in their second. It was also found that the native-Dutch speaking participants performed significantly more strongly in their second language that the native-French speaking students. However, these differences in

proficiency notwithstanding, the results of the analysis revealed that there was a significant main effect for the text version, such that "the impact of the text condition (connective vs. signalling phrase vs. implicit) applied equally to the two language groups (French and Dutch speaking) and it was not different in the first or second language of the participants" (p. 750). Specifically, it was found that the participants performed significantly better on the comprehension questions when they had read texts with connectives which explicitly denoted the logical relations than when these relations had been left implicit, although no significant differences could be found between the texts containing conjunctions and those containing longer signalling phrases. From these results, Degand and Sanders (2002, p. 751) concluded that "linguistic markers help readers construct a coherent cognitive representation of the information in the text", regardless of whether they are reading in the L1 or their L2.

The results of the studies reviewed above thus suggest that second-language readers can, under certain circumstances, derive benefit from the presence of conjunctions as explicit cohesive devices signalling the logical structure of a text. There is important evidence to suggest that the Threshold Hypothesis may apply to the processing of conjunctions in text, such that higher proficiency readers benefit more from such explicit cohesive devices than those at a lower level of proficiency. However, there is also evidence to suggest that the presence of conjunctions in a text may also provide support to readers with weaker syntactic knowledge, by helping them perceive the logical relations between the ideas in the text, thereby helping them to establish coherent discourse-level mental representations of the meaning of the text.

2.5.2 Processing explicit anaphoric markers of coreference in second-language reading comprehension

Explaining how readers process texts in order to establish textbase representations, Kintsch and Rawson (2005) stated that one of the primary means by which propositions may be cohesively related to each other is coreference, whereby two or more textual elements refer to the same semantic entity. The most common type of coreference is anaphor, which constructs a relationship between a newly introduced entity and one which has been previously established (Kintsch, 1998). In linguistic terms, such anaphoric relations are

established either grammatically through the use of personal or demonstrative pronouns, or lexically, through the use of, amongst other things, the kind of complex noun phrases (e.g. the orbital motion of electrons) described in some detail above (Halliday & Hassan, 1976). In psycholinguistic terms, coherence is in part established through a process called anaphor resolution, whereby a reader identifies the semantic relationship pertaining between an anaphor and its antecedent sufficiently well to *integrate* the information encoded in that anaphor into their existing representation of the text, thereby updating their mental model of the text's meaning (Kintsch & Rawson, 2005). Thus, it can be seen that the ability to correctly resolve anaphors is of central importance to successful reading comprehension, and, furthermore, that this ability rests to some extent upon the reader's capability to process the surface linguistic forms in which those anaphors are encoded.

While there has not been a tremendous amount of research conducted into L2 discourse processing in general, or the effects of anaphor resolution on L2 reading comprehension in particular (Grabe, 2009; Koda, 2007), a few relevant studies do exist. For example, Demel (1990) investigated the relationship between overall reading comprehension and comprehension of personal pronouns in a narrative text, working with English as a Second Language (ESL) students in an American University. For the study, participants first read a copy of a short story, and then wrote a summary of the text. The extent to which they had comprehended the story was judged by the degree to which they accurately recalled its main points. Secondly, the participants were presented with a copy of the text in which the pronouns used in the story had been underlined, and were instructed to match the pronouns to their antecedents. Analysis of the results indicated that "there is a significant relationship between overall comprehension and the comprehension of coreferential ties for L2 readers" (Demel, 1990, p. 275). Interestingly, a procedurally almost identical study by Demel (1994) conducted with university level second-language readers of Spanish produced extremely similar results, thus suggesting that the relationship between anaphor resolution and reading comprehension is not unique to English.

While both of Demel's studies were conducted using narrative texts, Pretorius (2005) identified a similar relationship between successful anaphor resolution and reading comprehension in the context of reading for academic purposes. In a study involving 68 ESL students in a South African university, Pretorius used an anaphor identification task in which participants had to match anaphors to their antecedents in texts of the type the students read for a psychology course. In this study however, rather than measuring the students' comprehension of the individual texts read, Pretorius compared their success in the anaphor resolution task to their overall level academic success (as measured by their final exam scores), arguing that the students' ability to succeed academically on their courses depended to a large degree on their ability to successfully read the relevant course material. The results of the study found that there was indeed a highly significant relationship between the participants' overall academic performance and their ability to resolve the anaphors in the texts read. Furthermore, in contrast to Demel (1990), who found no link between L2 proficiency, the ability to resolve anaphors and reading comprehension, Pretorius (2005) found a significant relationship between L2 proficiency and both successful anaphor resolution and overall academic success, reporting that, "together, L2 proficiency and anaphoric resolution contributed 54% of the variance in academic performance" (p. 532). This link between L2 proficiency and the ability to resolve anaphors in Pretorius' study may have resulted from the fact that she tested a much wider range of anaphoric cohesive devices than Demel (1990, 1994), who tested only personal pronouns. Pretorius, by contrast, investigated the participants' ability to resolve five kinds of anaphoric device; namely: repetitions, personal pronouns, synonyms, and determiners (either *this* or *these* used on their own rather than as part of a longer noun phrase). Her results suggested firstly that some types of anaphoric reference were easier to resolve than others, and that the ease with which they were resolved depended in part upon their grammatical complexity. Specifically, pronouns were the most commonly resolved, followed (in order) by synonyms, repetitions, paraphrases and determiners. Interestingly, Pretorius stated that she expected repetitions to be the easiest type of anaphor to resolve. However, it transpired that they were harder to resolve than both pronouns and synonyms. It is worth noting in this regard that Pretorius (2005, p. 528) defined repetitions as,

"anaphoric ties where the anaphor is an exact repetition of an antecedent item or a close repetition, in that a derivational morphological change occurs", citing as an example the coreferential pair 'hostile' and 'this hostility'. In explaining why this type of repetition may have proved harder to resolve than expected, Pretorius suggested it may have been due to deficiencies in the students' ability to interpret derivational morphology in English. It is also perhaps noteworthy that what Pretorius called 'paraphrases' were the second hardest type of anaphor to resolve. In her study, paraphrases were defined as anaphors containing "the determiner this or such followed by a single noun that paraphrases or summarizes the contents of the antecedent" (p. 530). In this regard, they are similar to Francis' (1994) retrospective labels (see Section 2.1.2.). Taken together, these two findings provide some support for the concern that where coreferential cohesive links in a text are achieved through either the use of morphological inflection or the use of discourse encapsulating noun phrases, this may cause comprehension difficulties for ESL readers at important points in the development of a text. More concerning still, the results of Pretorius' study also suggested that difficulty in resolving anaphors was strongly associated with lower levels of academic success. Furthermore, this relationship was strongest for the more complex forms of anaphoric devices, including repetitions and paraphrases. Given that use of these types of linguistically complex anaphoric devices is one of the principal characteristics of academic writing (Halliday, 1993a, 1993b), this may go some way to explaining the relationship between anaphor resolution and academic achievement identified in Pretorius' study.

2.6 Conclusion

So, where does this leave us? L2 reading research generally supports the idea that there is some kind of linguistic threshold which much be reached before readers can efficiently comprehend text. While on the one hand this may feel intuitively obvious, the exact nature of this threshold remains elusive. Furthermore, while the belief that second-language readers must learn to transfer their pre-existing reading skills from their first to their second language has been adhered to for many years, this belief has now been challenged.

Walter's (2007) suggestion that successful second-language reading comprehension depends less on the transfer of comprehension skills from first language to the second, and more on accessing fundamental cognitive comprehension processing skill through the second language has found support in a number of other studies. Specifically, researchers working with both adults and children have suggested that that in order to fully comprehend narrative texts, it is necessary to establish situation model representations of their meaning; that is, second-language readers must be able to integrate the information in the text with information they already hold in long-term memory in order to understand how it relates to the world beyond the text. It may thus be the case that the linguistic threshold which learners must meet before they are able to read efficiently in the second language is set at the level which allows them to access these fundamental, amodal cognitive processes which allow for the establishment of discourse-level representations of the text. Nonetheless, although studies of narrative processing have suggested that the ability to establish situation model representations is critical for the comprehension of narrative texts, as far as could be ascertained, no studies have investigated how such an ability may be related to the reading of expository texts. Moreover, although evidence exists to show the importance of establishing situation model representations, Kintsch's (1998) Construction-Integration model of reading comprehension makes clear that these representations do not exist independently, but rather as composite mental structures composed of information in the text which triggers the activation of information in long-term memory. Consequently, questions still remain about how second-language readers are able to establish the textbase representations generally needed for situation models to occur.

Additionally, in respect of a linguistic threshold, it is still unclear which aspects of linguistic knowledge are critical for the accessing of certain cognitive processes. While vocabulary and grammar knowledge have been repeatedly shown to account for the largest proportion of variance in second-language reading success, the research has also made clear that there are other factors at play. One such factor is the issue of textual cohesion. Research into how second-language readers process conjunctions in both expository and

narrative texts has shown that, under certain circumstances, they can derive processing benefit from the presence of such explicit cohesive devices. Yet here also there may be a linguistic threshold that must be met. In respect of anaphoric markers of coreference, evidence exists to suggest that there is direct relationship between the ability to process pronouns and other coreferential noun phrases and the ability to comprehend second-language writing. Given that the current study was motivated by concerns that native-Chinese speaking students enrolled in UK universities may struggle to learn from the texts they read as a result of difficulties with linguistic processing, the evidence reviewed above suggests these concerns may be justified. What is more, Halliday (1993b) suggests that native-Chinese speakers may struggle to comprehend the types of complex anaphoric noun phrases which Pretorius (2005) found were (*i*) hardest for second-language readers to resolve and (*ii*) most closely linked to academic success at the university level. Specifically, in Pretorius' study, complex coreferential noun phrases, which serve as the kind of retrospective labels Francis (1994) identified as serving a vital cohesive function in academic tests, proved the most difficult to resolve, despite containing a determiner which explicitly indicated they were anaphoric in nature. Given that the use of determiners in this way is especially common in academic writing (Gray, 2010), the processing of such explicit coreferential noun phrases is arguably an area of second-language reading research that is deserving of more attention.

Furthermore, while a great deal of scholarship exists which has explored the various components which underpin second-language reading comprehension, the types of discourse-level representations they may result in, as well as how explicit cohesive devices may contribute to the process, studies have typically used quantitative analysis alone to establish what relationships may exist between these various factors. However, to date few studies exist which shine a light on *why* these relationships may exist, by qualitatively investigating how second-language readers interact with explicit coreferential noun phrases as they seek to establish discourse-level mental representation of academic texts. Thus, when investigating second-language reading comprehension, a need arguably exists to combine quantitative data used to identify differences

between readers at different levels of proficiency, with qualitative data analysis designed to investigate how L2 readers of academic English consciously interact with the texts they read, and what they understand those texts to mean. Building on suggestions by Walter (2007) and Ushiro et al (2017, 2020) that a reader's ability to establish discourse-level mental representations of a text may be tested by proxy by investigating their ability to identify incongruent words in a text, the following thesis reports on a mixed method study which sought to shed light on the following four questions:

RQ1. To what extent are native-Chinese speaking readers of English able to identify words in coreferential noun phrases which are incongruent with an idea expressed in its antecedent?

RQ2. To what extent does the presence of a determiner in a coreferential noun phrase help native-Chinese speaking readers of English identify words in that noun phrase which are incongruent with an idea expressed in its antecedent?

RQ3. To what extent do native-Chinese speaking readers at different levels of reading proficiency differ (i) in their ability to identify incongruent words in coreferential noun phrases, or (ii) in the extent to which they are helped to identify incongruent words in such a noun phrase by the presence of a determiner?

RQ4. How might any differences in the ability of students of different proficiency levels to identify incongruent words be explained by a qualitative analysis of their reading behaviour?

Chapter 3: Methodology

3.1 Introduction - Methodology

The following sections firstly describe the research design that was used to address the four questions posed at the end of Chapter 2, including a description of its theoretical and epistemological basis, the participants, the research instruments used, as well as the data gathering procedures for both the eye-tracking experiment and the think-aloud interviews.

3.2 The research design: A mixed methods approach

For ease of reading, much of the detail about the eye-tracking experiment and the think-aloud interviews are given in their respective results chapters (Chapters 4 and 5). The following section provides an overview of the research design with a view to explaining its foundations, the methodologies used, the epistemology that the study rested upon, as well as the materials and research instruments employed. One of the key features of the study reported in this thesis is the adoption of a mixed methods design. The use of mixed methods as a 'third paradigm' has grown evermore prevalent in the social sciences in general, and the field of education in particular, as researchers have moved away from the paradigm debates which marked much of the 1980s and 1990s (Creswell & Tashakkori, 2007; Croninger & Valli, 2009). Greene et al. (1989, p. 256), with admirable simplicity, defined mixed method research designs as:

> those that include at least one quantitative method (designed to collect numbers) and one qualitative method (designed to collect words), where neither type of method is inherently linked to any particular inquiry paradigm.

While this is the definition which has been adopted in the current study, it is important to note that it is not without controversy. Indeed, from the outset, the use of mixed methods designs has been met with caution, concern and even outright rejection from a variety of quarters and for a variety of reasons (Mark, 2015). The following section will address some of the concerns, as well as provide both theoretical and philosophical justifications for the use of a mixed method design in the current thesis.

3.2.1 Mixed methodologies or multiple methods – the epistemology of pragmatism

In an article intended to herald the arrival of mixed methods as the soon to be dominant approach to research in the field of education, Johnson and Onwuegbuzie (2004, p. 15) described their use as a "third research paradigm [which can] bridge the schism between quantitative and qualitative research". In so doing, they were attempting to counter what has come to be known as the Incompatibility Thesis. This was the argument, advanced by such paradigmatic purists as Guba (cited in Johnson and Onwuegbuzie, 2004, p. 14) who sought to reject the very possibility of mixed methods research, arguing that "accommodation between paradigms is impossible ... we are led to vastly diverse, disparate, and totally antithetical ends". The basis of this decidedly extremist view lay in what researchers such as Guba felt to be the incompatible world views which served as the philosophical foundations for different research traditions. Historically, researchers in the Social Sciences have aligned themselves along a number of axes, adopting positions that some view as representing polar opposites: *quantitative* versus *qualitative*, *subjectivist* versus *objectivist*, *positivist* versus *constructionist*. Those who, for convenience, may be referred to as the Positivists, although who in recent times are perhaps more likely to self-identify as Post-Positivists (Mark, 2015), adopt an objectivist world view, arguing social reality exists as an (almost) human-independent unitary entity, which can be explored and quantified using experimental techniques "in order to discover and confirm a set of probabilistic causal laws that can be used to predict general patterns of human activity" (Neuman, 2014, p.97). In contrast, the Constructionists argue that " there is no 'objective' social reality 'out there'" (Hesse-Biber, 2018, p. 547), and instead adopted "a more subjective, culture-bound, and emancipatory approach to studying individual behaviours and social phenomena" (Tashakkori & Teddlie, 2003). In advancing the use of mixed methods research designs, advocates such as Greene et al. (1989) and Johnson and Onwuegbuzie (2004) urge educational researchers to move away from viewing quantitative and qualitative approaches to research as existing in such binary opposition to each other. Rather, they argue that researchers should learn to view qualitative and quantitative research techniques as options that are both available to the researcher,

depending on the type of social issues they are seeking to explore, and which may be combined to in order to build on the relative strengths of each approach in order to compensate for their respective weaknesses.

However, although the design adopted for the current thesis conforms to what Johnson and Onwuegbuzie (2004) would recognise as mixed methods, the extent to which the qualitative analysis, that is to say that part which deals with non-numeric data, would be recognised as such by researchers who are strongly grounded in the traditions of qualitative research is questionable. Rather, it may be more accurate to describe the design used here as what Mark (2015, p. 21) refers to as a "multi-methods". Explaining this concept, Mark argues that while mixed methods research was initially considered to be an approach which, by definition, included both a quantitative and a qualitative element, this need not necessarily be the case. Rather, he argues that researchers need only be guided in the choice of methods by what is most likely to provide them with the data they need to answer they questions they ask. Consequently, while it may be that a mixture of quantitative and qualitative techniques is the most appropriate design, under different circumstances a researcher may choose to mix two or more quantitative techniques, or alternatively two or more qualitative ones. In the current study, this is indeed what happened, with one part of the data being gathered using a strictly quantitative approach, eye-tracking, and another from the use of think-aloud interviews (see Section 3.3 for details), which were subjected to both quantitative *and* qualitative analysis.

The reason that what is termed qualitative research in this study might not be accepted as such in some quarters arises less from the methods employed, as the ends that research intends to serve. In explaining the ends of much qualitative research, Hesse-Biber (2018, p.547) states it aims to investigate "how social life is reproduced and privileged by those who occupy positions of power, often in the name of liberating and exposing social injustice", and is frequently conceptualised through a Marxist-emancipatory, Feminist or in some sense overtly politicised perspective. To this end, the argument that there is no reality beyond that which is socially constructed has been intended to counter a perceived historical trend in which the social structures that reflect the interests and attitudes of certain groups, often characterised as *rich white men,*

are simply accepted as objective reality. In terms of the current study, while I do not take issue with this view, or with the aims it seeks to achieve, it is important to note that this is not my purpose here.

Rather, the aims of this thesis are more closely aligned with the thoughts of John Dewey, one of the leading proponents of pragmatism as an epistemology suited for scientific inquiry (Russel, 1946). That Dewey was in much of his academic and professional life focused on education, an inherently practical field of inquiry, may have been instrumental in shaping his view of the proper ends of science, which, in the elegant prose of his age, he described as follows:

> The office of physical science is to discover those properties and relations of things in virtue of which they are capable of being used as **instrumentalities**; physical science makes claim to disclose not the inner nature of things but only those connections of things with one another that determine outcomes and hence can **be used as means** (Dewey, 1929, p. v - emphasis added)

While it may now be inaccurate to say that the modern physical sciences do not attempt to investigate the "inner nature of things", in the field of education it is perhaps less controversial to state that the proper ends of academic inquiry ought to be directed towards improving practice and learning. Much of the discussion in the current thesis revolves around issues relating to cognition and comprehension and consequently makes frequent use of terms such as 'activation' and 'representation' to describe what is alive and most prominent in a reader's mind at a given time. In this regard, it follows the terminology of psycholinguistic researchers such as Gernsbacher (1997) and Kintsch (1998) who both describe reading comprehension as arising from the activation of memory 'nodes' which may be 'integrated' to form 'cognitive structures'. However, while these terms have become common currency, from a strictly ontological sense, it is not clear, at least to me, what their exact status, what perhaps Dewey would call their "inner nature", might be. Nor, for that matter, do I hold it to be of consequence, in so far as the purposes of the current research is concerned. A contemporary critic of Dewey's pragmatism, Bertram Russel (1946, p. 731) characterised Dewey's epistemology as the view that:

> whether a belief is good or bad depends on whether the activities which it inspires in the organism entertaining the belief have consequences which are satisfactory or unsatisfactory to it. Thus, a belief about some event in the past is to be classified as 'good' or 'bad', not according to whether the event really took place, but according to the future effects of the belief

Criticising this position, Russel imagined such an argument being applied to judgements about whether statements of historical fact are true, for example regarding the dates on which figures such as Columbus or Julius Caesar performed certain deeds, before concluding that such an approach would be absurd. Rather, Russel argued that in deciding whether such 'facts' are true or not, one should be guided only by the *causes* that lead one to the belief. For example, a cause of the belief that Columbus sailed to the New World in 1492 might be that there is a great deal of documentary evidence that such an event happened in that year. Even if it some positive effect could be achieved by believing it was true that Columbus sailed to the New World in 1776, this would not alter the veracity of the original belief that he did so in 1492. Seen in this way Russel position is self-evidently true, and Dewey's bordering on absurd.

However, while the argument that a proposition may be taken as true to the extent that it leads to a positive outcome may not hold up to scrutiny when evaluating the veracity of claims about the past, as indicated by the quotation above, Dewey's epistemology was not intended primarily as a means of judging the truth value of any given statement, but rather as a means of judging the extent to which the fruits of a scientific investigation can be used in practice. Applied to the current research, such a view might contend that even if the exact ontology of the terms of the debate might remain somewhat mysterious, or even if the extent to which eye-tracking or verbal reports allow researchers a view into what is 'really' happening in a reader's head remains in part unclear, this need not affect the validity of the conclusions reached. If the data which is gathered in this way allow educational researchers to (*i*) develop an understanding of how readers process text in order to establish comprehension, and (*ii*) use that understanding to develop teaching methodologies which can be shown to have beneficial outcomes when used in a classroom setting, then it is arguably the case that those understandings can be said to be 'true' to the extent that they need to be. By

way of analogy, although Newtonian conceptions of gravity have been superseded by Einsteinian ones, for the purposes of the vast majority or terrestrial engineering projects, Newton's laws may still be taken as true in so far as they allow for reliable predictions to be made, and thus for practical decisions to be taken, in the course of an engineer's work (Cox & Forshaw, 2010). Similarly, while I do not hold myself qualified to judge whether the key psycholinguistic concepts which underpin this research accurately reflect the 'reality' of cognitive processing, it is my hope that the conclusions reached may be judged by the extent to which they contribute to the development of evidence-based teaching practice that can bring measurable benefit to the students whose interests educational research ought ultimately to serve.

3.2.2 Strengths and weaknesses of mixed methods designs

Although, as discussed above, the design of this study may be considered to have adopted a multi-methods rather than strictly speaking a mixed method approach, given the prevalence of this latter term literature it is the one used throughout this thesis. As with all approaches to research, the use of a mixed methods design brings both advantages and disadvantages. In terms of the strengths, the main advantage is that it allows the researcher to draw upon the merits of the various methods employed, whilst also putting procedures in place to overcome their weaknesses (Johnson & Onwuegbuzie, 2004). However, as Mutch (2009) counsels, when designing any mixed methods project, it is important to stay abreast of all the relevant theoretical assumptions and practical issues which relate to all of the research methods involved. Given the degree of specialisation which quantitative and qualitative researchers build up in their respective fields, this is not a trivial issue to overcome (Creswell & Plano Clark, 2018). Even where researchers adopted a position of what Mark (2015, p. 23) refers to as "paradigmatic agnosticism", they nonetheless need to develop a good understanding of why the methods they chose are thought to reliably return the types of data they do, as well as how that data should be analysed. Consequently, mixed method designs are relatively labour-intensive in the development phase, as well as in the implementation.

3.3 The mix of methods used in this study: The overall design

Although there have been several attempts to establish frameworks to guide the design of mixed methods research (e.g. Johnson & Onwuegbuzie, 2004; Yanchar & Williams, 2006), the one deemed most appropriate for the avowedly pragmatist aims of the current thesis is that of Greene et al. (1989), primarily due to the lack of paradigmatic orthodoxy it demands for a definition of qualitative research. The framework they developed resulted from a meta-review of educational research papers which investigates how mixed methods designs had been used up to that date. In this way, Greene et al. identified five major types of mixed method designs, although of these only two were used in the current thesis; namely, *triangulation* and *complementarity*. Both of these techniques seek to draw benefit for what is perhaps the chief benefit of a mixed methods design, i.e. that they allow for the strengths of both methods to be drawn upon, while also allowing the research to account for the limitations inherent to each. Specifically, triangulation allows the researcher to gather multiple types of data on a single phenomenon, thereby improving the validity of both (Mark, 2015). Complementarity in a mixed methods design involves the use of quantitative and qualitative techniques to measure "overlapping but also different facets of a phenomenon, yielding and enriched, elaborated understanding of that phenomenon" (Greene et al., 1989, p. 258).

In terms of its design, the current study is composed of two key elements, an eye-tracking experiment and subsequent think-aloud interviews. As will be explained, the combination of these two methods allowed not only for the data obtained from the eye-tracking experiment to be triangulated by implementing a quantitative analysis of the transcripts from the think-aloud protocols, but also for those data to be explored in more detail using a more strictly qualitative approach. As will also be explained, such triangulation and elaboration are needed to overcome the limitations inherent to these two methodologies. The following two sections provide a detailed description of both eye-tracking and think-aloud interviews, in terms of the assumptions they rest on, their inherent strengths and limitations, as well as some practical issues involved

in their implementation. After this, details will be given about how these two methodologies were used to support each other.

3.3.1 Eye-tracking

Eye-tracking research in reading depends on two key assumptions (Conklin & Pellicer-Sánchez, 2016). The first, known as the eye-mind hypothesis (Just & Carpenter, 1980), states that the word currently being fixated by the eye is the one being processed. In its earliest, strongest form, the eye-mind hypothesis claimed that:

> the eye remains fixated on a word as long as the word is being processed. So the time it takes to process a newly fixated word is directly indicated by the gaze duration (Just & Carpenter, 1980, p. 330)

However, subsequent research has resulted in a degree of nuance being added. In a now classic review of much of the early work on eye movements in reading, Rayner (1998, p. 377) cites evidence which demonstrated that although most processing is completed while a word is being directly fixated, in some cases *spillover* effects occur, in which "time processing a word can "spillover" onto the next word". In addition, it has been found that the word directly to the right of the word being fixated can, in some cases, be subjected to pre-processing when the first three to four letters of the word fall within *parafoveal* vision, that is to say the area just outside the eye's centre of focus (Rayner, 1998). This can result in a word being processed 30 to 40 milliseconds faster than words which were not pre-processed (Rayner, 1975). There is also evidence to suggest that words which the reader did not fixate are also processed, for example, when they are extremely predictable within the context of the article (Staub & Rayner, 2007). However, all this notwithstanding, the eye-mind hypothesis has stood up well to scrutiny, such that as researchers, we can feel confident that "it is generally the case that how long readers look at a word is a fairly good reflection of the processing time associated with that word" (Staub & Rayner, 2007, p. 329).

The second assumption is that "the amount of time spent fixating an item reflects the cognitive effort required to process it" (Conklin & Pellicer-Sánchez, 2016, p. 454), so that longer fixations are taken to indicate that a reader experienced relatively more difficulty processing a word, while shorter fixations indicate that they processed a word with relative ease. This assumption rests on a body of research which has demonstrated that when a text is manipulated so that certain features should be more difficult to process, this reliably leads to increased fixations durations. For example, ambiguous words, such as homographs, are fixated longer than non-ambiguous words (Duffy et al., 1988; Sheridan & Reingold, 2012), while pronouns with an uncertain antecedent are fixated longer than those for which the referent is easily identifiable (Erlich & Rayner, 1983).

One of the primary advantages of eye-tracking over other research tools used to investigate online reading comprehension, such as self-paced reading, is that it allows for the collection of extremely rich data in a relatively naturalistic setting (Frenck-Mestre, 2005). While a degree of controversy exists in the literature as to what exactly the best eye movement measurements to collect are, and indeed what those measurements reveal about reading (Rayner, 1998), there are now clear recommendations as to how eye-tracking data should be categorised (Rayner et al., 2006). The primary distinction to be made is between *fixations*, which is to say when the eye is (more or less) still, and *saccades*, which is to say when the eye is in motion between one fixation and the next. While research has shown that it is only during a fixation that the eye is able to extract new information from a text (Rayner, 1998), there is also evidence to suggest that processing, for example lexical access processing of a word already fixated can still occur as they eye moves (Irwin, 1998). Saccades come in two main types, *progressive* and *regressive*, i.e. either to the right or the left respectively of the word being focused (at least in English and other left-right orthographies). Research has consistently shown that regressive saccades account for approximately 10% to 15% of eye movements (Conklin & Pellicer-Sánchez, 2016; Rayner et al., 2006), and while there are a number of reasons why they occur, including small corrections resulting from a progressive saccade overshooting the mark, regressive

eye-movements are assumed to indicate that a reader experienced comprehension difficulty of some kind, and so returned to read a word a second (or more) times (Rayner, 1998).

Although saccades inform some of the eye-tracking measures which can be used as dependent variables, for example the ratio of progressive to regressive saccades in respect of a target word, the most common types of measurements are calculated from fixations made on the text (Rayner et al, 2006). In an article providing an overview of the how eye-tracking can be used to investigate various facets of the reading process, Staub and Rayner (2007) state that the fixation measures used in eye-tracking experiments can be divided into two broad categories, *early* measures and *later* measures The early measures, which include *first fixation duration* and *gaze duration*, if the region of interest contains only one word, or *first pass time* if the region of interest is longer (e.g. an idiom), are argued to reflect initial, word access processing. First fixation duration records the time the eye spends on a word from first entering the word from the left, before any movement to the left or right. Gaze duration includes eye movements made *within* a word during the first visit, before the eye leaves the word to the left or right. Hence, gaze duration is typically, although not solely, used to measure processing of longer words which cannot be processed within one fixation (e.g. *supercalifragilisticexpialidocious*). By contrast, the late measures are argued to reflect integration processing, in which the reader integrates the meaning of the word just processed into their developing discourse-level representation of the overall meaning of the text. The late measures include *second pass reading time* (i.e. time spent rereading a word having already fixated it during first pass), *total reading time* (calculated by combining first and subsequent reading times), as well as *regression path duration,* sometimes known as *go past time*, (defined as the time that elapses when a reader's eye moves to the left of a word they have already fixated, before they eye moves to the right of the word for the first time). Although Rayner (1998) advised collecting data from a variety of measures on any given region of interest, arguing that all eye-tracking measures require the researcher to make "unjustifiable simplifying assumptions about the relationship between the measure being used... and cognitive processing

during reading" (p. 378), in recent years concerns have been raised that using too many different measures may lead to inflated risks of Type 1 statistical errors resulting from the need for multiple comparisons to be made (von der Malsburg & Angele, 2017). Consequently, where possible it is advisable for researchers either to specify whether they expect a predicted effect to be found in either an early or late measure, if not in a specific eye movement measure. Where such accurate prediction cannot be made Bonferroni corrections, or some similar procedure, should be applied (von der Malsburg & Angele, 2017).

Although eye-tracking as a methodology for investigating online reading has been shown to provide rich, valid, consistent data (Carter & Luke, 2018; Godfroid & Spino, 2015; Staub & Rayner, 2007), it is important to note that it also has a number of inherent limitations which must be borne in mind and accounted for. Firstly, when designing eye-tracking studies, it is important to control carefully for outside variables which can affect processing time of the target word or region of interest. Factors which have been shown to have a significant effect on processing times include word frequency, word familiarity and predictability due to such issues as frequent collocational pairing giving rise to priming effects (Staub & Rayner, 2007), for example where the phrase *a breakfast of bacon and...* would in Anglophone readers prime for the word *eggs*. In addition, where a word appears in a sentence can also affect processing times, with sentence final words typically being fixated longer due to what is referred to as the *sentence wrap-up effect* (Rayner, 1998), which is believed to indicate integration processing. Similarly, words at the beginning of sentences, or at the end of a line of text are more likely to be skipped (Frenck-Mestre, 2005). Hence, if eye-tracking experiments are to result in reliable conclusions, it is important for researchers to control carefully for all possible outside variables in order to ensure that any effects are observed are attributable to the experimental manipulation rather than an extraneous variable. Secondly, as Miller (2015) points out, during the course of reading people's minds naturally wander. As such, while it may be the case that increased fixations times reflect processing difficulty, they may simply reflect a temporary loss of attention, or the fact that the word read reminded them of a related idea which distracted their attention. This suggestion

that loss of attention can show up in the eye movement record in longer fixations was supported by a study by Reichle et al. (2010) which investigate what they called Mindless Reading, i.e. when the mind had wandered. The participants in the study read the entirety of Jane Austen's *Sense and Sensibility*, over the course of several days, and were asked to press the *Z* key if the found their attention had wandered. They were also prompted every four to five minutes, and asked to press the *Y* key if their attention was focused on the text at the time of the prompt, or the *N* key if their mind had wandered. The results of the study showed that for the first-fixation duration, gaze duration and total reading time measures the reading times were longer when then participants' attention had wandered versus when they were engaged in normal reading.

It is important to note that the vast majority of eye-reading studies do not require the participants to read texts even vaguely approaching the length of a Victorian novel, and so while the suspicion that the participants' attention may wander during the process should not be dismissed, it is perhaps unlikely to constitute a major issue in most studies. However, as Frenke-Mestre (2005, p. 178 – emphasis added) points out, all eye-tracking measures "can only be used as an *indication* of processing difficulty when compared to an equivalent, experimentally controlled alternative condition [such as Event Related Potentials]". Although there is a substantial amount of evidence in the literature to support the various assumptions which form the theoretical bedrock of eye-tacking research (Miller, 2015), in themselves processing times do not constitute absolute proof of specific cognitive processing having occurred. Finally, and perhaps of most relevance to the current study, although eye-tracking data can be used to demonstrate *that* processing difficulties may have occurred, or that specific differences exist between how participants in different groups behave differently, they cannot tell us *why* they occurred, beyond the tightly controlled parameters of the experimental study, or what a reader was actually thinking as they processed a text. To gather data of this kind, it is necessary to turn to a different research method

altogether. As will be argued below, the most suitable for the task are introspective verbal reports, and in particular the use of think-aloud protocols.

Finally, although eye-tracking has been argued to provide a relatively naturalistic view of the process of reading (Frenck-Mestre, 2005; Rayner, 1998), as Godfroid and Spino (2015, p. 900) point out, "participants in an eye-tracking experiment are hardly reading in their easy chair." The process of gathering data during eye-tracking sessions requires participants to sit for extended periods of time positioned with their head in a chin rest, staring at a computer screen, in a way that is often reported as being quite fatiguing. What is more, the requirement to frequently recalibrate the eye-tracking equipment, a sometimes lengthy, repetitive and frustrating process, can also result in the experience being yet more tiring still. Consequently, there is a risk that data gathered in this way may not in fact reflect a participant's reading behaviours entirely naturalistically. Furthermore, as Godfroid and Spino (2015) also make plain, the overtly technical set up of any eye-tracking session also makes it abundantly obvious to the participant that they are involved in an experimental procedure focused on their eye movements. While efforts can be made to mitigate against these effects, for example by providing breaks in between data gathering so allowing participants to rest their eyes, or giving instructions to read in their usual way, the possibility exists that the eye movement data gathered in this way might not provide an entirely accurate representation of the participants' typical reading patterns.

3.3.2 Think-aloud protocols

One of the most common research methods currently used in the field of both Second-Language Acquisition in general, and second-language reading comprehension in particular, are verbal reports (Leow & Morgan-Short, 2004). Verbal reports come in two major forms; concurrent verbal reports, also known as think-aloud protocols, and retrospective verbal reports, also known as stimulated-recall. However, both think-aloud and stimulated recall depend on the assumption that a reader's cognitive processes are available for introspection with a sufficient degree of clarity and saliency for them to be able to report them

to the researcher (Fox et al., 2011). While verbal reports have been used for a long time in the cognitive sciences to investigate a wide variety of complex cognitive operations, ranging from playing to chess, to gambling, to solving number and picture puzzles for well over one hundred years (Leow & Morgan-Short, 2004), from the outset doubts have been raised as to how reliable a tool they are (Kucan & Beck, 1997). Given the degree of controversy the use of verbal reports has been subject to, and the attendant number of caveats and frameworks that have been generated to guide their successful implementation, the following section will provide a review of the concerns that have been raised, solutions that have been proposed, and steps taken to ensure that the research reported on in this thesis was conducted in such a way as to ensure the validity of the results generated.

A number of possible issues have been raised in respect of how valid the data obtained by think-aloud protocols might be, including such issues as the extent to which tasks directions might influence the participant's behaviour, whether the interviewer might, by their very presence, alter the behaviour they are seeking to measure or whether participants might be unduly distracted by the sound of their own voice (Stratman & Hamp-Lyons, 1994). One of most important issues which has been raised in respect to the use of think-aloud interviews in the second language is what has come to be known as reactivity; that is, the extent to which the process of giving verbal reports alters the cognitive processes that are being investigated to the extent that they no longer reflect the behaviour under question (Leow & Morgan-Short, 2004; Raudszus et al., 2019).

The concerns regarding the reactivity of think-aloud interviews have been expressed in relation to a number of separate issues. One that has received particular attention is between what Ericsson & Simon (1980), one of the principal research teams behind the modern use of think-aloud protocols, refer to as *metalinguistic* and *non-metalinguistic* interviews. In non-metalinguistic interviews, participants are asked simply to verbalise their thoughts as they arise, without attempting to explain them in anyway. In metalinguistic interviews, on the other hand, the participant is encouraged to explain why they had the

thought, or in some way comment on their own thought processes. When they first proposed the use of think-aloud interviews as a viable mean for collecting data on cognitive processes, Ericsson and Simon (1980) argued that while non-metalinguistic interview techniques would not be reactive with the underlying thought processes, metalinguistic techniques would add an additional layer of cognitive activity which would interfere with processes being measured. Subsequent research has broadly shown this to be correct (Bowles, 2010; Hu & Gao, 2017). However, it is important to note that the interactions are not necessarily negative, to the extent that the use of metalinguistic think-aloud interviews has frequently been shown to *improve* task performance in certain areas (Fox et al., 2011; Wade, 1990), and have now even be recommended as an instructional technique for teachers of reading (Kucan & Beck, 1997).

In order to investigate these issues, a number of studies have been carried out to investigate the validity of think-aloud protocols for use in research into Second Language Acquisition in general, and second-language reading in particular. For example, Leow and Morgan-Short (2004) set out to investigate the reactivity of think-aloud interviews in relation to second-language learners' reading comprehension, ability to learn new verb forms, and written production. They recruited 78 native-English speaking first-year university students who were enrolled on a Spanish course. Of these, 38 were assigned to a group that engaged in a think-aloud task, while 39 were placed in a control group, who engaged in silent reading. The participants in the think-aloud group were only requested to verbalise their thoughts as they arose, without being asked to comment on them, or to focus on any particular aspect of the language of the text. The participants read a short story in Spanish which was 384 words in length. The text contained 17 target verbs, in the Spanish impersonal imperative. Three weeks prior to the experiment itself, the participants were given a pre-test of the target form, composed of a recognition task and a controlled writing task. In addition, two days before the experiment took place, the participants were given a vocabulary matching worksheet, designed to familiarise them with the content words in the short story they would read. After reading the text on the day of the experimental procedure, the participants were presented with 11

comprehension questions written in English, which took the form of both multiple-choice and short answer questions. In addition, they also completed a 17-item test designed to assess how well they had learned the target form. Finally, they also completed a fill-in-the-blank task to assess their ability to reproduce the target form in writing. Using a *t*-test to compare the participants' scores, Leow and Morgan-short found that no significant differences existed between participants in the think-aloud group and those in the silent-reading group in terms of their comprehension of the article, their subsequent recognition of the target verb form, or their ability to reproduce it in writing. From these results, Leow and Morgan-short (2004, p. 50) concluded that:

> Thinking aloud while performing an L2 reading task of 384 words did not appear to have detrimental or facilitative effects on comprehension, intake, or controlled written production when compared to a non-think-aloud group performing the same task. In other words, thinking aloud was nonreactive in this study, and these results provide empirical support for the validity of the use of such procedures to gather concurrent data on learners' cognitive processes.

A second study which investigated think-aloud interviews as a suitable tool for second-language reading research was conducted by Godfroid and Spino (2015), who explicitly compared the reactivity of think-aloud interviews as compared with eye-tracking. For the study, 116 second- and third-year students at Belgian university who had either upper-intermediate (B2) or lower-advanced (C1) English proficiency, as measured by the Common European Framework of Reference (Council of Europe, 2001), were recruited. Of these, 28 participants were placed in an eye-tracking group, 28 in the think-aloud group, and 46 in a silent-reading control group. Similarly to Loew and Morgan-Short (2004), the participants in the think-aloud group were not asked to explain or justify their thoughts, but only to engage in non-metalinguistic verbalisations. The participants were asked to read 20 paragraphs taken from a variety of authentic sources, including newspaper and magazine article, and instructed to read them "in a leisurely fashion, as if they were reading a novel for pleasure" (p. 907). Of these, twelve paragraphs contained experimental items, and eight simply served as fillers. The experimental items were twelve target words, nine of which were pseudowords, such as *flarrisation*, and three were real words, used as a control. The post-test designed to measure the

participants' comprehension of the content of the paragraphs they read consisted of 20 statements, presented in random order, about which the participants had to decide if they were true or false. They were also given the option to select 'I don't know' in order to guard against guesswork, although this was scored as an incorrect answer. The statements were written in such a way that the pseudowords were not relevant to judging if they were true or false. In addition to this comprehension test, the participants were also given a surprise vocabulary test after they had completed the reading task. In this test, the participants were shown sentences drawn from the article, with one of the target words blanked out, and given 15 seconds to read it. After this, they turned the page and were given another 15 seconds to choose the right word to fill in the gap from a list of 18 possible options. When analysing the data from their study, Godfroid and Spino (2015) subjected them to both superiority and equivalence tests, arguing that "only the latter are conceptually appropriate for demonstrating nonreactivity" (p. 896), on the grounds that in null hypothesis testing, failure to reject the null hypothesis "does not provide statistical support that this hypothesis is true, only that the alternative cannot be accepted, [meaning that] any claims that think-aloud interviews are not reactive when a researcher fails to reject the null hypothesis are theoretically unfounded because the negative result may simply result from a lack of power" (p. 903). In order to circumvent this issue, Godfroid and Spino argue for the use of equivalence tests as secondary measure on the grounds that:

> compared to superiority tests, the logic of equivalence tests is inverted: Researchers seek to reject the null hypothesis that two groups are different. Rejecting the null hypothesis in favour of the alternative would therefore constitute theoretically sound evidence that a given methodology is indeed nonreactive. Thus, superiority tests seem to be the preferred statistical method to show that a given methodology is ***reactive***, but they are not sufficient to demonstrate nonreactivity. Conversely, equivalence tests are not ideal to show that a methodology is reactive, but they can provide theoretically grounded support for ***nonreactivity***.

Using this approach, the results from both sets of statistical analyses indicated that neither the think-aloud nor the eye-tracking procedures had any reactivity with the participant's ability to comprehend the texts read. With respect to the vocabulary learning task, the results from both analyses suggested that there was

a small positive effect from the think-aloud on that participants ability to remember the pseudowords. With respect to the eye-tracking however, the results diverged slightly, as the superiority testing analysis suggested that they eye-tracking was nonreactive with the participants' ability to remember the pseudowords, while the equivalence testing suggesting that there was small negative effect. The results of this study thus supported earlier studies (Bowles, 2010; Kucan & Beck, 1997) which have suggested that think-aloud interviews do not impact on participants' ability to comprehend texts, suggesting that they are a valid method for investigating the cognitive processes surrounding reading comprehension. However, they also suggest that use of think-aloud may react positively with the participants' ability to learn from texts. In the context of the current study, this last point is worth bearing in mind, in the light of some tentative findings from a pilot study reported in Stratman and Hamp-Lyons (1994, p. 102), who found that the use of think-aloud interviews " appeared to slightly enhance the detection of faulty pronoun references", although they also reported that it "appeared to make little difference in subjects' ability to detect phrase-level redundancies and word-level errors". Given that the research reported in this thesis investigated the participants' ability to identify incongruent words in a text, it is unclear to what to extent the use of think-aloud may positively enhance their ability to detect the incongruency as they read. Similarly, it is also not clear to what extent the use of eye-tracking may impede their ability to do so. As will be explained below, it was in part to address this issue that a mixed methods research design was used.

However, although the investigations presented above demonstrate that think-aloud interview can be used as a valid instrument to research a reader's cognitions as they process a text, it is important to also recognise that, as with all research methods, there are a number of important limitations. Firstly, as with all qualitative research techniques in which the researcher is engaged in a direct social interaction with the participant, think-aloud interviews constitute an active process in which the role of the researcher in shaping the participants' responses cannot be entirely mitigated against (Holstein & Gubrium, 1995). Although the aim of think-aloud protocols is to give the researcher a view into the participant's cognitive

processes as they read, the way the interview is framed and conducted will inevitably have some degree of impact on the very processes they see to measure. For this reason, in order to maintain consistency across participants, it is important for researchers to give clear and consistent instructions to all of the participants (Leow & Morgan-Short, 2004; White et al., 2007) [see Section 3.7 for the instructions given to participants in both the eye-tracking experiment and think-aloud study]. In a similar vein, as Mann (2011) points out, given that any interview situation creates its own context, one in which the researcher is to some extent the dominant participant, it is important for their voice to included in any transcripts in situations where they directly influenced the direction of the conversation. For this reason, in the analysis of the think-aloud interviews presented in Chapter 5 of this thesis, where participants were asked, and were thus responding to, a direct question from the interviewer over and above the general instructions they were all given, the question is included in the excerpts from the transcripts presented as evidence.

A second limitation inherent to the use of think-aloud interviews is that they necessarily create an inauthentic situation (White et al., 2007), to the extent that it is not usual for university students to voice their thoughts to an outside observer as they read. Consequently, it may be that they engage in reading strategies that they usually would not, and thus potentially establish mental representations of the text that they might otherwise not have done. As explained above, although it has been argued that engaging in think-aloud tasks does not impact on participants' ability to comprehend the texts they read (Godfroid & Spino, 2015), research has shown that they can have a beneficial effect on their ability to detect certain types of errors in a text (Stratman & Hamp-Lyons, 1994). Hence, in respect of the current study, it may be that the context of reading that was created by asking the participants to engage in a think-aloud interview may have resulted in them employing strategies they might not otherwise have done, or detected errors that might otherwise have gone unnoticed had they been reading in a more naturalistic setting. While this reality must be accepted, it cannot be mitigated against.

Similarly, while think-aloud interviews provide a window into the participant's conscious thought processes, this view is necessarily limited by the extent to which the participant is either willing or able to share *all* of their conscious thoughts, as well as by the fact that the line between conscious and unconscious thought is hard to be sure of as much thought is fleeting, and thus not necessarily available for conscious introspection (Dörnyei, 2007). Within the fields of both first and second-language research it has long been recognised that certain aspects of reading, such as word access, phonological processing or syntactic integration are automatised, and thus not open to introspection, while others, such as the use of metacognitive reading strategies are more purposeful, and thus salient to the reader (Grabe & Stoller, 2015). Thus, while think-aloud interviews are frequently used within second-language strategy research (Dörnyei, 2007), for example to investigate the use of reading comprehension strategies, they are less able to provide insight into the underlying automatised processes which may have motivated the use of those strategies in the first place. It is in large measure to address this inability of think-aloud interviews, or other forms of verbal report, to investigate both automatised and conscious processing that they are increasingly used in tandem with eye-tracking, so that researchers may gather data on different aspects of reading comprehension within the boundaries of a single study (Brunfaut & McCray, 2015; Godfroid & Spino, 2015).

Finally, one practical issue which relates specifically to the use of think-aloud interviews in second-language research is the relationship between the first and second language as mediums of expression, and the relationship between what might be termed the first and second culture in respect of how the participant approached the task they are presented with (White et al., 2007). The most obvious issue is that the requirement to verbalise their thoughts in a second language might add extra cognitive burden on the participants, especially those at a lower-level of language proficiency (Leow et al., 2014), with the consequence not only that this may affect their ability to conduct the task efficiently, but also that it is hard to know as researcher to what extent their ability to verbalise their thoughts was due to the task demands, or simply that they may not have sufficient command of the second language to express themselves clearly

and with nuance (Dörnyei, 2007). With a view to circumventing this problem, in the current project participants were told they could speak in either English, Chinese, or a mixture of both. In terms of conducting the interviews and analysing the data this did not cause difficulty as I am a fluent Chinese speaker, and thus comfortable in both languages. Nonetheless, as in all second-language research, the issue of cross-linguistic influence can never be fully discounted (Koda, 2005).

3.3.3 How the eye-tracking experiment and think-aloud interviews were combined in the current study.

Following Walter (2007) and Ushiro et al. (2017) [see Section 2.4], the research design used for the current study used an Anomaly Detection Paradigm to investigate the extent to which native-Chinese speakers are able to establish coherent, discourse-level mental representations of academic texts. As previously stated, the study was composed of two parts; an eye-tracking experiment followed by think-aloud interviews, a combination that is employed with increasing frequency in the second-language literature (Godfroid & Spino, 2015). The combination of these two methods was intended to build on the strengths, while off-setting the limitations inherent to each, by combining them in a mixed methods design which aimed at both triangulation and complementarity. In terms of the eye-tracking component, as Hessel et al. (2021) point out, any task which involves a secondary task over and above simply reading a text, such as thinking aloud, or even a relatively non-invasive task such as pushing a button when an anomaly is detected (Walter, 2007), "comes with additional metacognitive or linguistic demands. It is not clear whether such tasks capture comprehension monitoring as it happens during the course of reading itself" (p. 160). The use of eye-tracking is able to circumvent this, as a result of being a relatively non-invasive technique in which the participant is required to read the text, on the assumption that any "attempt to correct anomalous text, this should be reflected in the eye movement record" (Hessel et al., 2021, p. 160). This contention is supported by a number of first language eye-tracking studies (e.g. Braze et al., 2002; Rayner et al., 2004). To this end, in the current study, a number of incongruent words were placed in coreferential noun phrases embedded in four academic texts (see Section 3.5 for details of the target noun phrases and the texts they appeared in),

and data were gathered on the participants' eye movements as they processed them (see Section 4.3 for details of the specific eye-movement measures used in this study).

However, although eye-tracking can provide data at an extreme high degree of temporal resolution to *indicate* that a reader experienced processing difficulty when encountering an incongruent word in a text, as explained in Section 3.3.1, it cannot provide definite proof. Furthermore, while eye-tracking can provide a very strong indication *that* a reader experienced some kind of processing difficulty, it cannot illustrate *why* this happened. In order to overcome these inherent limitations, the think-aloud interviews were used to both triangulate and elaborate on the results of the eye-tracking experiment. Specifically, approximately one week after completing the eye-tracking phase of the study, a sub-sample of participants (see Section 3.4 for details of the participants) were invited to take part in a think-aloud interview in which they reread one of the articles they had read during the eye-tracking. The transcripts from the interview were then analysed quantitatively to discover whether or not they had been able to identify the incongruent words in the text (see Section 5.4.1 for details of this analysis) to see if this provided the same or different results from the eye-tracking experiment. Secondly, a thorough qualitative analysis was also conducted on the transcripts to examine the extent to which any difficulties the participants had reading the text may have influenced their ability to identify the incongruent words in the text (see Section 5.5 for details of this analysis). This secondary analysis thus sought to address the question of why any such differences occurred. However, in respect of the findings from the think-aloud interviews, it is important to introduce a note of caution. As discussed above, metalinguistic think-aloud protocols, of the sort used in the current study, have been shown not to impact on the participants' ability to comprehend a text. As such, contrary to fear that they may add to the cognitive load the participant may bear when attempting to comprehend a text (Raudszus et al., 2019), think-aloud interviews are a suitable method to investigate the types of mental representations readers are able to establish as a product of reading. However, there is also evidence to suggest that giving verbal reports during a task may lead to improved performance. During the think-aloud interviews

conducted for this study the participants were asked how they knew certain coreferential nouns were related to ideas stated earlier in the text. It cannot be discounted that their ability to perform this task was improved by the process of verbalising their thoughts as they did so (Bowles, 2010; Kucan & Beck, 1997).

3.4 The sampling procedure and the final sample

The participants who took part in this study were all either students or academic visitors at a university in the south of England. In order to recruit participants, posters were placed around the university advertising the study. By way of an incentive, participants were offered either £10 or ninety minutes of free English language support for participation in the eye-tracking experiment. As such, this study sampled opportunistically from the available population. On the one hand, it may be argued that the offer of free English language support, which was overwhelmingly the preferred option, may have led to a bias in the sample towards academically motivated students. On the other hand, the fact that this study sought to gather data on overseas students who were undertaking the arduous task of studying in a second language, suggests that the target population is likely to be academically motivated population to begin with. As such, it is likely that the sample gathered in this way accurately reflects the target population of native-Chinese speaking students enrolled in UK universities at least tolerably well. Altogether, this procedure led to the recruitments of 82 participants for the study. The majority were studying on pre-sessional English courses, undergraduate courses, or both taught and research postgraduate courses. However, in addition to these, there were also three students who were academic visitors who took part in the study. When recruiting participants, attention was paid to make sure that the sample was composed of participants at different levels of proficiency (see Table 3.1 for details); however, there was inevitably a smaller number of students with either the highest or lowest grades.

After the eye-tracking experiment, 24 participants were invited to take part in the think-aloud interviews, in exchange either for another £10 or another ninety minutes of English support. As far as was possible,

participants were selected for invitation in such a way as to ensure that readers from across the proficiency range were represented.

Table 3. 1: Participants by reading proficiency as measured by IELTS

IELTS Reading Score	Number of Participants
4.5	1
5	3
5.5	11
6	11
6.5	15
7	14
7.5	11
8	8
8.5	4
9	4

However, in practice, it was found that not many of them were willing to take part in the second part of the study, and so ultimately it was necessary to sample opportunistically. Full details of the sample used for the think-aloud interview is provided in Chapter 5 where the results from the study are discussed in detail.

3.5 Materials: The texts used

The reading materials to be used for this study were designed to be as close to authentic academic texts as possible. However, given the need to *i*) ensure they included certain linguistic features, and *ii*) be comprehensible to participants at a variety of levels of proficiency, it was not possible to use fully authentic materials. In the end, four texts were prepared, based on articles taken from an intermediate level academic reading course – **Making Connections 2** (McEntire & Williams, 2013). These articles were then adapted in order to include the regions of interest this study primarily focuses upon.

The texts contain between approximately 500 and 750 words each, and contained either three or five paragraphs. Three of the texts contained eight regions of interest each. The fourth contained twelve regions of interest. It was decided to only include one longer text in an attempt to maximise the number of regions of interest which could provide data for the study, whilst at the same time minimising the strain on the participants, as reading on an eye-tracker is quite tiring. Altogether, the texts contained thirty-six regions of

interest (see Table 3.2. for all regions of interest. See Appendix A for the full texts). Each region of interest was formed from a pair consisting of an antecedent and a later coreferential noun phrase. For example, in Text 1, the first antecedent (bold and underlined) appears in the line:

> Deaths from **cardiovascular disease (CVD), which includes heart attacks and high blood pressure**, first began to decrease in the 1960s in Western countries…

The subsequent **coreferential noun phrase** which refers back to it (bold and underlined) appears in the next sentence:

> …death rates from **THESE [heart/brain] diseases** fell by over 35 percent between 1980 and 1997

Each coreferential noun phrase contained a head noun (in the above example, *disease*) which was identical in all experimental conditions, and was the key target region for which the various fixation durations used as dependent variables were calculated. In addition, the head noun was modified by an adjective which was either **congruent** or **incongruent** with the semantic context established by the antecedent (in the above example *heart* is congruent, and *brain* is incongruent). Each participant saw only one of these adjectives in the text they read; that is, one version of the noun phrase. The third constituent of the noun phrase was a demonstrative or comparative pronoun, **this**, **these** or **such**, which were either present or absent. Thus, each target noun phrase could appear in one of four combinations:

Factor 1	**Factor 2**	**Key region of interest**
determiner	congruent word	TARGET NOUN
determiner	incongruent word	TARGET NOUN
no determiner	congruent word	TARGET NOUN
no determiner	incongruent word	TARGET NOUN

As mentioned in Section 3.3.1, two factors which have been shown to have a major effect on fixation durations in eye-tracking studies are the length and frequency of a word (Rayner, 1998). For this reason, it

was decided to make each of the 36 target head nouns in the coreferential noun phrases as similar to each other in terms of their length and frequency, with a view to reducing the number of external factors which could influence how the processing times recorded and so confound the results. Note that for each item, the head nouns themselves were identical across conditions. In the case of word length, this was relatively straightforward. The 36 target nouns had a mean length of 9.22 letters each, and a standard deviation of 1.38 suggesting they were very similar in this regard. However, in the case of word frequency it proved much harder to keep each of the head nouns similar to each other. This resulted from the need to choose contextually appropriate words for each of the texts. To investigate the frequency of each word, the BNC corpus was used (https://www.sketchengine.eu/british-national-corpus/). The mean frequency per million was 100.94. However, the standard deviation was 104.82, thus highlighting a large degree of variation. This variation proved unavoidable in the context of extended sections of academic discourse. For example, text 4 contained both the words patients (frequency per million 151.39) and symptoms (frequency per million 32.6). Since the topic of the article used for text 4 was health and wellness, it was not possible to avoid use of these words. Given that language naturally varies in terms of the frequency of words, it was hoped that these differences would not affect the validity of the results too seriously. Consequently, while this inability to control for the frequency of the head nouns as the dependent variable does constitute a limitation to the study, the fact that the study used identical target regions means that any impact from variation in word frequency is likely to be negligible.

Table 3. 2: List of target coreferential noun phrases and their antecedents per text

Text 1: Cardiovascular disease – A tale of good news and bad news.	
Antecedent	**Coreferential noun phrase**
cardiovascular disease (CVD), which includes heart attacks and high blood pressure	THESE [*heart*/*brain*] diseases
Similar decreases have occurred in other Western countries	THESE [*encouraging*/*discouraging*] reductions
heart attacks, high blood pressure and strokes	THESE [*dangerous*/*mild*] illnesses
heart attacks, high blood pressure and strokes	SUCH [*health*/*language*] problems

high-tech equipment, better drugs and more experienced surgeons	THESE [*improving*/*worsening*] techniques
a decrease in physical activity, an increase in smoking, and a change to a less healthy diet	THESE [*risky*/*harmless*] behaviours
CVD/Cardiovascular disease	THIS [*fatal*/*minor*] disease
The surgical treatment of heart disease	SUCH [*surgical*/*simplistic*] treatment

Text 2: Heart disease and changing attitudes

Antecedent	**Coreferential noun phrase**
heart diseases, including high blood pressure or heart attacks	SUCH [*health*/*musical*] difficulties
smoking tobacco, poor diet or lack of exercise	THESE [*negative*/*positive*] behaviours
many people now understand that there is a connection between heart disease and lifestyle	THIS [*growing*/*lack of*] awareness
access to doctors on whose advice they could rely	SUCH [*medical*/*military*] expertise
opinions are changing for the better about who is primarily responsible for maintaining and improving public health	THIS [*encouraging*/*annoying*] improvement
gentle exercise, like walking, or swimming	THESE [*moderate*/*extreme*] activities
the number of people dying from heart disease has fallen sharply	THESE [*improving*/*worsening*] outcomes
people have started to adopt more health-conscious living habits	THESE [*healthier*/*unhealthier*] lifestyles

Text 3: Medicine and genetic (DNA) research: promise and problems

Antecedent	**Anaphoric noun phrase**
research in this field has allowed scientists to identify the genes associated with some serious genetic diseases	THESE [*exciting*/*worrying*] advances
some serious genetic diseases	SUCH [*dangerous*/*pleasing*] disorders
identify the genes associated with some serious genetic diseases	THESE [*innovative*/ *ancient*] discoveries
some serious genetic disorders, such as Cystic Fibrosis	SUCH [*harmful*/*harmless*] conditions
severe loss of muscle control and almost total memory loss	THESE [*serious*/*minor*] symptoms
an increase in funding for research in this area	THIS [*growing*/*decreasing*] interest
new ways to treat many genetic illnesses	SUCH [*genetic*/*traditional*] techniques
Ten months later, to their parents' delight, the children had completely recovered	THESE [*encouraging*/*upsetting*] outcomes

Text 4: New approaches to health and wellness

Antecedent	**Anaphoric noun phrase**
his heart is weak, and he has extremely high blood pressure	THESE [*dangerous*/*valuable*] symptoms
Caroline Silva & Zhang Bao	THESE [*lucky*/*unlucky*] patients
In the coming years however, his situation is likely to change for the better	THIS [*welcome*/*unwelcome*] improvement
smaller, more powerful sensors; increases in cloud-based computer power and storage; and wireless internet	THESE [*advanced*/*ancient*] technologies
mHealth solutions	THESE [*amazing*/*useless*] devices
data about people's health	THIS [*medical*/*financial*] information
patients wear sensors on their bodies, for example on a watch or skin patch	THESE [*wearable*/*office*] machines
transmit data about their babies' heart rate and other health details continuously and wirelessly to a laboratory or doctor	THIS [*reliable*/*unreliable*] transmission
the medical services can respond immediately	SUCH [*immediate*/*delayed*] response
the sale of these mobile sensors	THESE [*powerful*/*terrible*] sensors
cardiovascular disease, which kills around three million people a year	THESE [*unnecessary*/*necessary*] fatalities
service which is difficult to provide to patients who live outside of large cities	THESE [*significant*/*insignificant*] challenges

Data in the eye-tracking study were only gathered on how the participants processed the head noun in each of the target coreferential noun phrases. The participants' ability to identify the antecedents was explored in the think-aloud interviews. Because of the time-consuming nature of conducting and transcribing think-aloud interviews as well as the detailed nature of the analysis, the participants only read one article (**Text 1**) during the interview.

3.6 The research instruments used to investigate individual differences

One of the foci of the current study was to explore the extent to which readers at different levels of proficiency differed in their ability to identify incongruent words in a text, or whether the presence of a determiner helped them to do so. As such, a number of measures of individual differences which have been shown to be predictive of reading comprehension outcomes (Jeon & Yamashita, 2014) were used to gather

data (see Section 2.3 for a discussion of individual differences). Specifically, this included a vocabulary test, a grammar test, and test of working memory capacity in both English and Chinese. It is important to note that vocabulary knowledge, grammar knowledge and working memory capacity are not the only factors which have been found to predict second-language reading comprehension ability (Bernhardt, 2010). However, constraints of time during the data collection sessions meant that it was not possible to test for other issues.

3.6.1 The reading comprehension measure

The participants' level of reading comprehension was measured by asking them for their latest IELTS reading score. Using this as a measure had a number of advantages and disadvantages. The main advantage was that using their pre-existing scores meant it was not necessary to conduct a specific test of reading comprehension during the data gathering session. For reasons of practicality, it was decided that data for the eye-tracking experiment and the measures of individual difference would all be gathered in one session, for fear that otherwise participants might drop out between sessions. It was felt that ninety minutes was the longest reasonable amount of time that participants could be expected to stay, meaning time was extremely short, and tests of reading comprehension take time to administer. A second advantage to using the IELTS reading scores is that the test used has been widely tested for validity (e.g. Bax, 2013; Schoepp, 2018; Weir et al., 2012), and so it was felt that it represents an adequately valid measure of reading proficiency. Thirdly, as IELTS scores are commonly used as the benchmark for deciding whether or not to offer non-native speakers a place at UK universities, it was hoped that using them as the measure of reading proficiency for the current study would allow for more realistic judgements to be made about how likely non-native students at different IELTS band scores are to be able to learn from the texts they are presented with as a part of their courses.

This notwithstanding, it is undoubtedly the case that using IELTS scores also brought a number of not inconsiderable disadvantages. Firstly, and perhaps most importantly, some of the participants' scores were considerably out of date. This was particularly the case for participants who were studying for PhDs having

previously completed a master's degree in the UK, as they were not required to provide an IELTS score as part of their most recent application process. Moreover, even for some of the undergraduate and master's students, the fact that they took part in the study in the summer term meant that they had already been studying in the UK for most of an academic year. Consequently, it is to be expected that their level of reading proficiency may have increased due to practice. As such, it is possible that for a substantial number of the participants their level of proficiency was higher than that indicated than their IELTS scores.

The second disadvantage to using IELTS scores as a measure of reading proficiency is that it is measured in categorical band scores, rather than as continuous data. While categorical data can be use in a multiple regression analysis, they are less powerful and less sensitive than the kind of continuous data that were used to measure the other individual differences. Consequently, for ease of analysis, the IELTS scores were only used to divide the students into three proficiency levels which were used as a factor in the factorial ANOVAs (see Section 4.7 for a full discussion). Despite these weaknesses, the time restrictions on the data gathering procedure meant that the use of IELTS score was the only practical option.

3.6.2 The language tests

The study gathered data on the participants' level of vocabulary knowledge and grammar knowledge. Measurements of vocabulary and grammar knowledge were included as predictors because research on L2 reading has consistently shown these two elements are important predictors of reading comprehension ability (Jeon & Yamashita, 2014; Zhang, 2012). Furthermore, language proficiency was one of the factors Pretorius (2005) identified as being related to successful anaphor resolution. As vocabulary and grammar knowledge are integral parts of language proficiency, it is important to gather data on these elements.

3.6.2.1 The vocabulary test

The vocabulary test used was Schmitt et al.'s (2001) Vocabulary Levels Test Version 2 (VLT2). The VLT2 is designed to test knowledge of vocabulary at different levels, stratified according to how frequently an item is used in English. Specifically, the levels are the 2000, 3000, 5000, and 10,000 word level, with for example

items in the 2000 level being drawn from what corpus studies have found to be among the 2000 most frequently used words in English (Nation, 2013). In addition, it tests knowledge of Coxhead's (2000) Academic Word List, which is a collection of 570 word-families argued to appear frequently enough in a variety of academic disciplines as to constitute a class of vocabulary in its own right (Nation, 2013). The VLT2 contains 150 test items and tests a learner's knowledge of the core semantic content of each word. To complete the test, participants are required to match a sample of words taken from each frequency level to a definition presented alongside them. They do this by writing a number which corresponds to each word next to the definition they feel is appropriate (see below for an example). In each case, there are six words presented, but only three definitions, in order to present participants with a foil. For example, in the sample test item below taken from Schmitt et al. (2001, p. 85), it can be seen that the participants are required to match the word *scheme* (5) to the definition *plan*, the word *option* (4) to the definition *choice*, and the word *integration* (3) to the definition *into joining something into a whole.*

1 debate
2 exposure **5 plan**
3 integration **4 choice**
4 option **3 joining something into a**
5 scheme **whole**
6 stability

Importantly, the definitions are always written using words from a lower frequency band than the word being tested (e.g. the definitions for words in the 5000-frequency level are from the 4000-frequency level and below). As such, the VLT2 sets out to test a learner's breadth of vocabulary knowledge.

It was decided that breadth of knowledge was more important to measure than depth of knowledge due to the high volumes of vocabulary knowledge which are needed for successful reading comprehension of academic texts (Schmitt et al., 2001). Furthermore, success in academic reading has been argued to depend on knowledge of a large body of what has come to be termed 'academic vocabulary' (Nation, 2013). Thus, the VLT2 test seems well suited to the needs of this project (see Appendix B for the full test). Although tests

do exist which seek to measure both breadth and depth of knowledge (e.g. Computer Adaptive Test of Size and Strength (CATSS) - Aviad-Levitzky et al., 2019), constraints of time during the data gathering session meant it was not possible to use such lengthy tests.

Constraints of time also meant that it was not possible to conduct validity and reliability testing for the Schmitt et al (2001) test. However, the test has already undergone rigorous validation in the paper from which it was taken. The VLT is designed to measure vocabulary knowledge of L2 students. It is predicated on the following research-based assumptions about vocabulary size (Nation, 2013):

- The most frequent 2000 words in English provide the bulk of lexical resources necessary for everyday communication
- The most frequent 3000 words in English provide the vocabulary level required to begin to read authentic text
- Knowledge of the most frequent 5000 words in English is sufficient to able to read authentic texts
- Knowledge of the most frequent 10000 words in English can be considered reflective of a 'wide vocabulary' and is reported to be thought necessary for university level study in L2

The VLT explores vocabulary knowledge at these different frequency levels by asking participants to discriminate between a choice of vocabulary items by matching them to descriptive sentences as explained above. Schmitt et al (2010) carried out a large scale study of over 800 L2 university students, from over 20 countries, who they asked to complete the VLT and checked back their findings with at least some of the students. They found that the VLT:

> Supplies a profile of vocabulary frequency levels which are highly scalable. Factor analysis suggests that the test is essentially unidimensional. Personal interviews indicate that examinees accept the test and that answers on the test do reflect underlying lexical knowledge (p. 79)

3.6.2.2 The grammar test

The test took the form of a twenty-two item grammaticality judgment test. Reliability testing demonstrated that the test was reliable (Cronbach's alpha = .837) The twenty-two test items were presented to the participants in randomised order. The participants were required to decide if each sentence was

grammatically correct or incorrect. Where they judged a sentence to be incorrect, they were further required to attempt to correct it, as in the example below:

The police tried to arrested the man, but he escaped.
arrest

(see Appendix C for full test)

The grammar test used was adapted from one used in DeKeyser (2000). It comprises three main sections: *verb tenses*, *cohesion*, and *word order*, and included a total of twenty-two test items.

The **verb tense** section was split between two categories: past simple and present/past perfect continuous. These two tenses were chosen to represent a 'basic' and an 'advanced' verb tense, as experience suggests this is the order in which the past simple and present/past perfect continuous tenses are typically presented in ESL courses. The verb tense section of the test contained nine test items in total, split between five test items for the past simple tense and four for the perfect continuous tense. There was a disparity between the number of test items as reliability testing conducted after the pilot test indicated that the value for Cronbach's alpha could be increased by removing one of the items in the perfect continuous tense section.

The **cohesion** section was split between two categories: *pronouns* and *articles*. These two categories were also chosen to represent basic and advanced knowledge, a decision based on my twelve years' teaching experience which suggests that the pronoun system is typically easier to learn than the article system for many Chinese learners. This section contained nine test items in total, split between five in the pronouns category and four in the articles category. Again, there was a disparity as reliability testing indicated that the value for Cronbach's alpha could be increased by removing one of the items in the articles section.

The **word order** section tested knowledge of syntactic construction and focused on the position of relative clauses. This was in order to include knowledge of at least one form of post-modification. The final word order section of the test was very short, containing only four test items. This was because in pilot testing five of the items were answered correctly in 100% of cases, meaning they had no power to discriminate. Furthermore, reliability testing also suggested Cronbach's alpha could be increased by the removal of one of the items in the advanced section. Once again, constraints of time meant it was not possible to write and pilot new items.

3.6.3 Working memory capacity test

Although the evidence is somewhat mixed, individual differences in working memory capacity (WMC) have been suggested to have an important relationship with both L2 sentence processing (H. Zhou et al., 2017) and inferential processing during discourse comprehension (Rai et al., 2011). The relationship between WMC and discourse comprehension may result from the role linguistic markers of coreference play as 'processing instructions' within a text, indicating which semantic elements need to stay active in memory (Ariel, 2001b; T. J. Sanders & Gernsbacher, 2004).

According to Juffs and Harrington (2011, p. 138) working memory, "involves both the storage and processing of information", with the latter being "responsible for directing attentional resources and the interplay between the storage domains and executive control". Given the complexity of discourse processing, it was decided to take a measure of both storage and processing capacity in working memory. To this end, a forward digit span task was used to test storage capacity (Juffs & Harrington, 2011) and a backwards digit span test was used to test processing capacity (Kormos & Sáfár, 2008). The WMC test was conducted in both English and Chinese to establish a picture of the participants' L1 and L2 WMC to see if these differentially related to their reading comprehension ability, their proficiency at resolving anaphors in text or detecting semantic anomalies. A digit span was used rather than a word span test because of concerns that word span tests may not be independent from measures of reading comprehension (Kintsch,

1998; Oakhill et al., 2005). The test that was used was composed of two sections, and was identical in the English and Chinese version, save for the language used (see Appendix D for a full copy of the test).

3.7 Procedures

The following section provides details of the data gathering procedures for the eye-tracking experiment and the think-aloud interviews.

3.7.1 Eye-tracking

On first arriving, participants were asked to hand in a signed copy of the consent form and information sheet they had been sent previously. If they did not have these, they were provided with fresh copies. Once this was completed, they were given a full explanation of how the session was to be structured. Once they were ready to begin, they completed the vocabulary test, the working memory capacity tests and the grammar tests, in that order. This was done as the vocabulary test was the longest, and it was hoped that doing this first while they were fresh would reduce the chance of them becoming fatigued and so allow for better data to be collected. The working memory capacity tests were delivered first in Chinese and then in English. Both parts of the test were delivered by me, a native-English speaker with a strong knowledge of Chinese. In order to ensure the participants would not be affected by not understanding my Chinese pronunciation, I first counted to ten in Chinese. All the participants reported that they could understand me without difficulty. Before each of the tests, the participants were given clear instructions, and completed a few practice items to make sure they had understood what they had to do.

3.7.1.1 The eye-tracking data collection

The eye-tracking data itself was designed using SR Research's Experiment Builder software. The data were collected using an Eye-Link Portable Duo eye-tracking machine. During the session, the participants were first given an explanation of what was expected of them. Next, they were asked to place their chin on the chin rest, and the equipment was placed so that their eyes were 45cm from the screen. When they were ready, they pressed a button to start the calibration procedure.

The procedure itself consisted of the participants reading four articles. Before each article, the eye-tracker was calibrated to locate their eye position on the screen, and the results then validated. The calibration and validation procedure are run automatically by the Experiment Builder software. After each article, the participants looked away from the eye-tracker and answered five simple comprehension questions about the text just read. These were done on paper as prolonged eye-tracking is uncomfortable, and it was felt best to allow them to rest the eyes between the articles. These questions did not form part of the analysis, but were used only to test if they in fact read and understood the articles, or merely scanned their eyes over them. Altogether the eye-tracking took around 30 minutes to complete per participant.

Before they began reading, the participants were given the following instructions orally:

> During this experiment you will see four articles. After each article, you will be given some simple comprehension questions to answer. However, these questions do not require you to remember any detailed information. They are only intended to check if you have read the articles and understood their main meaning. They are not like the questions you will have seen in an IELTS test. Please just read the articles in the same way you might read a text book when learning about a new subject.
>
> Three of the articles have three screens, and one of them has five screens. There is no real difference, one is just longer. At the end of each screen please press any key on the controller to get the next screen. Please be careful not to press the button until you are ready, as we cannot return to previous screens.

3.7.2 Think-aloud

As explained in Section 3.3.3, the think-aloud interviews were conducted with a subsample of the eye-tracking participants as a follow up to the eye-tracking experiment, in order to triangulate and elaborate on the results from that experiment. The think-aloud interviews consisted of two main parts; the think-aloud interview itself, and a secondary anaphor resolution test (see Section 5.2 for details). The process was conducted in four steps, as described below.

3.7.2.1 Instructions

The first step was to explain to the participants how the think-aloud procedure works. They were given the following instruction:

> I am interested in exploring how Chinese students read academic articles in English. For this task, I am going to ask you to read an article, out loud, and explain to me what you are thinking while you try to understand its contents. When you tell me what you are thinking, you may use either English or Chinese, or a mixture of both. Whatever feels most comfortable. While you are reading, I will not say anything, except to remind you to keep telling me your thoughts. However, if you have any questions please feel free to ask me at any time. I will now give you a demonstration of how to explain your thoughts while reading.

All participants were given the same instructions in order to standardise the proceedings as much as possible.

3.7.2.2 Demonstration

In Step 2, I demonstrated to the participants how to verbalise their thoughts while reading. In order not to prejudice their natural approach to reading a text in English, I read a text in Chinese to them, voicing my thought processes as I did so, with a particular focus on comprehension building, monitoring and regulation. For example, this involved modelling such strategies as guessing the meaning of words from context or identifying semantic links between sentences. This approach was trialled during the pilot study and found to be highly effective. A text in Chinese was used so as to avoid, as far as possible, biasing the participants' own reading by showing them which aspects of an English text it might be useful to focus on. Nonetheless, it cannot be ruled out that the demonstration in some way affected the participants' performance during their think-aloud interviews (see Section 3.3.2 for a discussion of the limitations of think-aloud interviews)

3.7.2.3 The think-aloud task

In step 3, the participants read **Text 1** from the eye-tracking study (see Appendix A for a full version of the text). This text was chosen as it was felt to be sufficiently challenging to elicit a response from the strongest participants whilst at the same time being within the capability of the weakest (Bowles, 2010). During this phase I did not speak, asides from to remind the participants to keep voicing their thoughts.

3.7.2.4 The anaphor resolution task

In step four of the interview, the participants were specifically asked a series of questions about the eight target noun phrases in the text they read:

What does this word (the main noun in the target noun phrase) refer to?	*used to test the participants' ability to integrate the noun phrase with its antecedent*
How do you know this?	*used to explore what knowledge resources the participant drew upon to integrate the noun phrase with its antecedent*
What does the word mean in Chinese?	*used to test if the participant had correctly understood the word*
What adjective did the writer use to describe this word?	*used to draw the participant's attention to the incongruent word*
Why do you think they chose this adjective?	*used to encourage the participant to focus on the meaning of the incongruent word*

This last step was conducted as think-aloud interviews have been argued to return the best data when readers experience difficulty processing the language of a text, as it is under these circumstances that their thoughts are most salient to them (White et al., 2007). As such, it was hoped that if the higher proficiency readers in the sample did not say enough to allow for analysis during the think-aloud interview, then it would be possible to gather secondary data during the anaphor resolution test. Furthermore, it was hoped that the use of a secondary *off-line* test would allow for comparison to be made about how the participants

processed the incongruent words in the target noun phrases when their attention was specifically draw to them, as opposed to when they were engaged in the cognitively taxing activity of online reading.

3.7.3 The qualitative think-aloud analysis – Elaborating on the results from the eye-tracking experiment

Part of the aim of the current thesis was to explore how second-language readers at different proficiency levels establish discourse-level representations of academic texts. To this end, the study used an Anomaly Detection Paradigm, predicated on the assumption that if a reader is able to integrate a coreferential noun phrase with its antecedent, they should be able to identify a word in that noun phrase which is incongruent with an idea expressed in the antecedent. In the eye-tracking experiment, the influence of proficiency was tested for statistically, to see firstly if any relationship could be found to exist between the vocabulary knowledge, grammar knowledge and working memory capacity, as these are believed to be key components of reading success. Secondly, the participants were grouped into three proficiency levels – *low*, *medium* and *high* – to see if difference could be found in the eye movement record of how they interacted with the incongruent words embedded in the text. However, as explained in Section 3.3.1, while the eye-tracking can provide data as to what happened as the participants processed the target noun phrases, this technology cannot tell us why it happened. For that, it was necessary to use think-aloud interviews, in order to explore the participants' moment-by-moment interaction with, and comprehension of, the text they read.

Much previous research has uncovered the comprehension strategies that second-language readers use as they process a text (e.g. Abbott, 2006; Anderson, 1991; Block, 1986; Ghavamnia et al., 2013; Tsai et al., 2010; Yapp et al., 2021). Following Fox (2009), the current study explored the relationship between the comprehension regulation strategies that the participants used and the subsequent mental representations they were able to establish as a result. It was initially intended to investigate the impact that regulation strategies, used either individually or in clusters had on the types of mental representations they were able to form as they read, in terms of Kintsch's (1998) Construction-Integration model's categories of proposition, textbase or situation model representations (see Section 2.4 for details of the Construction-Integration

model). However, while it was possible to do this by conducting a close line-by-line analysis of how the participants read every sentence of the article, in practice there was not enough data available in relation to the eight target coreferential noun phrases, which were the main focus of the analysis, for any reliable conclusions to be formed. To complicate things still further, as the primary focus of this project was to identify under what condition the participants were able to identify incongruent words embedded in a text, where such an ability is taken as a proxy for the ability to establish a discourse-level representation of that text, this meant that there were only four target noun phrases in the text used for the think-aloud interviews that could be used to analyse how the participants dealt with the incongruencies.

Consequently, the nature of the analysis changed. Rather than trying to examine how particular comprehension regulation strategies affected the participants' ability to establish different types of mental representations when they encountered an incongruent word, when a participant showed evidence of having used *any* comprehension regulation strategy, regardless of what it was, this was taken simply to indicate that they had experienced difficulty processing the language of the text. This was based on the assumption that the use of a regulatory strategy comes with a cost in terms of cognitive load, and as such is unlikely to be used unless the reader needs to overcome a specific problem. Hence, in relation to the goal of investigating the extent to which the participants' level of reading proficiency related to their ability to identify the incongruent words in the text, the analysis of whether they had used a comprehension regulation strategy was used to identify whether they had experienced difficulty processing the text. That is to say it was used to provide symptomatic evidence that a problem had occurred, although not to document how the strategy functioned as a remedial response. It is important to note that the decision to use the analysis of comprehension regulation strategies in this way was practical rather than principled in nature, and resulted from the need to integrate the data from the think-aloud and eye-tracking components, rather than from a belief that an analysis of the remedial impact of comprehension strategies is not important when investigating how second-language readers process incongruent words in a text. In fact, the analysis of

how the use of comprehension regulation strategies affected the types of mental representations the participants were able to form yielded a number of interesting results before it was discontinued in favour of the analytical approach ultimately reported in this thesis. That, however, is a story for another paper.

3.7.3.1 Preparing for the qualitative think-aloud analysis: The coding frameworks

As explained in Section 3.3.3, the think-aloud interviews were subjected to both quantitative and qualitative analysis in order to first triangulate and then elaborate on the data from the eye-tracking experiment. The analytical procedure for both these elements of the study are explained in Chapter 5, as it was felt that it would be most useful for the reader if the analytical procedure and the results were presented in close proximity. However, given that the qualitative analysis depended on the creation of coding frameworks which closely guided the analysis, it was felt to be important to explain in some detail how those frameworks were established. That information is in this section. However, for ease of reference, details of how the codes were applied is summarised again in the think-aloud results chapter.

While much of the research on both first and second-language reading comprehension has focused on the underlying skills which predict reading proficiency, in recent years there has been an increasing recognition that "assessing the outcome of processes at the text level, such as textbase memory and situation model quality, also provides information about component skills, which can then be related to performance on standardized reading comprehension tests" (Raudszus et al., 2019, p. 108). For example, Fox, (2009, p. 199 - emphasis added) conducted a meta-review of 45 published research articles relating to reading comprehension in which she explained her aim was:

> to address a central question linking reader characteristics, ***process***, and ***product***: How do reader characteristics relate to what we see from readers in terms of conscious processing while reading and learning from informational text and to the nature and quality of the mental representations they construct?

Defining what she meant by the product of reading, Fox (p. 197 – emphasis added) argued that:

> Relevant aspects of products included mental representations of text (***textbase***, ***situation model of text***, ***situation model of phenomenon***, and author model) and the quality of those representations.

Hence, she defined the products of reading comprehension directly in terms of the components of Kintsch's (1998) Construction-Integration model, to the extent that she specified the need to investigate the extent to which readers had established textbase or situation model representations, albeit with the added component of whether the reader is able to recognise the author's purpose in producing the text. When deciding which studies to include in her meta-review, Fox selected research in which the participants had engaged in both a think-aloud task which allowed the investigators a view into the processes they had engaged in as they read, and a secondary measure or the type of product they had been able to establish, including such tasks as immediate or delayed free recall of the text (e.g. Kardash & Howell, 2000; Linderholm & van den Broek, 2002), responses to comprehension questions (e.g. Moore & Scevak, 1997) ranking of documents based on credibility as historical sources (Wineburg, 1991), and evaluations of the author's argument (Fox et al., 2005; Schooler et al., 1996)." (p. 207). Using these studies, Fox then extrapolated the types of mental representations the participants had been able to establish as a result of how they performed on these secondary measures. For example, when assessing the impact of age or education level on a reader's ability to establish mental representations of a text, Fox (p. 218 – emphasis added) argued that:

> In these studies involving the role of school level, readers aimed variously and with varying degrees of success at construction of textbases; situation models of individual texts; situational models of a phenomenon, domain, or event across multiple texts; and models of the text as the product of the choices of an author. More experienced and more successful readers produced more coherent and complete recall reports and summaries... and demonstrated better performance on literal comprehension questions..., where such tasks ***might be expected to relate to the construction of a textbase*.**

Similarly, reviewing the relation between relevant prior knowledge and the types of mental representations the readers were able to establish, Fox (p. 230 - emphasis) stated that:

> In these studies involving the role of relevant prior knowledge, readers aimed variously at construction of different levels of mental representations and had varying degrees of success in these constructive activities. Relevant prior knowledge in the form of general world knowledge... and appropriate topic knowledge... ***enabled readers to construct more complete and coherent textbases as evidenced by their higher level of gist recall of the text than readers lacking such knowledge.***.. In addition, the possession of relevant prior knowledge enabled readers to construct an elaborated situation model of the text more successfully, ***as shown by their performance on tasks assessing their ability to interpret and infer beyond the direct information in the text***.

In this way, Fox (2009) attempted to draw a direct line between reader characteristics, in these cases either in terms of age, school level or relevant prior knowledge, and the types of mental representations they were able to establish of an expository text. However, in the majority of the studies that Fox included in her meta-review [although see Linderholm and van den Broek (2002) for an exception], the product of the reader's processing of the text was measured using a post-test delivered after they had finished reading. Thus, in some sense it could be argued that these various studies all provided insight into the *finished product* the readers were able to establish. However, they do not provide a view of how the product of reading comprehension, that is to say a reader's discourse-level mental representation of the meaning of a text, develops as they process the text on a word-by-word or line-by-line basis. It was with this end in mind that the qualitative think-aloud analysis reported on in this study was conducted. Specifically, the analysis attempted to investigate the participants' processing of the text from two perspectives, *process* and *product*, using the same transcripts from the think-aloud interviews as the source of data in both cases. As explained previously, in terms of analysing the *process*, a coding framework was developed to identify when a reader experienced difficulty processing a text, based on whether they employed any comprehension regulation strategy, regardless of what that strategy was. In terms of product, a coding framework was developed using Kintsch's (1998) Construction-Integration model as the basis for identifying what types of mental representations the participants were able to establish as they read. Due to constraints of time and space, and the need to integrate the think-aloud and eye-tracking components of the study, this analysis was only applied to the four target incongruent noun-phrases that were embedded in the text used for the think-

aloud interviews. The following sections will provide a description of how the two coding frameworks were developed.

3.7.3.2 Developing the comprehension regulation coding framework used to identify processing difficulties

The qualitative think-aloud analysis used a primarily top-down, deductive approach (Gu, 2014) as the aim was not to discover new comprehension regulation strategies, but rather just to identify when a recognised strategy had been used, where this was taken as a proxy for some kind of processing difficulty. The first step in developing the coding framework was to review the existing literature on reading comprehension strategies (Abbott, 2006; Anderson, 1991; Block, 1986; Ghavamnia et al., 2013; Tsai et al., 2010; Yapp et al., 2021). From these frameworks, only strategies which were argued to be used during online reading comprehension were selected for inclusion, while strategies such as those intended to be used during pre-reading (e.g. looking up vocabulary before reading), or those which focused on test taking (e.g. reading the questions before reading then scanning for the answers) were not included, as the participants did not engage in any of those types of activities in the current study. This process carried on until no new strategies could be identified.

3.7.3.3 Trialling and finalising the coding framework.

The first step in trialling the coding framework was to fully code one transcript to check if the strategies in the framework were seen to have been used in the participants' transcripts, and if any extra comprehension regulation strategies which were not included could be discovered (Gu, 2014). This resulted in the addition of fourteen new strategies. The revised comprehension regulation strategy framework contained a total of 37 strategies (see Table 3.3). While it is customary in reading or other learning strategy research to precisely define each strategy so that they may be differentiated, in the current study the employment of a comprehension regulation strategy by the participants was used only as a proxy measure to indicate that they had experienced some type of processing difficulty. Consequently, the regulation strategies were only loosely defined, as the identification of any strategy, regardless of what it may have been, was interpreted in

the same way, i.e. as indicating that the participant had experienced reading comprehension difficulty. As such, the addition of tighter definition would not have added to the accuracy of the findings of the current study.

Table 3. 3: Comprehension regulation strategies used to identify where a participant experienced processing difficulty

Reader revises their understanding of the text's meaning in the light of recently read information
Reader ties to guess the meaning of a word/phrase using metalinguistic knowledge or word meaning in order to repair their understanding
Reader ties to guess the meaning of a word/phrase using metalinguistic knowledge of grammar in order to repair their understanding
Reader ties to guess the meaning of a word/phrase using context in order to repair their understanding
Reader ties to visualise the scene or information they are reading
Reader skips a difficult section because they judge it to be unimportant
Reader skips a difficult section because they are unable to solve the difficulty
Reader compares a difficult word/phrase to its L1 equivalent to see if this brings clarity
Reader translates an extended section of the text into their L1 to try to understand it
Reader attempts to reformulate a section of the text in order to bring clarity
Reader compares a difficult word/phrase with another L2 word (e.g. synonym) to see if this brings clarity
Reader uses background related to the topic of the text to overcome comprehension difficulty
Reader uses metalinguistic knowledge of morphemes to overcome comprehension difficulty with word meaning
Reader uses metalinguistic knowledge of discourse structure to overcome comprehension difficulty
Reader uses metalinguistic knowledge of lexical cohesion to overcome comprehension difficulty
Reader uses metalinguistic knowledge of grammatical cohesion to overcome comprehension difficulty
Reader uses metalinguistic knowledge of structural cohesion to overcome comprehension difficulty
Reader rereads section of the text to overcome comprehension difficulty
Reader acknowledges comprehension problem, then reads on in order to regulate comprehension
Reader proposes possible meanings for an unclear element in the text in order to overcome a comprehension difficulty
Reader looks for possible rhetorical intent on the part of the author to overcome comprehension difficulty (e.g. sarcasm)
Reader uses metalinguistic knowledge of how a relative clause relates elements in a clause complex to regulate their comprehension
Reader uses metalinguistic knowledge of complex noun group construction to regulate their comprehension
Reader uses metalinguistic knowledge of tense or aspect in the verb group to regulate their comprehension

Reader parses the sentence in order to regulate their comprehension (e.g. identifies subject, verb, object, main clause)

Reader explicitly identifies words/phrases/ideas they do understand in a section they otherwise don't understand

Reader identifies that a word is referring to a previously expressed idea and searches for the referent

Reader formulates a simple summary of the meaning of a sentence or phrase that they have struggled to comprehend

Reader ignores words or phrases they don't know

Reader makes an inference in order to regulate comprehension when they don't have a very clear understanding of a word/sentence

Reader identifies the rhetorical structure of the text and uses this to regulate their comprehension

Reader identifies the logical structure of a sentence or section of text (e.g. reason-result)

Reader regulates comprehension by linking unknown element to a point made previously

Reader regulates comprehension by linking information in adjacent sentences

Reader uses semantic relationship between clauses/sections to regulate comprehension (e.g. negative-negative, negative-positive)

Reader identifies keywords to regulate comprehension

Reader purposefully scans ahead in order to ascertain the semantic/grammatical structure of the text so as to ease later reading

Reader uses punctuation to regulate comprehension

The next step was to perform an inter-rater reliability test to make sure that the codes could be applied with a reasonable degree of objectivity. To this end, three transcripts were selected, one from each of the three reading proficiency groups, representing a quarter of the total. Initially, the plan was to use short extracts from a large number of participants. However, in practice, this was found not to be a satisfactory approach, because the article was composed of three paragraphs which developed the main ideas and themes of the article such that the information in the later paragraphs assumed knowledge of information in the earlier ones. Consequently, it was reasoned that the reading behaviours and strategies the readers engaged in might change as they progressed through the text. Therefore, it was decided that using full transcripts would provide a more accurate result in terms of inter-rater reliability.

The process of finding a suitable candidate to perform as an inter-rater was not straight forward. This was because they needed to have not only strong knowledge of reading strategy use in L2 populations, but also to be fluent in Chinese, as this was the language used predominantly by the lower-intermediate participants.

In practice, such a candidate could not be found. As an alternative, a native-Chinese speaker who was highly familiar with second-language listening strategies agreed to participate. This was felt to be an acceptable substitute as there are substantial similarities between these two areas. Furthermore, this arrangement also meant that in addition to verifying the application of the strategies, the second rater was also able to provide a reliability check of my understanding of what had been said in Chinese.

This notwithstanding, although there are areas of cross-over between second-language reading and listening strategy use, there are also substantial differences. With this in mind, rather than the second inter-rater recoding the transcripts independently, a process that would have required more preparation in terms of familiarising himself with the theoretical basis of the coding framework than was practical, it was decided that he would instead review three transcripts which had already been coded. Using the coding framework as a guide, next to each code that had been applied he marked either **AGREE**, **DISAGREE** or **NEEDS EXPLANATION**. Once he had performed this task, a meeting was held to discuss the result. The results indicated that there was broad agreement between the two raters in respect of which strategies the readers had used (83.5%). Consequently, the list of strategies was judged as being clear enough to be applied with the degree of objectivity required for the present study. Detailed examples of how comprehension regulation strategies were identified in practice are provided in Chapter 5.

3.7.4 The Construction-Integration model of reading comprehension framework

As stated, the coding framework which was used to analyse the types of mental representations the participants established as they processed the language of the text was based on Kintsch's (1998) Construction-Integration Model of Reading Comprehension. This was felt to be preferable to Gernsbacher's (1997) Structure Building Framework, the model used by Walter, (2007), as while the two models adopt a similar view of the fundamental cognitive processes which underpin comprehension, the Construction-Integration (CI) model is more explicitly focused on text processing (Gernsbacher & Faust, 1995). The CI model argues there are various elements which may be present in a reader's mental representation of a

text, and while it would be inaccurate to imagine these as existing separately in the mind of the reader, or developing in a strict linear fashion, for ease of reference it can be useful to consider them in this way (Kintsch & Rawson, 2005). The Construction-Integration model proposes that a reader's mental representations are composed of three main elements. The first of these are what Kintsch (1998) refers to as *propositional* representations. Loosely put, these refer to the representations that the reader establishes of the ideational content of individual sentences, and that are formed as the reader accesses and integrates the various lexical and grammatical elements of which the sentence is composed. The second level, he explains, is the *textbase*. The textbase refers to the discourse-level representations the reader is able to establish of the information which is directly encoded in the language of the text. The textbase itself is composed of two sub-elements: the *microstructure* and the *macrostructure*. The microstructure is formed as the reader establishes an integrated, detailed representation of sentence-level propositions which are in close proximity to each other, as indicated by the cohesive devices used by the writer. The macrostructure is a higher-order memory representation, in which the meaning of discrete sections of the text, as represented in the microstructure, are subsumed to form a representation of the main topics discussed in the text. As such, macrostructure representations are much less detailed than microstructure representations, frequently expressible in just a few words, but are, perhaps as a result, much more available for recall after reading is complete (Kintsch, 1998). In more conventional ESL/EFL terminology, the microstructure could be considered as a local representation, while the macrostructure may be understood as a more global representation, terms often used in L2 reading research and practice (Grabe & Stoller, 2015). The final element Kintsch argues must be present for a complete representation of the meaning of a text to be established in the mind of a reader is a *situation model.* This is the part of the reader's representation which is formed from their pre-existing knowledge of, and attitudes and opinions towards, the situation being discussed in the text. This information is activated in the reader's long-term memory as they process the language of the text, and is fused with the information directly stated to become part of a new memory structure.

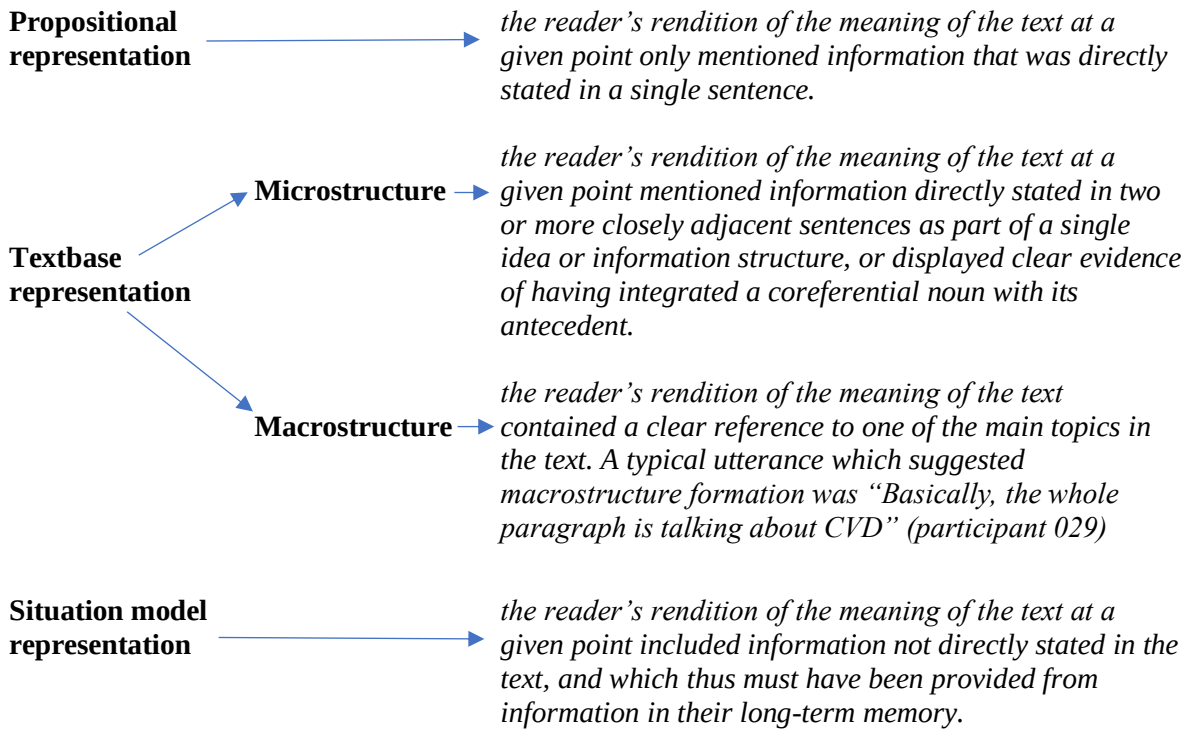
The role of the framework developed in reference to the CI model was to allow judgements to be made about the representations the readers were able to establish as they processed the language of the text by analysing what they thought the text meant as they read each sentence, based on their verbalisations. Inevitably, it was not possible to know with certainty exactly what type of representations the participants were able to establish, and so the framework was only intended to allow for evidence-based inferences to be drawn according to a predetermined set of criteria (see Figure 3.1). That is, the types of mental representations the readers were judged to have formed were based on the information that was in the rendition of the meaning of the sentences they gave. As discussed in Section 3.3.2, one limitation of using think-aloud protocols is that the researcher only has access to such thoughts as the participant gives voice to. Consequently, the possibility always exists that participants formed mental representations which they did not verbalise, or did not verbalise in full (Seliger, 1983). As when judging whether or not the participants had identified the incongruent words in the text or integrated the target nouns with their antecedents, only those instances where it was possible to clearly discern the type of representation established were selected for analysis.

The first major division in the coding framework was between *propositional* and *microstructure* representations. Participants were judged as having only established a propositional representation if their rendition of the meaning of the text at a given point only contained information which was directly stated in a single sentence. They were judged as having established microstructure representation if their rendition of the meaning of a section of text included information from multiple sentences fused into a single idea structure, or they displayed clear evidence of having integrated a coreferential noun with its antecedent in the way described above.

Ancillary to this, judgements were also made as to whether the participants had established macrostructure representations of the text. This was typically the hardest judgement to make as it was rare that direct evidence presented itself. Rather, it was necessary to infer the establishment of macrostructure

representations from what the participants said. This was done in two ways. Firstly, it was frequently found that when students were asked how they knew a noun was coreferential, they would say something along the lines of "Basically, the whole paragraph is talking about CVD" (Participant 029)", indicating that they had established a representation of this as the main topic of the article. Secondly, it was judged on occasion that when a reader experienced difficulty processing an incongruent word, that this could only have resulted from them having established a macrostructure representation of one of the main topics of the article. For example, when Participant 032 read the incongruent noun phrase *language problems*, despite having struggled with the language of the immediately prior text, she nonetheless showed clear evidence of having identified the incongruent word. The reason *language* was incongruent in this context was that it clashed with the main discourse topic of the article, which was *health problems*. As such, it was inferred that Participant 032 must have established a macrostructure representation of this main topic, and thus experienced processing difficulty when she tried to integrate the incongruent noun phrase *language problems* into this macrostructural representation.

Figure 3. 1: The coding criteria used to judge the type of mental representations established



The second major division was between textbase and situation model representations. The participants were judged as having established textbase representations if their rendition of the meaning of the text only contained information directly stated in the text. They were judged as having established situation model representations if their renditions of the text contained information that was not directly stated in the text, and thus must have been activated in long-term memory.

It is hoped that using these criteria brought a degree of objectivity to the analysis, by allowing concrete judgements to be made about how to classify the type of representation that was established in any given instance. However, when reading these results, it must at all times be remembered that the ability to know with certainty the exact nature of a reader's mental representations is necessarily limited, and at times inferences need to be drawn (White et al., 2007). It is always possible that a reader may have established textbase or situation model representations, but not provided evidence of this in their rendition of the text's meaning. As such, any findings reported in this regard should be viewed as being only indicative of the types of mental representations the readers were able to establish, rather than as being definitive.

3.8 Summary

In order to address the research questions posed at the end Chapter 2, this study reported in this thesis adopted a mixed methods approach, composed of an initial eye-tracking experiment and subsequent think-aloud interviews. The eye-tracking experiment was used to gather data on how the participants processed incongruent words embedded in coreferential noun phrases. The role of the think-aloud interviews was two-fold. Firstly they were used to triangulate with the data from the eye-tracking experiment. This was done as although eye movement data can give a strong indication that a reader experienced processing difficulty, it cannot provide proof positive, as the researcher can never be sure exactly why they fixated any given word for the time they did. The use of the think-aloud thus allowed another view on how they interacted with the incongruent words in the text. Secondly, the think-aloud interviews were used to elaborate on the data from the eye-tracking experiment, by providing insight into any specific difficulties they participants may have

experienced as they read the text, and what, if any, effects this had on the types of mental representation they were able to establish as they read, or their ability to identify the incongruent words embedded in the text.

3.9 Ethics statement

The study reported on in this thesis was conducted in full conformity with the University of Reading's rules for ethical conduct. Before both the pilot and the full study permission was sought from, and granted by, the University Research Ethics Committee who gave a favourable ethical opinion for conduct. Prior to participation, all participants were provided with a information sheet describing what would happen in the data gathering session, what the purpose of the research was, how they could withdraw consent from the process, as well as the complaints procedure should they wish to make ones. The participants were asked to complete a consent form expressly giving their consent for every stage of the data gathering processes. Finally, prior to both the pilot and the main study a Health and Safety review was conducted, in full conformity with the University of Reading's Health and Safety policies.

Chapter 4: Results – The eye-tracking experiment

4.1 Introduction

The aim of this study was to investigate the extent to which native-Chinese speaking readers of academic English texts are able to integrate the meanings of **coreferential noun phrases**, i.e. two or more noun phrases which refer to the same idea or entity, in order to establish discourse-level mental representations of the texts they read. Following Oakhill et al. (2005) and Walter (2007), the research design of this project was underpinned by two key assumptions:

> **Assumption 1**: A reader's ability to integrate a coreferential noun phrase with its antecedent can be tested by investigating their ability to identify an incongruent word in a coreferential noun phrase which contradicts an idea expressed in the antecedent noun phrase.

As explained in Section 2.2, the English language contains a lexicogrammatical system of cohesive devices which serve to indicate to a reader how the ideas and entities in a text are logically or semantically related. While there are several types of cohesive devices, the category most relevant to the processing of coreferential noun phrases in academic writing are **determiners** (Francis, 1994; Halliday, 1994a). Determiners include such words as *this*, *those*, and *such*. Given that the presence of a determiner in a noun phrase explicitly indicates that it is coreferential with a previous noun phrase, at least in so far as the grammar is concerned (Halliday & Matthiessen, 2014), it arguably follows that the presence of a determiner ought to aid the reader to psycholinguistically integrate a coreferential noun phrase with its antecedent as they read. This gives rise to the second assumption which underpinned the research design of this project:

> **Assumption 2**: The presence of a determiner in a coreferential noun phrase ought to help a reader integrate that noun phrase with its antecedent, and so to identify any incongruent words that noun phrase may contain.

As explained in Section 3.3.3, in order to further investigate these issues, an eye-tracking experiment was conducted that sought to shed light on the following three questions:

RQ1. To what extent are native-Chinese speaking readers of English able to identify words in coreferential noun phrases which are incongruent with an idea expressed in its antecedent?

RQ2. To what extent does the presence of a determiner in a coreferential noun phrase help native-Chinese speaking readers of English identify words in that noun phrase which are incongruent with an idea expressed in its antecedent?

RQ3. To what extent do native-Chinese speaking readers at different levels of reading proficiency differ (i) in their ability to identify incongruent words in coreferential noun phrases, or (ii) in the extent to which they are helped to identify incongruent words in such a noun phrase by the presence of a determiner?

This chapter will report the results of that eye-tracking experiment. Questions one and two will be addressed simultaneously in Section 4.6, by reporting the results of **2x2 repeated measures factorial ANOVAs** that were performed. The two factors were (*i*) *Determiner* vs *No Determiner* and (*ii*) *Congruent* vs *Incongruent*.

Question three will be addressed in Section 4.7, by reporting the results of two statistical analyses. Firstly, it reports the results of **a second round of 2x2 repeated measures factorial Anovas in which the sample was split into three proficiency levels**. The two factors were again (*i*) *Determiner* vs *No Determiner*, and (*ii*) *Congruent* vs *Incongruent*. Secondly, it reports the results of **multiple regression analyses** that were performed in order to investigate to what extent individual differences in terms of *vocabulary knowledge*, *grammar knowledge*, *short-term working memory in English*, and *short-term working memory in Chinese* predicted differences in how the participants processed the target nouns based on the presence or absence of a determiner in the noun phrase, or there was either a congruent or an incongruent word in the noun phrase. Before these results are presented, some brief information is provided about the sampling procedure used, the sample gathered, any data cleansing that was performed, as well as descriptive statistics.

4.2 Results from the measurements of individual difference

As mentioned in Section 3.6, prior to the eye-tracking session, each of the participants' most recent IELTS reading score was recorded, and they completed a vocabulary test, a grammar test, and working memory tests in both English and Chinese. The results of these various tests is shown in Table 4.1, alongside scores for the five comprehension questions completed by participants after reading each of the four articles in the

eye-tracking session. All scores, with the exception of the IELTS reading scores, have been converted to percentages for ease of reference.

Table 4. 1: Average IELTS reading scores, individual difference, and comprehension of the eye-tracking texts – all averages are means except for the IELTS reading scores which shows the mode

	Average	*SD*	*Max*	*Min*
IELTS Reading Score	6.5	NA	9	4.5
Vocabulary (%)	62.59	15.57	92.52	19.73
Grammar (%)	43.51	19.55	86.36	4.55
Working memory: English (%)	58.82	12.07	96.67	36.67
Working memory: Chinese (%)	88.74	8.58	103.33	66.67
Comprehension (%)	66.37	15.90	94.74	31.58

Inspection of Table 4.1 indicates that overall the participants performed well on the comprehension questions they answered after they read each of the eye-tracking texts, indicating that they did understand the articles they read. Nonetheless, there were twelve participants who scored lower than 50%. These lower scores may suggest that these readers did not understand the texts fully. However, where a participant did answer fewer than three of the five questions after each article accurately, they were asked to give a brief description of the article they had just read. In each instance they were able to explain the overall gist of the article. Hence, although it may be considered unconventional, with a view to maintaining maximum possible statistical power , it was accepted that all of the participants had understood the main meaning of all the articles. Nonetheless, the possibility remains that some participants may not have fully comprehended the articles, and that this may have affected the results.

4.3 Data preparation in Data Viewer

The data were first prepared using SR Research's Data Viewer software (SR Research, 2018). The first step was to merge or delete fixations whose durations fell above 1500 milliseconds or below 80 milliseconds. This was done because fixations which are too short, 80 milliseconds or less, suggest that no information could have been extracted by the reader; on the other hand, fixations which are too long, 1500 milliseconds or more, suggest that the reader was engaged in processes other than reading (Rayner, 1998). Furthermore,

small random movements (microsaccades) of the eye may result in the eye tracker recording a number of very short fixations when in fact the reader was processing one word (Rayner, 1998). To perform this cleansing, spatially and temporally adjacent fixations (measured in degrees of angle or milliseconds respectively) which were separated by no more than a pre-specified distance were merged. In Data Viewer, this cleansing procedure is achieved using the software's Four-stage fixation cleansing function (see Appendix E for a full description of this process)

Once the data had been cleaned, they were exported from Data Viewer using the Interest Area Report function (SR Research, 2018, p. 135). This function allows the user to specify which variables to extract data for, and to assemble a report in which each row contains the eye movement data pertaining to one word in the text. For the current analysis, four measures were chosen. The first two, **first fixation duration** and **gaze duration**, comprised the **early measures**. Early measures are argued to reflect a reader's initial processing of a word, including such features as phonological, morphological and lexical access (Rayner, 1998). In Kintsch's (1998) terminology, these reflect the Construction phase of the Construction-Integration (CI) model (see Section 2.4 for a description of the CI model). The **first fixation duration** measures the duration of the first fixation made on a word or region of interest. **The gaze duration** measures the duration of all first pass fixations on a word or region of interest before the eye moves to another region. This second measure is necessary as in the case of long, multisyllabic words, it is common for the reader to make multiple fixations in order to process the entire word. The second two measures, **selective regression path duration** and **second pass reading time**, comprised the **late measures**. Late measures are argued to reflect comprehension processes, where the word is integrated into the reader's developing mental model of the text, and therefore the second phase of the Construction-Integration process (Staub & Rayner, 2007). The **selective regression path duration** measures the amount of time the reader spent reading and rereading the target word before they moved to the next word to the right. Selective regression path duration is similar to the more commonly used **regression path duration,** which measures all of the time that passes from when a

reader's eye first lands on a word until it moves to the next word to the right, including time spent reading previous words or regions in the text. It was decided to use the selective regression path duration measure because while the selective regression path duration reflects the fact that the reader reread the previous text, evidenced by the fact that they fixated the target word more than once before moving their eyes to the right, it does not include the time the reader actually spent reading the previous text. As a result, from a statistical point of view, the selective regression path duration measure contains less variance, as it is not always clear how much text individual participants needed to reread, and thus how much time they spent doing so. In pilot testing, it was found that the data gathered using the regression path duration measure had extremely large standard deviations, suggesting that the means did not represent a good model of the participants' reading times (Field, 2013). Consequently, this measure was not used for the full study. Finally, the **second pass reading time** measures the time spent rereading a target word or region after the reader's eye had already moved past it to the right at least once. As with all eye-tracking measures, longer reading times are argued to be indicative of processing difficulty (Conklin & Pellicer-Sánchez, 2016). Longer reading times in the early measures are argued to indicate difficulty with initial processing of a word, while longer reading times in the late measures are argued to indicate difficulties integrating the word into the reader's developing mental model of the meaning of the text (Rayner, 1998).

4.4 Data preparation in Excel

Because the report produced by Data Viewer contained data on every word in the four texts used, it was necessary to use Excel to create a data file which only contained information on the thirty-six target noun phrases the analysis focused upon. This was achieved using the Macros-VBA function in Excel. This process resulted in a data table 2938 rows long, with each row containing the eye-tracking data from one participant reading one target word. Given that there were eighty-two participants, and each should have read thirty-six target words, there should have been a total of 2952 rows, meaning that data for fourteen of the target words were missing. It is not clear why this happened, although it is likely the results were removed during

Data Viewer's four step cleansing process. Furthermore, inspection of the data table revealed that 564 target words had not been fixated (indicated by a value of 0 appearing in the first fixation duration column), meaning that no usable data had been gathered on them. It was decided to remove these non-fixated words from the final data set so that they would not distort later statistical analysis. This resulted in a final data set of 2374 target words across the 82 participants.

4.5 Descriptive statistics and data cleansing

This section reports the descriptive statistics for each of the four eye-tracking measures, as well as any cleansing procedures that were applied. Given the large sample size that was collected for this project, as well as the large number of target words which were analysed, the assumption of normality is argued to have been met by virtue of the Central Limit Theorem (Field, 2013).However, Kolmogorov-Smirnov tests were also run, and P-P plots and boxplots produced, and where necessary logarithmic transformations were applied. Because these data were analysed using Factorial repeated measures ANOVAs, the assumptions of linearity and homogeneity of variance were not tested (Field, 2013).

4.5.1 Descriptive statistics prior to the ANOVAs

The 2x2 repeated measures factorial ANOVAs performed had two factors, a **determiner** factor and a **congruency** factor. Each of these factors had two levels. The two levels for the determiner factor were *Determiner* vs. *No Determiner*, reflecting the fact that the target noun phrases either contained, or did not contain, a determiner. The two levels for the congruency factor were *Congruent* vs. *Incongruent*, reflecting the fact that the target noun phrases contained either a congruent or an incongruent word. In each case, the dependent variable was the relevant fixation duration on the main noun in the target noun phrase, depending on which eye movement measure was being used (see Section 3.5 for full details of the target noun phrases). As a result, there were four possible combinations in which a target noun could appear:

Factor 1	Factor 2	Dependent variable
determiner	congruent word	READING TIME OF TARGET NOUN
determiner	incongruent word	READING TIME OF TARGET NOUN
no determiner	congruent word	READING TIME OF TARGET NOUN
no determiner	incongruent word	READING TIME OF TARGET NOUN

The ANOVA was conducted separately for each eye movement measure. Before conducting the ANOVAs, descriptive statistics were generated for these four combinations for each of the four eye-tracking measures used. The results are shown in Tables 4.2 to 4.5:

Table 4. 2: Descriptive statistics and Kolmogorov-Smirnov test for first fixation duration

			K-S test	
	M (ms)	*SD*	*D*(82)	*Sig.*
Determiner-Congruent condition	244	40.85	.087	.188
Determiner- Incongruent condition	243	44.47	.087	.184
No Determiner-Congruent condition	260	49.47	.068	.200
No Determiner-Incongruent condition	259	54.54	.089	.160

Table 4. 3: Descriptive statistics and Kolmogorov-Smirnov test for gaze duration

			K-S test	
	M (ms)	*SD*	*D*(82)	*Sig.*
Determiner-Congruent condition	382	107	.123	.004
Determiner- Incongruent condition	369	105	.117	.008
No Determiner-Congruent condition	390	104	.079	.2
No Determiner-Incongruent condition	400	121	.095	.064

Table 4. 4: Descriptive statistics and Kolmogorov-Smirnov test for selective regression path duration

			K-S test	
	M (ms)	*SD*	*D*(82)	*Sig.*
Determiner-Congruent condition	439	127	.095	.065
Determiner- Incongruent condition	468	176	.123	.004
No Determiner-Congruent condition	447	128	.106	.023
No Determiner-Incongruent condition	484	131	.058	.200

Table 4. 5: Descriptive statistics and Kolmogorov-Smirnov test for second pass reading time

			K-S test	
	M (ms)	*SD*	*D*(82)	*Sig.*
Determiner-Congruent condition	281	206	.116	.008
Determiner- Incongruent condition	362	267	.102	.034
No Determiner-Congruent condition	300	238	.135	.001
No Determiner-Incongruent condition	381	263	.132	.001

4.5.2 Data cleansing

The data from all four of the eye-tracking measures contained a number of outliers. Furthermore, as revealed by the results of the Kolmogorov-Smirnov tests reported above, some of the distributions deviated significantly from normality. In order to remove these outliers, as well as to ensure that the data conformed as closely as possible to a normal distribution, the data were logarithmically transformed. In all four eye-tracking measures this resulted in normally distributed data (see Appendix F for the transformed values). Furthermore, there were a number of instances where, even after logarithmic transformation, outliers remained in some of the distributions. In these cases, the data were trimmed. To decide how best to do this, two studies were consulted. Firstly, an L1 discourse processing study, Mak and Sanders (2013), which investigated the processing of cohesion relationships in discourse, was used. In addition, a study by Bott and Gattnar (2015), which looked at grammatical processing by L2 readers was consulted, to see if a different approach had been taken. Comparison of these two studies revealed that both had trimmed outliers which fell more than 2.5 standard deviations above or below the mean. As this is the same approach recommended by Field (2013), and is the one commonly used in eye-tracking studies (Holly Joseph, personal communication) it was decided that in the current study outliers falling more than 2.5 standard deviations above and below the mean would be trimmed from the data.

All of the repeated measures factorial ANOVAs conducted for this analysis used the data sets that resulted from the cleansing processes described above. Although the analyses were performed using the

logarithmically transformed data, for ease of reference, only the raw means in milliseconds are reported below (although see Appendix F for tables with the logarithmically transformed values).

4.6 Results: Research Question 1 and Research Question 2

The following section will report the results of the 2x2 repeated measures factorial ANOVAs conducted to address the first of the two research questions posed above:

> ***RQ1. To what extent are native-Chinese speaking readers able identify words in coreferential noun phrases which are incongruent with an idea expressed in its antecedent?***
>
> ***RQ2. To what extent does the presence of a determiner in a coreferential noun phrase help native-Chinese speaking readers identify words in that noun phrase which are incongruent with an idea expressed in its antecedent?***

Prior to the analysis, the following three predictions were made:

i.	The presence of a **determiner** in a noun phrase with a congruent word should resulted in ***shorter*** reading times of the main noun in the target noun phrase than where there is **no determiner**.	*This is because the determiner should explicitly indicate that the noun phrase is coreferential with a previous noun phrase, and so should be integrated.*
ii.	The presence of an **incongruent** word in a noun phrase should resulted in ***longer*** reading times of the main noun in the target noun phrase than where there is a **congruent** word.	*This is because the presence of the incongruent word should cause processing difficulty, resulting in longer processing times.*
iii.	Where there is a **determiner** AND an **incongruent** word in a noun phrase, this should result in even longer processing times of the main noun in the target noun phrase than where there is **no determiner**. This should result in a significant interaction between the determiner and congruency variables	*This is because the presence of the determiner should explicitly indicate that the noun phrase is coreferential with a previous noun phrase, thereby highlighting the incongruency, resulting in increased processing difficulty*

The results for each of the four dependent variables (i.e. eye-tracking measures) are reported separately. It must be noted that for all of the tests conducted in this section, SPSS did not return a *p*-value of Mauchly's test of sphericity as there were zero degrees of freedom. However, it should also be noted that no differences were observed for the *F* values where sphericity was assumed, or where any corrections (e.g. Greenhouse-Geisser) were applied. Accordingly, the assumption of sphericity is argued to have been met (Field, 2013). The effect size used was Partial Eta Squared (η_p^2). When categorizing effect sizes as small,

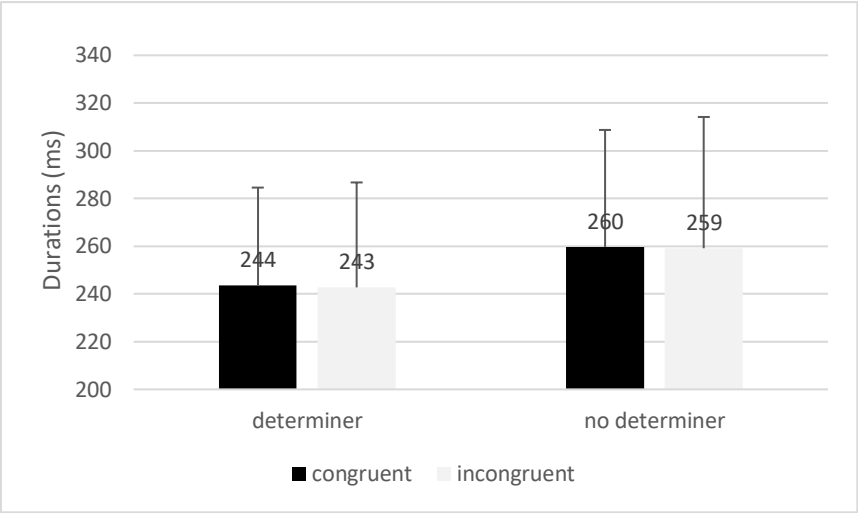
medium or large the following values were used: **Small** – .01, **Medium** – .06, **Large** – .14 (http://www.psy.gla.ac.uk/~steve/best/effect.html). In order to reduce the chance of a Type 1 error occurring, Bonferroni corrections were calculated. Following von der Malsburg and Angele (2017), the number of comparisons to be included in the correction included the number of eye-tracking measures used, which was four. In order to account for the high degree of correlation which typical exists between eye-tracking measures (von der Malsburg & Angele, 2017), part of the calculation to arrive at the new alpha included an adjustment for correlation. Following Joseph et al. (2021), the correlation was set at 0.6. Although this is relatively conservative, given that correlations can be much higher between certain eye-tracking measures (von der Malsburg & Angele, 2017), it is felt that a more conservative measure would also account for the fact that the actual correlations were not known. The adjusted alpha was calculated using an online tool, the SISA Bonferroni Calculator (www.quantitativeskills.com). The result of the Bonferroni correction indicated that the new alpha should be set at ***p* = .013**.

4.6.1 Early measure: *First fixation duration*

The 2x2 repeated measures factorial ANOVA conducted on the data from the first fixation duration measure indicated that there **was a significant main effect of determiner** $F(1) = 21.25$, $p = <.001$ with a **large effect size**, $\eta_p^2 = .218$. There was no significant main effect of congruency $F(1) = .032$, $p = .858$, $\eta_p^2 = .000$, and no significant interaction between the determiner and congruency factors $F(1) = .003$, $p = .957$, $\eta_p^2 = .000$.

This reveals that in the earliest of the four eye-tracking measures used, only the determiner variable produced a significant main effect, with shorter fixations when there was a determiner in the target noun phrase versus when there was not (see Figure 4.1).

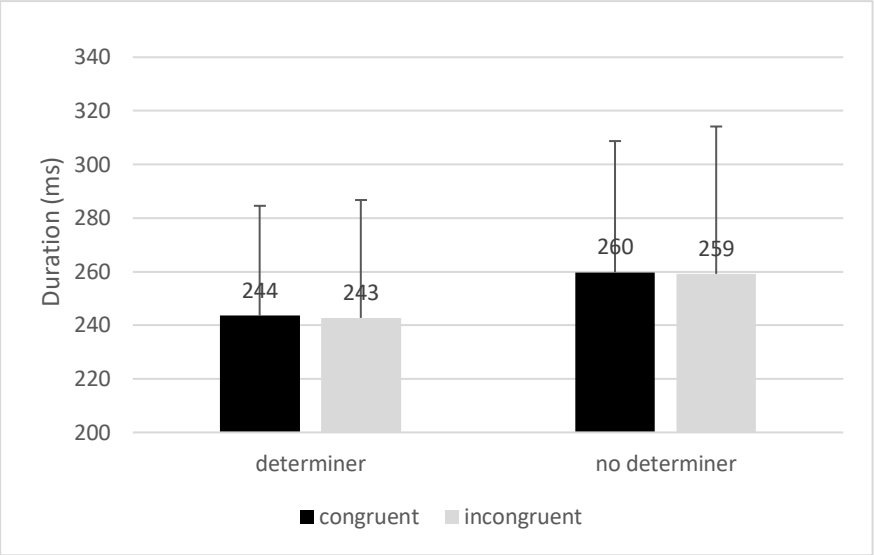
Figure 4. 1: Mean first fixation durations on the target nouns as a function of congruency and presence of a determiner (units in milliseconds) – error bars show standard deviation



4.6.2 Early measure: *Gaze duration*

The 2x2 repeated measures factorial ANOVA conducted on the gaze duration measure indicated that there was a significant main effect of determiner F(1) = 6.68, *p* = .012, with a medium-large effect size, η_p^2= .080. There was no significant main effect of congruency F(1) =.016, *p* = .899, η_p^2= .000, and no significant interaction between the determiner and congruency factors F(1) =.987, *p* = .324, η_p^2= 0.13. Hence it would appear that the determiner factor had only a very early and fleeting effect on reading times in the very earliest of the measures, and had faded substantially by the time of even the gaze duration measure.

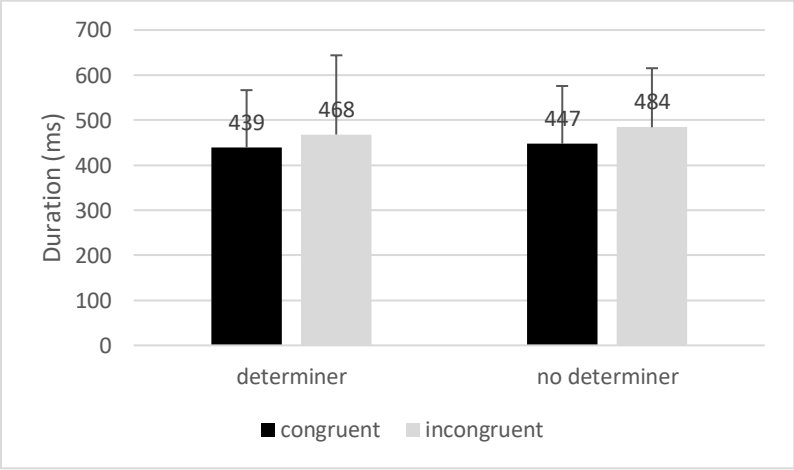
Figure 4. 2: Mean gaze durations on the target nouns as a function of congruency and presence of determiner (units in milliseconds) – error bars show standard deviation



4.6.3 Late measures: *Selective regression path duration*

The 2x2 repeated measures factorial ANOVA conducted on the selective regression path duration measure indicated that there was no significant main effect of determiner $F(1) = 2.351$, $p = .129$, $\eta_p^2 = .030$. and no significant main effect of congruency $F(1) = 5.451$, $p = .022$, $\eta_p^2 = .066$, although there was a medium-large effect size,. There was also no significant interaction between the determiner and congruency factors $F(1) = .912$, $p = .342$, $\eta_p^2 = .012$.

Figure 4. 3: Mean selective regression path durations on the target nouns as a function of congruency and presence of determiner (units in milliseconds) – error bars show standard deviation

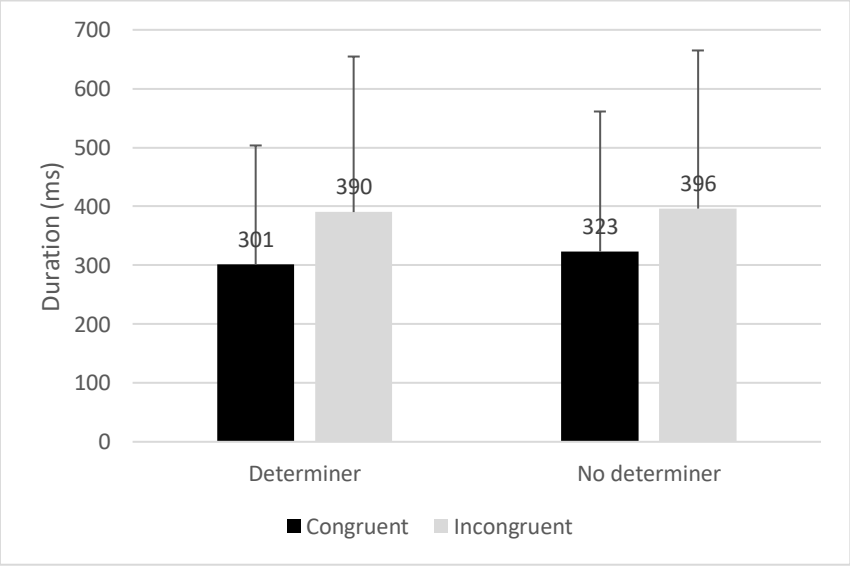


4.6.4 Late measure: *Second pass reading time*

The 2x2 repeated measures factorial ANOVA conducted on the second pass reading time measure indicated that there was no significant main effect of determiner $F(1) = .069$, $p = .793$, $\eta_p^2 = .001$. However, there was a significant main effect of congruency with a large effect size $F(1) = 13.38$, $p = <.001$, $\eta_p^2 = 16.2.$, with longer processing times when there was an incongruent than a congruent word in the target noun phrase. There was no significant interaction between the determiner and congruency factors $F(1) = .027$, $p = .870$, $\eta_p^2 = <.000$.

Hence, in the latest of the measures, there was a significant main effect for the congruency factor. As predicted, there were longer processing times when there was an incongruent word in the target noun phrase versus when there was a congruent word. This indicates that, overall, the participants did experience difficulty with integration processing when there was an incongruent word in a target noun phrase, suggesting they had identified the incongruent words in the text.

Figure 4. 4: Mean second pass reading time on the target nouns as a function of congruency and presence of determiner (units in milliseconds) – error bars show standard deviation



4.6.5 Summary

The analysis in this section set out to answer the first two research questions posed at the start of this chapter:

RQ1: To what extent are native-Chinese speaking readers able identify words in coreferential noun phrases which are incongruent with an idea expressed in its antecedent?

RQ2: To what extent does the presence of a determiner in a coreferential noun phrase help native-Chinese speaking readers identify words in that noun phrase which are incongruent with an idea expressed in its antecedent?

In answer to **research question 1,** it can be said that, overall, the participants were able to detect the incongruent words embedded in the text, as indicated by the significant main effect found for the congruency factor in the second pass reading time measure. In this respect, the predictions made above were confirmed, with significantly longer reading times being recorded when there was an incongruent word in the target noun phrase. It may be remembered from Section 3.3.1 that, as a late measure, the second pass reading time is argued to indicate integration processing. It is in the late measures we would expect to see longer processing times for noun phrases containing an incongruent word, as it is not until the reader integrates the incongruent noun phrase with its antecedent that the incongruency becomes apparent.

However, in answer to **research question two**, and contrary to expectations, it was found that there was no significant interaction between the determiner and the congruency factors. This suggests that although the participants were able to detect the incongruent words in the text, the presence of a determiner did not help them to do so. Rather, it was found that although there was a significant main effect for the determiner factor in the first fixation and gaze duration measures, with the presence of a determiner leading to shorter processing times of the target noun, this effect faded rapidly, and was no longer in evidence in later measures. This result indicates that although determiners may facilitate initial processing of a coreferential noun, they play little to no role in subsequent integration processing.

4.7 Results: Research question 3

The following section reports the results of statistical tests that were conducted to shed light on the third research question posed at the start of this chapter:

RQ3: To what extent do native-Chinese speaking readers at different levels of reading proficiency differ (i) in their ability to identify incongruent words in coreferential noun phrases, or (ii) in the extent to which they are helped to identify incongruent words in such a noun phrase by the presence of a determiner?

In order to address this question, two statistical tests were used. Firstly, the original sample of 81 participants was divided by proficiency level, according to their IELTS reading scores, to form three groups: ***Lower-intermediate***, ***Intermediate***, and ***Upper-intermediate***. The 2x2 repeated measures factorial ANOVAs were then reconducted on each of these groups independently. Details of these groups are provided in table 4.6.

It should be noted at the outset that this approach is not without controversy. A more typical method of investigating the effect of proficiency on how participants perform in respect of a given experimental factor would be to include their proficiency scores as either a third factor, or as covariates. If a general effect of proficiency were detected in this kind of omnibus testing, the next step would be to further interrogate the data, by splitting the sample into smaller groups in order to see how participants at specific proficiency levels performed (Field, 2013). This approach was not used in the current study, however, as it was felt that IELTS reading scores were not suitable for inclusion in this type of statistical test. This is because the IELTS band scores to which test takers are assigned are composites of four other scores, the details of which are opaque, at least to observers working outside of the testing organisations. Consequently, given the opaque nature of the original data which informed how the participants had been allocated to a given IELTS band grouping, it was felt that including these band scores in rigorous statistical tests such as ANOVAS or ANCOVAs would amount to asserting an almost spurious level of accuracy, as the broad nature of the scores do not allow for fine-bone distinctions to be made between participants at different levels.

Table 4. 6: The three IELTS proficiency levels

Proficiency level	IELTS Reading scores	Number of participants per group
Lower-intermediate	4.5 – 6	26
Intermediate	6.5 – 7	29
Upper-intermediate	7.5 – 9	27

Nor indeed, need this constitute a problem. This is because IELTS scores are not *intended* to be used as statistical measures, but rather to allow educational institutes to divide students into proficiency groups so that they may receive lessons aimed at their level of need. Inevitably, these class groups are by their nature broad, as in most educational settings class sizes are too large to allow attention to be given to *exactly* meeting the individual needs of all of the students in the groups. This is simply a reality, and one which the current study has tried to address head on. If academic research is to reflect the world on which it seeks to gather data, there is arguably a need for it to reflect as closely as possible the types of practices, decisions and compromises that must be made in that world. For this reason, it was felt that the use of IELTS grades as a measure of reading proficiency was justified, as it allowed the participants to be divided into the types of groupings that equate to 'real world' classroom settings. It was felt that this justified the use of a different approach to analysing the data than is typically used in this type of research, most particularly as the findings from the eye-tracking analysis were triangulated with those from the analysis of the think-aloud interviews.

This notwithstanding, as discussed in Section 3.6.1, it is important also to note that the use of IELTS scores as a measure of reading proficiency also resulted in a number of limitations. Firstly, from an accuracy point of view, it was found that the IELTS scores for many of the PhD students who participated in the study were out of date. This was because they had recently completed a master's degree in a UK university, and so were waived the need to provide an IELTS score when entering their PhD courses. Similarly, the timing of the data collection meant that many of the participants who were studying for a master's degree were coming towards the end of their courses, and so had not taken an IELTS test for approximately one year. Thus, given that they had studied at post-graduate level in English for a substantial period of time since their last IELTS

test, there is a strong likelihood that these participants were more proficient readers than their IELTS scores indicated. This should be borne in mind when considering the findings reported below.

A second limitation arises from the very fact of dividing the participants by proficiency levels at all, as the creation of such group inevitably involves setting arbitrary cut-off points. Not least, part of the judgement in setting the cut offs for the current proficiency groups was guided by the need to keep the groups roughly equal in size. However, it was also felt that the way the participants were allocated to their respective proficiency levels conformed to a large extent with the guide that IELTS provides for educational institutes to help them set entrance requirements for their courses (see Table 4.7). This suggests that the rationale for dividing the students into proficiency groups used in this study was at least similar to that used by universities and testing organisations, and thus was not wholly idiosyncratic. This notwithstanding, the fact remains that dividing participants into proficiency groups imposes a somewhat artificial framework upon them, and had the participants been divided differently this may have affected the results accordingly.

Table 4. 7: Guidance for setting entry requirements provided by IELTS (https://www.ielts.org/for-organisations/setting-ielts-entry-scores)

Band Score	Linguistically demanding academic courses	Linguistically less demanding academic courses	Linguistically demanding training courses	Linguistically less demanding training courses
7.5 – 9	Acceptable	Acceptable	Acceptable	Acceptable
7	Probably acceptable	Acceptable	Acceptable	Acceptable
6.5	English study needed	Probably acceptable	Acceptable	Acceptable
6	English study needed	English study needed	Probably acceptable	Acceptable
5.5	English study needed	English study needed	English study needed	Probably acceptable

A third important limitation to these results involves the issue of statistical power. The overall number of participants involved in this study was not sufficient to allow the three proficiency groups to be large enough

to maintain a high degree of statistical power once the overall sample had been subdivided (Field, 2013). This reduction in power potentially caused a number of problems, which should be borne in mind when reading the following results. Firstly, a reduction in power inevitably leads to an inflated risk of a Type 2 error occurring (Field, 2013); that is, the chance of not finding an effect when one does exist, and so not rejecting the null hypothesis when in fact it would have been statistically sound to have done so. However, there is also an increased risk of a Type M error occurring. According to von der Malsburg and Angele (2017, p.129), Type M errors occur when:

> an effect in the correct direction is detected but the size of the true effect is overestimated ... The error tends to be larger when power is low, [to the extent that] the size of Type M errors [is] approximately twice as big in data sets with low numbers of subjects and items than in data sets with high numbers of subject and items.

What is more, in eye-tracking data, the issue of low power is likely to be amplified still further in certain of the eye-tracking measures, with late measures (e.g. second pass reading time) being more likely to suffer from low power than early measures (e.g. first fixation duration). This is because all words which are processed by a reader will be fixated at least once, but not all words will necessarily be read more than once. In consequence, the likelihood is that there will be a lower number of data points in the data sets from the late measure, and as such, the risk of Type M errors occurring in the data for these late measures is especially acute. For this reason, the effect sizes reported below should be considered with this caveat in mind. Finally, the simple fact is that as the size of a sample reduces, so too the likelihood that it reflects the true population similarly diminishes, and consequently the generalisability of the results. As a result of these various limitations, the results from the ANOVAs reported below should be regarded with some degree of caution.

In addition to conducting this second set of 2x2 repeated measures factorial ANOVAs, **multiple regression analyses** were conducted in order to investigate the extent to which the participants' performance was predicted by their *vocabulary knowledge*, *grammar knowledge*, *working memory capacity*

in English, *working memory capacity in Chinese*. Full details of the research instruments used to generate scores for the individual differences can found in Section 3.6.

4.7.1 2x2 repeated measures factorial ANOVAs - sample regrouped by proficiency level

This section reports the results of the ANOVAs for the four eye-tracking measures. Prior to the analysis two predictions were made.

i.	Lower proficiency participants will not show a significant main effect for the **determiner factor**, while higher proficiency participants will.	*In both instances, this is because the Linguistic Threshold Hypothesis (see Section 2.3.2) suggests that readers must be above a certain level of linguistic proficiency before they can comprehend the texts they read. In this study, the ability to detect incongruencies in a text is assumed to act as a proxy for the ability to establish discourse-level representations of that text, that is, to comprehend it. As such, it is predicted that the lower proficiency readers will not identify the incongruent words. Furthermore, their relatively low level of linguistic proficiency should preclude them from using the determiner as an aide to processing.*[2]
ii.	Lower proficiency participants will not show a significant main effect for the **congruency factor**, while higher proficiency participants will	

Because no interactions were found in the ANOVAs described in Section 4.6, no interactions were predicted for in the current analysis. Rather, the question to be addressed is the extent to which lower-level readers were able to make use of determiners as an aide to initial word processing, as indicated by a significant main effect in either the first-fixation duration or gaze duration measures.

Similarly to the ANOVAs reported in Section 4.6, Bonferroni corrections were made to the alpha levels accepted for significance, following the same procedure. This again resulted in a revised alpha of ***p* = .013**.

4.7.2 Early measure: First fixation duration – 2x2 repeated measures factorial ANOVA - sample regrouped by proficiency level

Inspection of Table 4.8 reveals that for the first fixation measure, once the participants' reading proficiency was accounted for, only the upper-intermediate group showed a significant main effect of determiner, although the intermediate group almost reach significance, with a large effect size. None of the three groups showed a significant main effect of congruency , or any significant interactions.

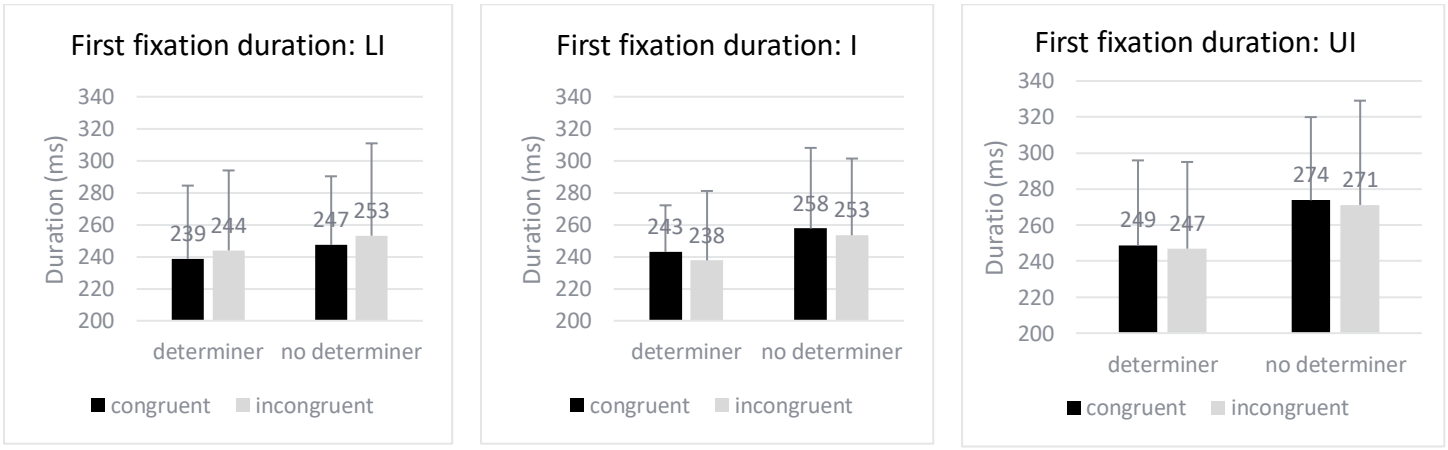
[2] The assumption that the lower-proficiency readers in this study also had a lower level of linguistic proficiency is supported by the fact that there were strong corelations between the IELTS reading scores and the results of both the vocabulary test (r = .7) and the grammar test (r = .6).

Table 4. 8: First fixation duration 2x2 repeated measures factorial ANOVA - sample regrouped by proficiency level (alpha set at p = .013)

		F	*Sig.*	η_p^2
Lower-intermediate	determiner condition	3.036	.095	0.12
	congruency condition	1.057	.315	0.04
	determiner * congruency	0.357	.556	0.02
Intermediate	determiner condition	6.972	.014	0.21
	congruency condition	0.085	.773	0.00
	determiner * congruency	0.712	.406	0.03
Upper-intermediate	determiner condition	14.341	.001	0.37
	congruency condition	0.046	.832	0.00
	determiner * congruency	0.135	.716	0.01

In the upper-intermediate group, the target nouns were processed faster when there was a determiner in the noun phrase than when there was not (see Figure 4.5).

Figure 4. 5: Mean first fixation durations on the target nouns as a function of congruency and presence of a determiner separated by proficiency (units in milliseconds) – error bars show standard deviation



4.7.3 Early measure: Gaze duration – 2x2 repeated measures factorial ANOVA - sample regrouped by proficiency level

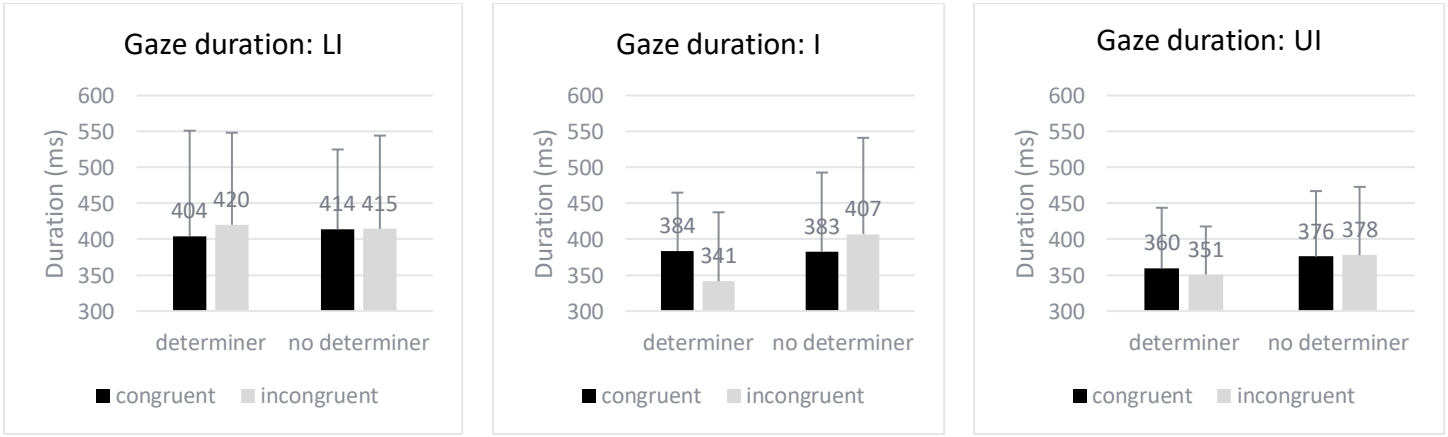
Inspection of Table 4.9 reveals that there were no significant main effects of either factor in any of the proficiency groups, nor any significant interactions. This suggests that in the upper-intermediate group the

effect of the determiner was extremely short lived, and had faded by the time of even the gaze duration measure.

Table 4. 9: Gaze duration 2x2 repeated measures factorial ANOVA - sample regrouped by proficiency level (alpha set at p = .013)

		F	*Sig.*	η_p^2
Lower-intermediate	determiner condition	1.442	.243	0.06
	congruency condition	1.037	.320	0.05
	determiner * congruency	0.593	.449	0.03
Intermediate	determiner condition	2.895	.100	0.10
	congruency condition	0.762	.390	0.03
	determiner * congruency	5.754	.024	0.18
Upper-intermediate	determiner condition	3.004	.095	0.10
	congruency condition	0.11	.743	0.00
	determiner * congruency	0.044	.835	0.00

Figure 4. 5: Mean gaze durations on the target nouns as a function of congruency and presence of a determiner separated by proficiency (units in milliseconds) – error bars show standard deviation



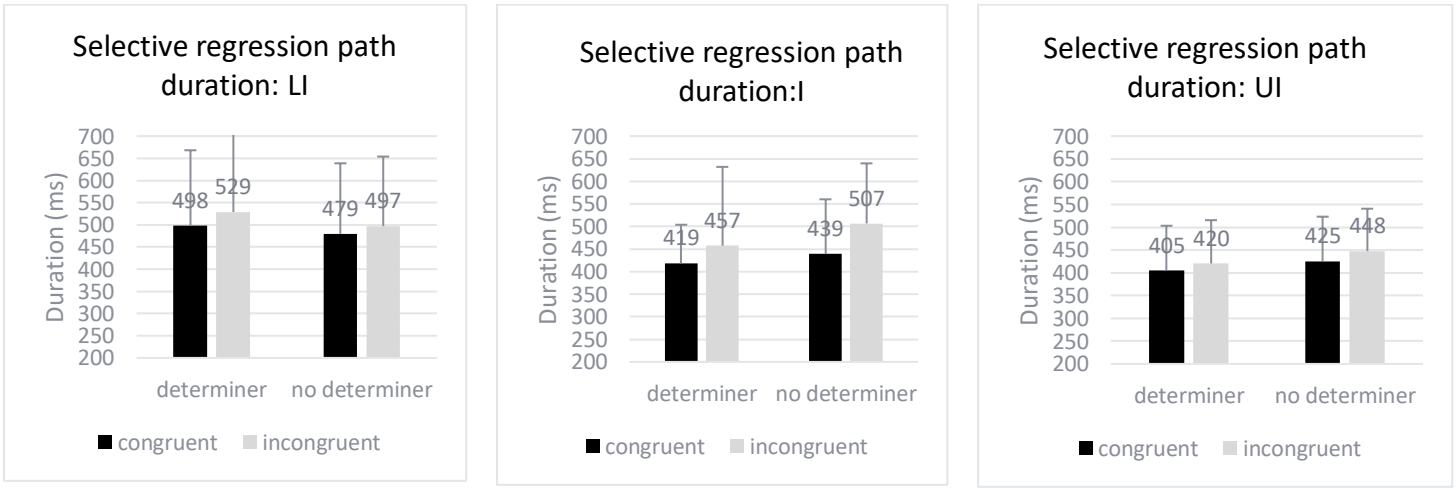
4.7.4 Late measure: Selective regression path duration – 2x2 repeated measures factorial ANOVA - sample regrouped by proficiency level

Inspection of Table 4.10 reveals that there were no significant main effects of either factor in any of the proficiency groups, nor any significant interactions for the selective regression path duration measure.

Table 4. 10: Selective regression path duration 2x2 repeated measures factorial ANOVA - sample regrouped by proficiency level (alpha set at p = .013)

		F	*Sig.*	η_p^2
Lower-intermediate	determiner condition	0.300	.589	.01
	congruency condition	0.180	.676	.01
	determiner * congruency	0.012	.912	.00
Intermediate	determiner condition	5.238	.030	.16
	congruency condition	4.802	.037	.15
	determiner * congruency	1.946	.174	.07
Upper-intermediate	determiner condition	2.902	.100	.10
	congruency condition	1.769	.195	.06
	determiner * congruency	0.039	.845	.00

Figure 4. 6: Mean selective regression path durations on the target nouns as a function of congruency and presence of a determiner separated by proficiency (units in milliseconds) – error bars show standard deviation



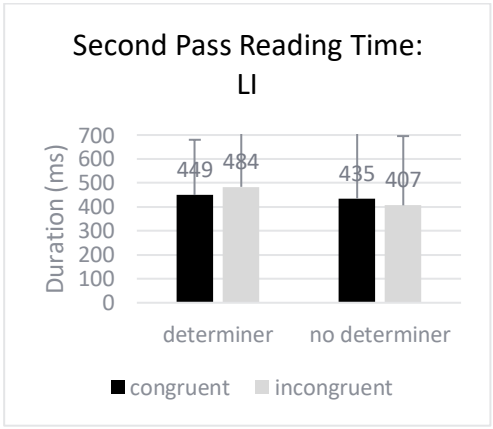
4.7.5 Late measure: Second pass reading time – 2x2repeated measures factorial ANOVA - sample regrouped by proficiency level

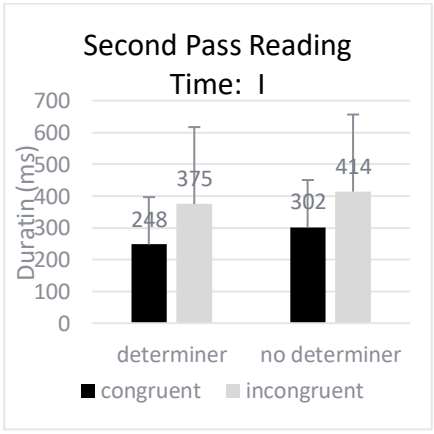
Inspection of Table 4.11 reveals that in the second reading pass measure, only the upper-intermediate group showed a significant main effect of congruency. None of the groups had a significant main effect for the determiner factor, and there were no significant interactions. The upper-intermediate group showed longer reading times when there was an incongruent word in the target noun phrase versus when there was a congruent word. However, the fact that there was no significant interaction between the determiner and congruency factors suggests that the presence of determiner in a noun phrase did not help the upper-intermediate readers to identify the incongruent words.

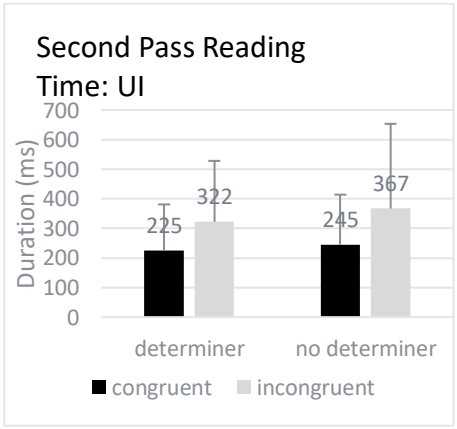
Table 4. 11: Second pass reading time 2x2 repeated measures factorial ANOVA - sample regrouped by proficiency level (alpha set at p = .013)

		F	*Sig.*	η_p^2
Lower-intermediate	determiner	0.420	.524	.02
	congruency	3.019	.096	.12
	determiner * congruency	0.381	.543	.02
Intermediate	determiner	0.937	.342	.04
	congruency	1.485	.234	.06
	determiner * congruency	0.850	.365	.03
Upper-intermediate	determiner	0.162	.692	.01
	congruency	13.836	.001	.41
	determiner * congruency	2.349	.141	.11

Figure 4. 7: Mean second pass reading times on the target nouns as a function of congruency and presence of a determiner separated by proficiency (units in milliseconds) – error bars show standard deviation







4.7.6 Summary of the 2x2 repeated measures factorial ANOVAs - sample regrouped by proficiency level

The results for the ANOVAs reported in this chapter showed a significant main effect for the determiner factor in the early first fixation duration measure, and a significant main effect for congruency factor in the late second pass reading time measure. However, the results when the proficiency level of the students was taken into account indicate that it was predominantly among upper-intermediate level readers that these effects were to be found. This suggests that there may be a linguistic threshold which native-Chinese speaking readers must pass before they can comprehend a text well enough to detect any incongruencies it may contain. The results also suggest that the while the upper-intermediate readers were able to identify the incongruent words in the text, the lack of any significant interactions between the determiner and the congruency factors indicate that the presence of a determiner in a target noun phrase did not help them to do so.

4.8 Robust multiple regressions

The results for the ANOVAs reported above revealed a number of issues regarding the three factors involved. Firstly, it was found that the *determiner* factor only had a significant main effect in the earliest eye-tracking measure, the first fixation duration. On the other hand, the *congruency* factor only had a significant main effect in the latest measure, the second pass reading time. However, once reading proficiency was factored for, the results suggested that these effects only pertained to the upper-intermediate readers. In order to investigate why this difference between the highest proficiency readers and the two lower proficiency groups may have occurred, multiple regression analyses were conducted to see which, if any, individual differences may have predicted the participants' performance. Given that the effects for the determiner factor and congruency factor were only apparent in the first fixation duration and second pass reading time respectively, the multiple regression analyses were only conducted on data from the measure where an effect was found. The dependent variables in each of the multiple regression models are shown in Table 4.12:

Table 4. 12: Dependent variables for the multiple regression analyses

	Eye-tracking measure	Dependent variable
Early measure	**First fixation duration**	*The difference in the average reading time of the target noun when there **was** a determiner in the target noun phrase versus when there was **not** a determiner in the target noun phrase*
Late measure	**Second pass reading time**	*The difference in the average reading time of the target noun when there was a **congruent** word in the target noun phrase versus when there was an **incongruent** word in the target noun phrase*

The following sections first present details of the measures of independent differences that were used, then the results of the multiple regressions are presented for these two eye-tracking measures separately.

4.9 Measures of individual differences

As mentioned above, there were four measures of individual difference used in this study:

- Vocabulary knowledge
- Grammar knowledge
- English working memory capacity
- Chinese working memory capacity

Full descriptions of all of these tests can be found in Section 3.6. Descriptive statistics for each of these measures are presented below.

4.9.1 Vocabulary knowledge

The participants' vocabulary knowledge was tested using the Vocabulary Level Test (Schmitt et al., 2001). The test is scored out of 147, with one point being awarded for each correct answer. Table 4.13 presents the mean, the standard deviation as well as the results for the Kolmogorov-Smirnov test for normality. As can be seen, the data were found to be normally distributed.

Table 4. 13: Descriptive statistics for the vocabulary test

			K-S test	
	M	*SD*	*D* (82)	*Sig.*
Vocabulary	92	22.89	.081	.200

4.9.2 Grammar knowledge

The grammar test was based on DeKeyser (2000). The test was scored out of 22. Table 4.14 presents the mean and standard deviation of the participants' scores, as well as the results of the Kolmogorov-Smirnov test. The results suggested the distribution was significantly non-normal. Attempts were made to address this departure from normality by transforming the data, but these were unsuccessful.

Table 4. 14: Descriptive statistics for the grammar test

			K-S test	
	M	*SD*	*D*(82)	*Sig.*
Grammar	9.57	4.30	.118	.006

4.9.3 English working memory capacity

The test for working memory capacity used for this study were generic digit span memory tests, scored out of 30 . Table 4.15 presents the descriptive statistics for the test when conducted in English.

Table 4. 15: Descriptive statistics for the English working memory capacity test

			K-S test	
	M	*SD*	*D*(82)	*Sig.*
English WMC	17.65	3.62	.109	.017

As can be seen, once again the distribution was not normal, and attempts were made to address this departure from normality by transforming the data were again unsuccessful.

4.9.4 Chinese working memory capacity

Table 4.16 presents the descriptive statistics for the test when conducted in Chinese. Once again, the data were found not to be normally distributed, and attempts to remedy for this using transformation were unsuccessful.

Table 4. 16: Descriptive statistics for the English working memory capacity test

			K-S test	
	M	*SD*	*D*(82)	*Sig.*

Chinese WMC	26.62	2.58	0.118	.006

4.9.5 Summary

Unfortunately, in all but one of these tests, the assumption of normality could not be met. As previously noted, the large size of the sample used in this study means that normality may be assumed due to the Central Limit Theorem (Field, 2013); however, the lack of normality in these data should nonetheless be borne in mind when considering the following results, as it may limit their reliability.

4.10 Robust multiple regression analyses: Results

In order to compensate for the non-normal distributions of the data for three of the four measures of individual differences used in this study, robust multiple regression analyses were performed (Field, 2013). In order to do this, the bootstrapping function was used in SPSS. The results are presented below in Tables 4.17 and 4.18. In each instance, the vocabulary and grammar scores were entered into the model together, as while collectively these two factors have been shown to predict around 50% of the variance in reading comprehension scores, studies vary as to which is the stronger predictor (Bernhardt, 2010). The English and Chinese working memory scores were added sequentially to see if Chinese working memory capacity accounts for any variance over and above English working memory capacity.

4.10.1 Correlations

Prior to performing the regression analyses, correlational analysis was performed on the four predictors, and each of the two dependent variables, shown in tables 4.17 and 4.18.

Table 4.17 Correlations between the four predictors and the determiner vs. no determiner factors - First fixation durations

	Determiner vs. No determiner	*Vocabulary*	*Grammar*	*English WMC*	*Chinese WMC*
Determiner *vs*. No determiner	1	-.166	-.126	.038	.114
Vocabulary	-.166	1	.534**	.202	.001
Grammar	-.126	.534**	1	.358**	.068
English WMC	.038	.202	.358**	1	.495**
Chinese WMC	.114	.001	.068	.495**	1

** Correlation is significant at the 0.01 level (2-tailed).

Table 4.18 Correlations between the four predictors and the congruent vs. incongruent factors - Second pass reading time

	Congruent vs. Incongruent	*Vocabulary*	*Grammar*	*English WMC*	*Chinese WMC*
Congruent *vs.* Incongruent	1	.072	.169	.153	-.035
Vocabulary	.072	1	.534**	.202	.001
Grammar	.169	.534**	1	.358**	.068
English WMC	.153	.202	.358**	1	.495**
Chinese WMC	-.035	.001	.068	.495**	1

** Correlation is significant at the 0.01 level (2-tailed).

Inspection of these two tables indicates that there were no significant correlations between either of the dependent variables and any of the four predictors. However, moderate correlations were found between the participants' vocabulary and grammar scores, and their English and Chinese working memory scores (WMC).

4.10.2 Multiple Regressions

Inspection of Tables 4.19 and 4.20 reveals that none of the independent variables significantly predicted variation in the difference in the reading times between conditions. Furthermore, not only was the range of scores between the lower and upper boundaries extremely large, but in every instance they also crossed zero, suggesting the true *b* could have been zero.

Table 4. 19: First fixation duration – Results of the robust multiple regression analysis used to investigate differences in the average processing time of the target noun based on the presence or absence of a determiner in the target noun phrase

Model		*b*	*SE*	95% CI		*Sig. (2-tailed)*	*Collinearity Statistics*		R^2	R^2 *Change*
				LL	*UL*		*Tolerance*	*VIF*		
1	Constant	-15.26	3.88	-23.16	-7.21	0				
	Vocabulary	-4.88	5.48	-15.61	6.17	.382	.72	1.4	.03	.03
	Grammar	-1.85	4.69	-11	7.67	.688	.72	1.4		
2	Constant	-15.26	3.9	-23.38	-7.19	0				
	Vocabulary	-4.93	5.5	-15.49	6.35	.378	.71	1.4		
	Grammar	-3.05	4.85	-12.19	5.99	.522	.65	1.54	.038	.008
	English WMC	3.43	3.59	-3.58	10.55	.328	.87	1.15		

	Constant	-15.26	3.92	-23.4	-7.02	0				
	Vocabulary	-4.74	5.57	-15.61	7.24	.399	.71	1.4		
	Grammar	-2.72	4.9	-12.04	6.52	.57	.64	1.55		
3	English WMC	1.59	4.51	-8.46	10.24	.723	.65	1.54	.045	.007
	Chinese WMC	3.41	4.65	-4.77	13.07	.455	.74	1.35		

The regression model shown in Table 4.19 , with vocabulary, grammar, English WMC and Chinese WMC entered as predictors, explained only 4.5% of the total variance, [F (1, 77) = 0.902, p = .467] in the difference between the average reading times of the target nouns when there either was or was not a determiner in the target noun phrase. Similarly, the regression model shown in Table 4.20, also with vocabulary, grammar, English WMC and Chinese WMC entered as predictors, explained just 5.1% of the total variance, [F (1, 77) = 1.045, p = .390], in the difference in reading times of the target nouns when there was either a congruent or an incongruent adjective in the target noun phrase. It can also be seen from both tables that in every step in the model, the R^2 change was extremely small, suggesting none of the predictors had any substantial effect on the outcome. As such, none of the measures of individual differences can be said to have predicted the participants' reading times of the target noun in respect of either the determiner or the congruency factors.

Table 4. 20: Second pass reading time – Results of the robust multiple regression analysis used to investigate differences in the average processing time of the target noun when there was either a congruent or an incongruent word in the target noun phrase

				95% CI			*Collinearity Statistics*			
Model		*b*	*SE*	*LL*	*UL*	*Sig. (2-tailed)*	*Tolerance*	*VIF*	R^2	R^2 *Change*
	Constant	1664.73	385.24	937.78	2331.82	0				
1	Vocabulary	-87.75	500.22	-1149.77	868.82	.865	.72	1.4	.007	.007
	Grammar	644.15	539.19	-326.31	1651.38	.24	.72	1.4		
	Constant	1664.73	389.54	914.49	2360.91	0				
2	Vocabulary	-93.29	504.37	-1161.36	853.03	.862	.71	1.4	.012	.005
	Grammar	512.57	548.78	-508.6	1558.3	.351	.65	1.54		
	English WMC	375.91	416.12	-463.54	1136.77	.359	.87	1.15		
	Constant	1664.73	390.1	892.43	2430.66	0				
3	Vocabulary	-120.1	501.08	-1265.51	903.8	.816	.71	1.4	.013	.001
	Grammar	468.21	567.59	-554.9	1485.29	.409	.64	1.55		

English WMC	628.15	477.7	-388.74	1572.45	.18	.65	1.54
Chinese WMC	-466.59	549.41	-1453.12	497.73	.397	.74	1.35

4.11 Research question 3: Summary

The analysis reported in this section sought to address the third research question posed at the start of this chapter:

RQ3: To what extent do native-Chinese speaking readers at different levels of reading proficiency differ in their ability to identify incongruent words in coreferential noun phrases, or in the extent to which they are helped to identify incongruent words in such a noun phrase by the presence of a determiner?

The results suggest that there were indeed differences between the proficiency levels, with only the upper-intermediate readers able to identify the incongruent words in the text. However, once again the results suggest that they were not aided in so doing by the presence of a determiner. As such, it would appear that the second assumption which underpinned this research design, namely that *the presence of a determiner in a coreferential noun phrase ought to help a reader integrate that noun phrase with its antecedent, and so to identify any incongruent words it may contain*, is not correct. Rather, the results suggest that the role of a determiner in a coreferential noun phrase is limited to aiding initial word access processing of the main noun in the phrase, with determiners playing little to no role in later integration processing.

Interestingly, the results of the multiple regression analyses did not identify any individual differences which might account for this variance in the extent to which participants at different levels of reading proficiency were, or were not, able to identify the incongruent words in the text. Consequently, we still do not know *why* these differences occurred. It is with addressing this question that the following chapter is concerned.

Chapter 5: Results – The think-aloud interviews

5.1 Introduction

The central aim of this thesis has been to investigate the extent to which native-Chinese speaking readers of academic English are able to integrate coreferential noun phrases, that is two or more noun phrases which refer to the same idea or entity within a text, as they seek to establish discourse-level representations of the texts they read. One of the key assumptions which underpinned the design of this study, drawing on Walter (2007), was that that the ability to identify an incongruent word in a text can be taken as a proxy for the ability to integrate coreferential noun phrases in successive sections of text in order to establish a discourse-level representation of the text's overall meaning. This is because if a reader is able to integrate the meaning of two coreferential noun phrases, then it should be obvious to them if one of the words in the second noun phrase is incongruent with an idea expressed in the first. The eye-tracking study described in the previous chapter sought firstly to investigate the extent to which the participants were able to identify incongruent words embedded in a text, and secondly to what extent the presence of a grammatical determiner (e.g. *this*, *those*) might help them to do so. As explain in Section 4.7, the participants in the study were divided into three proficiency levels, according to their IELTS reading scores, and placed in either the lower-intermediate, the intermediate or the upper-intermediate group. The results from the eye-tracking results yielded three key findings:

i) Neither the lower-intermediate nor the intermediate readers reacted to the incongruent words embedded in the text (i.e. they did not experience processing disruption as indicated by longer reading times of the target noun). This suggest that they were unable to monitor their comprehension well enough to identify incongruent words in a text.

ii) The upper-intermediate readers did react to the incongruent words, displaying longer reading times in the second pass reading time measure. This indicates that they did identify the incongruent words, and so experienced difficulty integrating the target nouns into their developing discourse-level representation of the text. This suggests they were able to monitor their comprehension well enough to identify incongruent words in a text.

iii) Although the upper-intermediate readers did react to the incongruent words in the text, contrary to expectations, the presence of a determiner in a target noun phrase did not aid them in doing so, as indicated by the lack of any significant interaction between the determiner and congruency factors. This suggests that although they were able to monitor their comprehension well enough to identify incongruent words in a text, the presence of a determiner played little to no role in allowing them to do so.

These findings suggest that clear differences exist between how native-Chinese speaking readers at different levels of proficiency integrate coreferential noun phrases in English language academic texts. However, they do not tell us *why* these differences exist. This chapter reports the findings of a think-aloud study that was conducted in order to address one more key research question that arose from the results of the eye-tracking study. In Section 3.3.3, it was explained that part of the role of the think-aloud interviews was to explore the findings from the eye-tracking experiment in more depth. Given that it was not possible to know what would be found from the eye-tracking experiment, when the fourth research question was posed at the end of the literature review, it was left deliberately vague:

RQ4.How might any differences in the ability of students of different proficiency levels to identify incongruent words be explained by a qualitative analysis of their reading behaviour

Prior to the think-aloud interviews being analysed, the question was refined in the light of the results from the eye-tracking experiment. This allowed the analysis to focus on the most important question which came out of the eye-tracking:

RQ4. Why were the upper-intermediate readers able to identify the incongruent noun phrases while the intermediate and lower-intermediate readers were not?

5.2 The think-aloud study: Procedure

In order to conduct the think-aloud study, twenty-four participants who took part in the eye-tracking session were invited attend an interview between seven and ten days later. The interview was conducted in two stages. In the first part, the participants were asked to read the text *Cardiovascular Disease: A tale of good news and bad* (see Appendix A for a full version of the text). This was one of the four texts they had

read during the eye-tracking experiment. The participants saw the same version of the text they had read during the eye-tracking (see Section 3.5 for details of the counterbalancing of the text). While they were reading, they were asked to voice their thoughts in respect of four main points:

i)	How they thought the sentences in the text were related to each other	*used to investigate if they had integrated information in the target noun phrases*
ii)	How they knew the sentences in the text were related to each other	*used to investigate what knowledge resources they drew upon if they integrated information in the target noun phrases*
iii)	What, if any, comprehension difficulties they experienced as they read	*use to investigate what, if any, comprehension difficulties they faced*
iv)	What, if any, comprehension regulations strategies they used to overcome any such difficulties	*used to investigate what, if any, comprehension regulation strategies they favoured*

In part two of the interview, the participants did an anaphor resolution task (i.e. were tested on their ability to identify an anaphor's antecedent) in which they were asked the following questions about the eight coreferential target noun phrases that were embedded in the text (see Section 3.5 for full details of the target noun phrases):

i)	**What does this word (the main noun in the target noun phrase) refer to?**	*used to test the participant's ability to integrate the noun phrase with its antecedent*
ii)	**How do you know this?**	*used to explore what knowledge resources the participant drew upon to integrate the noun phrase with its antecedent*
iii)	**What does the word mean in Chinese?**	*used to test if the participant had correctly understood the word*
iv)	**What adjective did the writer use to describe this word?**	*used to draw the participant's attention to the incongruent word*
v)	**Why do you think they chose this adjective?**	*used to encourage the participant to focus on the meaning of the incongruent word*

The aim of the second part of the interview was primarily to ensure that data could be gathered on the higher proficiency readers in the study. This is because it has been found that think-aloud protocols are most suitable for gathering data on readers who are experiencing processing difficulties, as it is during these circumstances that their thought processes are most salient to them, and when processing ceases to be automatic, and conscious strategy use comes into play and hence is available for introspection (Bowles, 2010; Güss, 2018). It was hoped that in the event that the more advanced readers did not say enough during the first part of the interview to allow for analysis, their answers in the second part would provide additional data. However, in hindsight these fears proved to be unfounded. All of the participants were able to verbalise their thoughts as they read during the initial part of the interview, at least to some degree. Furthermore, the answers given to the questions in the second part provided additional insights for readers in all three proficiency groups. Although the participants were asked about all eight of the target noun phrases in the text, only the four noun phrases which contained an incongruent word were analysed, as the intention was to investigate what had caused differences between the different proficiency groups in terms of how they monitored their comprehension in respect of these items. The participants were asked about all eight target noun phrases for fear of making the presence of the incongruent words too salient to them if only those noun phrases were presented to them. All of the interviews were video recorded, with the participants' consent, so that analysis could be conducted not only on what they said during the interviews, but also on where they pointed at the text while they were reading.

5.3 The think-aloud study: Participants, transcriptions, and translation

All twenty-four of the interviews were transcribed either by hand or using an online transcription tool (https://www.iflyrec.com). Of those that were transcribed by hand, the majority were transcribed by me, a native-English speaker, while a sample (33%) were also transcribed by a native-Chinese speaker. The second transcriber was employed mainly because the task of transcribing in Chinese turned out to be too time-consuming for one person to complete alone. Both sets of transcriptions produced similar results. The

interviews which were transcribed using the online tool were subsequently checked by hand and corrections were made where needed to ensure accuracy. After the initial transcriptions were complete, each one was examined again and annotated to see where the participants had pointed at particular places in the text while they were reading. For example, if a participant said something along the lines of "I don't understand this word", or "I think this word means the same as this", the transcript was annotated to reflect which word they were pointing to as they said it. Initially, it was foreseen that all twenty-four of the transcripts would be analysed. However, during analysis it was found that the time available did not allow them all to be analysed in the depth required. Consequently, transcripts from only twelve of the participants were selected for inclusion in the final sample. In order to make sure the sample was as balanced as possible, four participants were selected from each of the three proficiency groups. Furthermore, from within each proficiency group transcripts were selected so that all of the IELTS scores in that group were represented. Details of this final sample are presented in Table 5.1:

Table 5. 1: Details of the twelve participants included in the final sample chosen for analysis

	Participant	IELTS reading score	Vocabulary score	Grammar score
Lower-intermediate group	Participant 017	6	60.54%	31.82%
	Participant 013	5.5	39.46%	36.36%
	Participant 032	5.5	42.86%	36.36%
	Participant 025	5	19.73%	31.82%
Intermediate group	Participant 071	7	72.79%	68.18%
	Participant 060	6.5	75.51%	45.45%
	Participant 004	6.5	55.10%	31.82%
	Participant 033	6.5	55.78%	50.00%
Upper-intermediate group	Participant 056	9	77.55%	81.82%
	Participant 029	8.5	87.07%	54.55%
	Participant 006	8	77.55%	72.73%
	Participant 009	7.5	83.67%	77.27%

During the think-aloud interviews, participants were told they could speak English, Chinese or a mixture of both. This was to enable them to express their thoughts in whichever language they found easiest, in the

hope that this would result in more accurate, detailed responses (White et al., 2007). Initially, it was planned to translate the transcripts into English, where necessary, prior to analysis. However, it quickly became apparent that this was impractical. This was because where the participants did use Chinese, they typically engaged in a high degree of code switching, which often resulted in utterances in a kind of interlanguage formed by mixing Chinese and English vocabulary and grammar. This was particularly the case when the lower-level participants were reading sentences they found hard to process. Although these mixed Chinese and English sentences were not hard to understand when I was reading them, it was not always clear how they should be translated as they contained grammatical inaccuracies in both languages. This meant that these utterances were difficult to translate in a way that was both comprehensible and true to what had actually been said. This made translation an extremely cumbersome and time-consuming process. As such, it was decided that, since the transcripts were easily comprehensible in their original format, it was best to analyse them untranslated, as this provided the most accurate view on what had in fact been said. However, where excerpts from the transcripts are presented below, those excerpts have been translated to aid the reader. The original, untranslated excerpts along with the translation are provided in Appendix G. It has been noted already that the participants frequently engaged in code switching both as they read the text aloud, and as they voiced their thoughts. In order to show where the participants switched languages, utterances which were initially made in Chinese are shown in italics while those made in English are in standard font.

5.4 Think-aloud: Findings

The following section reports the findings from the analysis of the think-aloud protocols. Firstly, a brief review is presented of the target noun phrases used in the texts. Secondly, the results are presented of a *quantitative* analysis designed to see how closely the findings from the think-aloud interviews aligned with those from the eye-tracking experiment. Finally, a detailed *qualitative* analysis of the participants' transcripts is presented, with the aim of elucidating what caused differences in their ability to identify the incongruent words in the text.

5.4.1 An overview of the research design

As explained in Section 3.3.3, the research design for this study used a complementary mixed methods approach, in which the think-aloud interviews were intended to both triangulate and elaborate on the data from the eye-tracking experiment. As explained below, this meant that the transcripts from the think-aloud interviews were subjected to both *quantitative* and *qualitative* analysis. It has been argued throughout this thesis that the ability to identify an incongruent word in a text can be taken as a proxy for the ability to integrate coreferential noun phrases with their antecedents. In order to test that assumption, a number of incongruent words were embedded in coreferential noun phrases in a text, so that the target noun phrases contradicted information stated in their antecedents. The first example of such an incongruent noun phrase appeared in the opening two lines of the article:

> (***i***) Deaths from *cardiovascular disease (CVD), which includes heart attacks and high blood pressure*, first began to decrease in the 1960s in Western countries, and have continued to do so ever since. (***ii***) In the United States, for example, statistics show that death rates from **these brain diseases** fell by over 35 percent between 1980 and 1997.

Here it can be seen that in sentence (***i***) the antecedent noun phrase (shown in italics) refers to *cardiovascular disease*, which is specifically defined as including *heart attacks* and *high blood pressure*. In sentence (***ii***), the target coreferential noun phrase (shown in bold and underlined) *these brain diseases* is incongruent with its antecedent, as it is referring to a different type of disease. The first analysis conducted on the think-aloud transcripts investigated two issues, and did so from a quantitative perspective. Firstly, did the participants display clear evidence of having integrated the main noun in the target noun phrase, in this case the word *disease*, with the antecedent noun phrase, in this case either *cardiovascular disease*, *heart attacks* or *high blood pressure*? Secondly, did the participants clearly identify the incongruent word in the target noun phrase, in this case the word *brain*? A limitation of think-aloud protocols as a research method is that the researcher only knows what a participant states aloud (Dörnyei, 2007). As such, it was not always possible to precisely determine if a participant did in fact identify an incongruent word, or integrate a target noun with its antecedent. With this is mind, only those instances where a participant displayed clear

evidence of having done so were used for analysis. When judging if a reader had integrated the main noun in a target noun phrase with its antecedent, two possible criteria were used. Firstly, there were occasions when a reader explicitly stated that the relevant noun phrases were coreferential. A good example of this was found in the transcript of Participant 029, from the upper-intermediate group, when she read the incongruent noun phrase (highlighted in bold and underlined) in the following sentence from the second paragraph of the text:

> Firstly, in spite of **worsening techniques** for treating heart related problems, heart disease is still the most common cause of death in most Western countries.

As can be seen from her transcript below, when reading this item, she explicitly stated that in the first paragraph the text had talked about improving treatments, thereby indicating that she had integrated the target noun phrase with its antecedent (relevant utterances highlighted in bold):

> **P029**: so this part see, so this is something I don't understand, it's worsening, so it means like bad, so it means, despite bad techniques for treating heart related problems, so **because the first paragraph they talk about, like say the treatment has improved**, **but here it talks about worsening techniques**...

Secondly, there were occasions when a reader pointed to the antecedent while they read the target noun phrase. For example, when Participant 004 from the lower-intermediate read sentence (*ii*) from the passage above, she said:

> **P004**: In the United States, for example, statistics show that death rates from these brain diseases fell [indistinct while reading] it just said an example that shows that this disease [**points to cardiovascular disease in sentence (*i*)]** in the United states, United States, and show some statistics to say how the, this disease happens in a period...

Here it can be seen that in explaining the meaning of sentence (***ii***), she pointed at the antecedent noun phrase in sentence (***i***), indicating that she had integrated the information in the two sentences. In order for a participant to be judged as having successfully integrated a target noun phrase with its antecedent, they needed to fulfil at least one of these criteria.

Similarly, when judging if a reader identified the incongruent word in one of the target noun phrases, they needed to display clear evidence of having done so. Again, two criteria were used. Firstly, there were

occasions when a participant explicitly stated that the word was incongruent. For example, when Participant 056, from the upper-intermediate group, read the incongruent noun phrase in the sentence:

> So, what are the factors that could have brought about **these discouraging reductions** in the rates of heart disease?

she explicitly stated that she that she was confused, and thought there was an error in the text, saying:

> **P056**: **… I am a little bit confused** about discouraging, because I think reduction in the rate of heart disease is something good for human beings, it is positive phenomenon, but why here it mention discouraging… **I think maybe it's an error or mistake that the writer has made**…

Secondly, there were occasions where although a reader did not explicitly state that there was an incongruent word in the target noun phrase, they did return to the word multiple times in way that suggested that they were experiencing difficulty integrating it into their representation of the text. For example, when Participant 032, from the lower-intermediate group, read the incongruent noun phrase in the sentence:

> What's more, for the many patients who in the past have already suffered **language problems** the news is also good

she returned to the phrase *language problems* several times, emphasising it as she did so in a tone that suggested confusion (utterances shown in italics were originally made in Chinese):

> **P032**: What's more, *in addition* for the many patients who in the past have already [indistinct while reads], *some additional patients*, *they*, *some past patients already*, suffered language problems, *already suffered this kind of language problem*, the news is good [indistinct], suffered a language problem, the news *is also good*…

In order for a participant to be judged as having identified the incongruent word in a target noun phrase they needed to have fulfilled at least one of these two criteria. The obvious limitation to this approach is that it is possible that a reader did integrate a target noun with its antecedent, or identify one of the incongruent words, without displaying evidence of having done so. As such, the findings presented below should be interpreted as being only indicative of how the participants processed the target noun phrases. However, in all of the transcripts there were instances where readers clearly did integrate at least some of the target noun phrases in way that contrasted with the way they read those noun phrases where no such evidence

was forthcoming. Furthermore, while not all of the participants identified incongruencies in the text, in those instances when a participant did noticeably encounter difficulties with an incongruent word, it was extremely evident that they had done so. As such, it is felt that these results do provide an accurate, albeit potentially incomplete, picture of how the participants processed the target noun phrases in the text.

Using these criteria, the transcripts of each of the participants were analysed to investigate whether the participants integrated the target noun phrases with their antecedents or identified the incongruencies embedded in the text. The results of this analysis are presented below in tabulated form. Results for the three proficiency groups are reported separately so that relevant differences may be highlighted.

5.4.2 The lower-intermediate group

It has been noted the results from the eye-tracking experiment suggested that the lower-intermediate group did not react to the incongruent words in the text. The quantitative analysis of the think-aloud interview supports this finding. Inspection of Table 5.2 indicates that in the first part of the interview, where the participants were reading the text aloud, only one of the participants, Participant 032, identified any of the incongruent words in the text, and she only identified one of the four. In Part two of the interview, when the participants were specifically asked about the target noun phrases, only one more, Participant, 017, was able to identify an incongruent word, and he also only noticed one. However, although for the most part the lower-intermediate readers did not identify the incongruent words in the text, inspection of Table 5.2 also shows that, with the exception of Participant 025, all of the participants *did* display clear evidence of having integrated at least two of the four target noun phrases with the antecedent during initial reading, with Participant 017 clearly integrating three of them. Furthermore, in the second part of the interview, the participants were markedly more successful at integrating the noun phrases, with even Participant 025 now showing clear evidence of having integrated three out of the four. This suggests that although the readers in the lower-intermediate group were unable to monitor their comprehension well enough to identify the

incongruent words, this was not, or at least not solely, because they were unable to integrate the target noun phrases with their antecedents.

Table 5. 2: Findings for the lower-intermediate group regarding noun phrase integration and incongruency detection

Lower-intermediate group					
	Incongruent noun phrase	**Incongruent word identified in part 1 of the interview**	**Evidence of noun phrase integration in part 1 of the interview**	**Incongruent word identified in part 2 of the interview**	**Evidence of noun phrase integration in part 2 of the interview**
Participant 017	*these brain diseases*	no	unclear	**yes**	**yes**
IELTS: 6	*mild illness*	no (word unknown)	**yes**	no (word unknown)	**yes**
Vocab: 60.5%	*worsening techniques*	no (word unknown)	**yes**	no (word unknown)	**yes**
Grammar: 31.8%	*this minor disease*	no (word unfamiliar)	**yes**	no (word unfamiliar)	**yes**
Participant 013	*these brain diseases*	no	no	no	**yes**
IELTS: 5.5	*mild illness*	no (word unknown)	**yes**	no (word unknown)	no
Vocab: 39.5%	*worsening techniques*	no	no	no	partial
Grammar: 36.4%	*this minor disease*	no (word unknown)	**yes**	no (word unknown)	**yes**
Participant 032	*these discouraging reductions*	no (word unknown)	no	no (word unknown)	**yes**
IELTS: 5.5	*language problems*	**yes**	**yes**	yes	**yes**
Vocab: 42.9%	*harmless behaviours*	no	**yes**	no	**yes**
Grammar: 36.4%	*such simplistic treatment*	no	no	no	**yes**
Participant 025	*these brain diseases*	no	no	no	**yes**
IELTS: 5	*mild illness*	no (word unknown)	no	no (word unknown)	**yes**
Vocab: 19.7%	*worsening techniques*	no (word unknown)	no	no (word unknown)	no
Grammar: 31.8%	*this minor disease*	no (word unknown)	unclear	no (word unknown)	**yes**

5.4.3 The intermediate group

In general, the quantitative findings for the intermediate group from the think-aloud protocols also aligned closely with those from the eye-tracking experiment. In this group, only two of the four participants displayed clear evidence of having identified an incongruent word in the text, and both of these participants only identified one each. In Part 2 of the interview, although Participant 004 was able to identify three of the

Table 5. 3: Findings for the intermediate group regarding noun phrase integration and incongruency detection

Intermediate group	**Incongruent noun phrase**	**Incongruent word identified in part 1 of the interview**	**Evidence of noun phrase integration in part 1 of the interview**	**Incongruent word identified in part 2 of the interview**	**Evidence of noun phrase integration in part 2 of the interview**
Participant 071 IELTS: 7 Vocab: 72.8% Grammar: 68.2%	*discouraging reductions*	**yes**	**yes**	**yes**	unclear
	such language problems	unclear	no	no	**yes**
	these harmless behaviours	no	no	**yes**	**yes**
	simplistic treatment	no (word unfamiliar)	**yes**	no	**yes**
Participant 060 IELTS: 6.5 Vocab: 75.5% Grammar: 45.5%	*discouraging reductions*	no (word unknown)	unclear	no	unclear
	such language problems	**yes**	**yes**	**yes**	**yes**
	these harmless behaviours	no	**yes**	no	**yes**
	simplistic treatment	no	no	no	**yes**
Participant 004 IELTS: 6.5 Vocab: 55.1% Grammar: 31.8%	*these brain diseases*	no	**yes**	no	**yes**
	mild illness	no	**yes**	**yes**	**yes**
	worsening techniques	no	**yes**	**yes**	**yes**
	this minor disease	no (word unfamiliar)	**yes**	**yes**	**yes**
Participant 033 IELTS: 6.5 Vocab: 55.8% Grammar: 50%	*these brain diseases*	no	**yes**	no	**yes**
	mild illness	no	unclear	no	partial
	worsening techniques	no (word unknown)	**yes**	no	**yes**
	this minor disease	no	unclear	no	**yes**

four incongruencies, the other participants were markedly less successful, with Participant 071 identifying two, Participant 060 only identifying one, and Participant 033 not identifying any at all. However, as inspection of Table 5.3 also makes clear, in this group *all* of the participants showed direct evidence of having integrated at least 50% of the target noun phrases with their antecedents during the first reading, and were almost entirely successful at doing so during the second part of the interview.

5.4.4 The upper-intermediate group

Finally, inspection of Table 5.4 indicates that in the think-aloud interviews the upper-intermediate readers were once again markedly more successful in identifying the incongruent words than either of the two lower proficiency groups. There were ten occasions where a reader displayed clear evidence of having

identified an incongruent word, and in one of the instances where a reader (Participant 009) did not identify an incongruency, this occurred because she misread the word *mild* as the word *mind* during her initial reading of the text. In Part 2 of the interview, she read the word correctly, and immediately noticed that it was incongruent with the previous text. As such, for the upper-intermediate group there was also a high degree of conformity between the results from the eye-tracking experiment and the think-aloud interviews.

Table 5. 4: Findings for the upper-intermediate group regarding noun phrase integration and incongruency detection

Upper-intermediate group

	Incongruent noun phrase	**Incongruent word identified in part 1 of the interview**	**Evidence of noun phrase integration in part 1 of the interview**	**Incongruent word identified in part 2 of the interview**	**Evidence of noun phrase integration in part 2 of the interview**
Participant 056	*these discouraging reductions*	**yes**	**yes**	**yes**	**yes**
IELTS: 9	*language problems*	**yes**	**yes**	**yes**	**yes**
Vocab: 77.6%	*harmless behaviours*	**yes**	**yes**	**yes**	**yes**
Grammar: 81.8%	*simplistic treatment*	no	**yes**	**yes**	**yes**
Participant 029	*these brain diseases*	no	no	**yes**	unclear
	mild illness	no	**yes**	no	**yes**
IELTS: 8.5	*worsening techniques*	**yes**	**yes**	**yes**	**yes**
Vocab: 87.1%	*this minor disease*	no	**yes**	**yes**	**yes**
Grammar: 54.5%					
Participant 006	*discouraging reductions*	no (word unfamiliar)	**yes**	no	**yes**
IELTS: 8	*such language problems*	**yes**	**yes**	**yes**	**yes**
Vocab: 77.6%	*these harmless behaviours*	**yes**	**yes**	**yes**	**yes**
Grammar: 72.7	*simplistic treatment*	**yes**	**yes**	**yes**	**yes**
Participant 009	*these brain diseases*	**yes**	**yes**	**yes**	**yes**
IELTS: 7.5	*mild illness*	no (word misread)	**yes**	**yes**	**yes**
Vocab: 83.7	*worsening techniques*	**yes**	**yes**	**yes**	**yes**
Grammar: 77.3%	*this minor disease*	**yes**	**yes**	**yes**	**yes**

However, despite the fact that in this group the participants were overwhelmingly successful at integrating the target noun phrases with their antecedents, there were still five occasions where a reader did not identify one of the incongruent words embedded in the text. Notably, the highest proficiency reader in the sample, Participant 056 did not identify one of the incongruent words, and Participant 029, the second highest proficiency reader in the sample, failed to identify fully three of the four.

5.4.5 Interim conclusions from the quantitative analysis

The findings from the think-aloud analysis presented thus far support the results from the eye-tracking experiment, at least in general terms. Overall, the participants from the upper-intermediate group were distinctly more successful than the participants from the two lower-proficiency groups at identifying the incongruent words in the text. However, the findings also suggest that while this was the general trend, it was not an absolute rule. There were three instances when readers from either the intermediate or lower-intermediate groups did identify an incongruent word, and five occasions when one of the upper-intermediate readers did not. This suggests that proficiency level is not the sole, or even necessarily the determinant, factor which influences a reader's ability to identify an incongruent word in a text. However, the question still remains as to what other factors may be at play. Furthermore, these results also call into question the first of the key assumptions which underpinned the research design of this project:

> **Assumption 1**: A reader's ability to integrate a coreferential noun phrase with its antecedent, in order to form discourse-level representations, can be tested by investigating their ability to identify a word in a coreferential noun phrase which is incongruent with an idea expressed in the antecedent.

The quantitative think-aloud analysis suggested that although the ability to identify the incongruent words in the text was confined to the upper-intermediate readers, participants at all proficiency levels enjoyed at least some success integrating the target nouns with their antecedents during initial reading, and all of them were overwhelmingly successful at doing so when engaged in the anaphor resolution task during the second part of the interview. As such, these findings indicated scant direct correspondence between the ability to identify an incongruent word in one of the target noun phrases and the ability to integrate the

main noun in that phrase with the antecedent, at least in absolute terms. Consequently, given that the ability to integrate a coreferential noun phrase with its antecedent does not, of itself, confer the ability to identify an incongruency which exists between them, the question remains as to what does. While the findings reported so far resulted from quantitative analysis of the data, addressing the issue of *why* differences existed between the participants' ability to identify the incongruent words in the text required a secondary round of qualitative analysis to be conducted. Details of that analysis, including the coding frameworks used, as well as the findings, are presented below.

5.5 A qualitative analysis of why participants either did, or did not, identify the incongruent words in the text

As described in Section 3.3.3, drawing on Walter (2007) the research design for this project used an Anomaly Detection Paradigm to investigate the extent to which native-Chinese readers are able to establish coherent discourse-level representations of English language academic texts. As explained throughout, the main focus was on exploring the ability of Chinese readers to integrate complex coreferential noun phrases with their antecedents. However, the quantitative findings presented above suggest there may be some tension between the ability to integrate coreferential nouns and the ability to identify incongruent words in a text. Although the above analysis revealed that there were some differences between the participants regarding their ability to integrate the main nouns in the target noun phrases, in general they all displayed some ability to do so, and were overwhelmingly successful at doing so during the anaphor resolution task. The primary difference which separated the participants was their ability to identify the incongruent words in the text. As such, for reasons of space, the main focus of this section will be on exploring why this was the case rather than on the issue of noun-phrase integration per se; the issue of what facilitated or impeded noun phrase integration will only be directly addressed where it relevant to this primary aim, although more will be said of the relationship between these two issues in the Discussion Chapter. Furthermore, the findings from the quantitative analysis reported above also suggested that although the ability to identity the incongruent words was primarily confined to the upper-intermediate readers, there were instances

where lower-proficiency readers did identify one of the incongruent words, and equally where one of the upper-proficiency readers did not. As such, in the analysis which follows the issue of proficiency will be briefly set aside as an organising factor, so that the focus of investigation may fall on the two key situations uncovered:

Condition 1: *The incongruent word was identified*

Condition 2: *The incongruent word was not identified*

5.5.1 The coding frameworks used in the qualitative analysis and how they were applied

As explained in Section 3.7, when analysing the transcripts from the think-aloud interviews qualitatively, two coding frameworks were used. The first sought to identify if the readers had encountered any processing difficulty, as indicated by the use of any comprehension regulation strategies. The second was used to explore the types of mental representations they were able to establish, defined by reference to Kintsch's (1998) Construction-Integration model of reading comprehension. Details of the codes that were used, as well as how they were applied, are presented below.

5.5.1.1 Difficulty processing the language of the text

Part of the analysis was undertaken to investigate why the participants either were, or were not, able to identify the incongruent words, by exploring the extent to which they had encountered difficulty as they read the text. This was done because previous research into both comprehension monitoring and coreferential noun phrase integration has found a link between both L2 reading and language proficiency (Hessel et al., 2021; Pretorius, 2005; Walter, 2004, 2007; Yapp et al., 2021). As such, it was felt important to investigate potential online processing difficulties the participants may have experienced. When judging whether or not a reader had encountered reading difficulty, the basic criterion used was whether or not they had employed some kind of comprehension regulation strategy. This was based on the assumption that readers are unlikely to divert attention and other cognitive resources to comprehension regulation unless they encounter difficulty of some sort. Where the analysis of a reader's transcript did not indicate that they

had employed any comprehension regulation strategies, and they were able to provide an accurate explanation of the meaning of a sentence or group of sentence, they were judged not to have experienced processing difficulty, and vice versa. Full details of the coding framework developed to identify use of comprehension regulation strategies, as well as how it was compiled and trialled are given in section 3.7.3.3.

One clear limitation of using the identification of explicit comprehension regulation strategy use as a proxy for processing difficulty is that it is possible that readers experienced difficulty but did not engage in comprehension regulation. Indeed, such a situation would be consistent with the literature, which has found that lower proficiency readers in particular frequently do not engage in comprehension regulation when they encounter difficulty (García-Rodicio & Sánchez, 2014; Kinnunen & Vauras, 1995). However, in practice it was found that since the participants were required to talk about their comprehension of the text, it quickly became clear if they understood it or not. Perhaps as a result, when they did not understand the text, the participants always visibly engaged in some degree of comprehension regulation, even if that meant simply declaring that they did not understand something before reading on in the text. It may be that this requirement to explain their comprehension of the text in this way resulted in them employing different regulation strategies than would otherwise have been the case. However, given that the aim was to identify processing difficulty, rather than to investigate comprehension regulation in depth, it is felt that this would not have negatively affected the findings from this study.

5.5.1.2 Examples of how comprehension strategy use was identified

In some instances, judging if a reader had engaged in comprehension regulation, and what strategy they used, was done on the basis of something they explicitly reported. In other instances, although a reader did not specifically state that they had used a given strategy, analysis of what they said allowed for a judgement to be made of the strategy they had used. A clear example of both these cases can be shown by the analysis of how Participant 025, from the lower-intermediate group, read the second sentence of the first passage of the text:

(*ii*) In the United States, for example, statistics show that death rates from **these brain diseases** fell by over 35 percent between 1980 and 1997

Analysis of the following excerpt from his transcripts showed that he had engaged in three comprehension regulation strategies.

P025: In the United States *this is in America*, for example, such as [indistinct] *I don't know this word* [points to 'statistics' in sentence (*ii*)] I don't know this meaning, but ***this word*** **[points to 'statistics' in sentence (*ii*)]** ***is the subject of this sentence … this is a verb*** **[points to 'show' in sentence (*ii*)],** ***this is followed by a*** **that** ***attributive clause****, … this is* death rates from this brain disease fell by, *this* death rate *is* the advantage of the death from this, *that is, this brain disease, so* ***I infer this*** [points at 'statistics' in sentence (*ii*)] *probably refers to an article or a report*

R: Why do you think that?

P025: *This is the subject* [points to 'statistics' in sentence (*ii*)] *the bit after this says... the bit after shows this brain disease's death rate, death rate, these years, exceeded 35% … in 1980 and 1997, it also says, these clauses after, how can I put this, this bit after, this is* that *it's an attributive clause, this is the main clause* [points to 'statistics' in sentence (ii)], *this, that is according to the age*, **(I)** ***can judge it is an article or a report, in order to have what can be considered rigorous data***

Firstly, he engaged in a strategy of **parsing**. Although he did not know the meaning of the word *statistics*, he immediately identified it as the subject of the sentence, then identified the word *show* as the verb and finally identified the presence of a *that-relative clause*, which he referred to in Chinese as an "attributive clause" (relevant utterances shown in bold and underlined). Hence, it can be seen here that his first step in regulating his comprehension of this sentence was to parse it into its basic grammatical elements. The second strategy he used was **to infer the meaning of the unknown word from context**, by suggesting that the unknown word *statistics* meant something like "article" or "report". In this instance, he explicitly stated that he had inferred this meaning from the fact that the sentence contains a lot of "rigorous data", and this type of information is most likely to be contained in a report (relevant utterances shown in bold). Finally, the fact that he was able to make this inference resulted from him **applying pre-existing knowledge of the characteristics of reports as a genre**.

Participant 025 explicitly acknowledged a comprehension difficulty before engaging in a comprehension regulation strategy. However, there were other instances where a reader was judged to have experienced some degree of processing difficulty, based on the fact that they engaged in comprehension regulation, even

when they themselves did not explicitly report having done so. A clear example of this was seen when participant 033, from the intermediate group, read the same sentence. As can be seen from the following excerpt from her transcript, after reading the sentence, Participant 033 appeared confident that she had understood the sentence, saying "I know the meaning of these words". Utterances originally made in Chinese are shown in italics

> **P033** ... in the United States, in US, so here is an example [points to for example sentence (*ii*)] in US, statistics shows the death rate from this [indistinct while reading] so, *statistics*, *a statistical agency shows, because the death rate from this brain disease* [points to brain disease in sentence (*ii*)] *from*, *at this* [points at 1980 and 1997 in sentence (*ii*)], this time to this time, between this time, *decreased*, *decreased to thirty five* percent, right, just because I know the, I know the meaning of these words

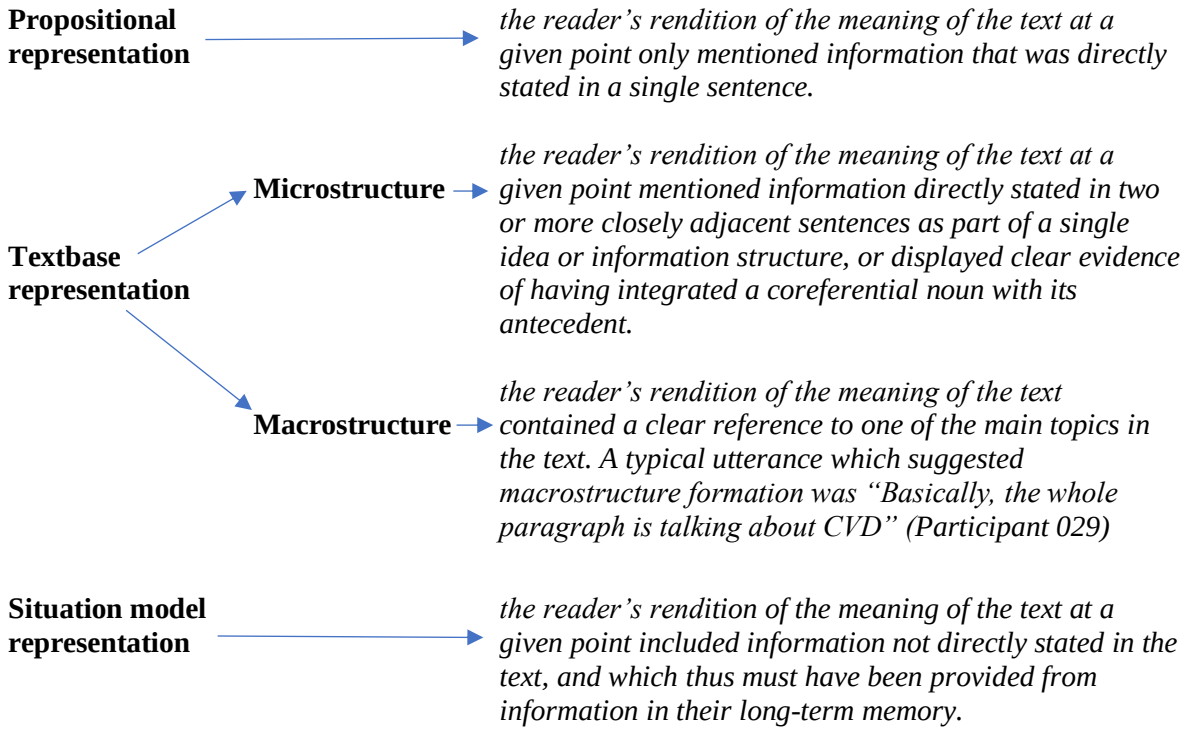
However, as can also be seen, when reading the sentence she frequently engaged in a strategy of **code switching**. Although it is not clear from the transcript why she did this, the fact that she felt the need to revert to Chinese to understand parts of the sentence was interpreted as suggesting that she had experienced some degree of difficulty processing the sentence in English, even if she was able to overcome this difficulty with relative ease. It should be noted though that in such instances as these, the judgement that a reader experienced processing difficulty was arrived at inferentially, and cannot be considered as reliable as those occasions when a reader explicitly stated that comprehension difficulty had occurred.

5.5.1.3 The Construction-Integration model of reading comprehension framework

The second layer of analysis used to explore why the participants differed in their ability to identify the incongruent words in the text focused on the type of mental representations they were able to establish as they read. As explained in section 3.7.4, the model of reading comprehension adopted for the current study was Kintsch's (1998) Construction-Integration model. Inevitably, it was not possible to determine with a high degree of accuracy exactly what type of mental representations the participants established as they read; rather, judgments were made from the verbalisations they gave *as* they read. It is perhaps worth noting at this point that this layer of analysis was added after the data collection had been completed. It resulted from the finding from the quantitative analysis of the think-aloud transcripts which, contrary to expectations,

indicated that while in general the lower-proficiency participants were not able to identify the incongruent words in the articles they read, they were able to integrate the target coreferential noun phrases with their antecedents. Consequently, this phase of the analysis was devised post hoc to investigate the extent to which the types of mental representations the lower and higher proficiency readers were able to establish as they read impacted on their ability to identify the incongruencies buried in the text.

Figure 5. 1: The coding criteria used to judge the type of mental representations established (repeated from methodology)



As explained in section 3.7.4, in an attempt to bring a degree of objectivity to the analysis, judgements as to what types of representations the participants formed as they read were made according to a strict set of criteria, as show in Figure 5.1 (repeated from section 3.7.4 for ease of reference). However, given that the participants were not directly asked what they believed the texts to mean, but rather described this incidentally as they explained how and why they thought the sentences were related to each other, it may be that they established mental representations, as defined by the framework shown in Figure 5.1, for which they did not show evidence in their transcripts. Consequently, while the analysis presented below is argued

to provide a clear and accurate picture of the types of representations the participants established, the possibility exists that it is at least partially incomplete.

5.5.2 The analysis

The following section describes the circumstances under which the participants' ability to identify incongruent words in the text was either empowered or impeded. The findings from the analysis indicated that the key determinant of whether or not the readers identified an incongruent word in the text was whether or not they had been able to establish a situation model representation of the relevant section of text. As will be seen, the establishment of a situation model representation typically required the reader to be able to process the language of the text efficiently. However, as will also be seen, there were also occasions where despite processing the text without difficulty, a situation model was not established, and other occasions where despite the reader experiencing processing difficulty, a situation model representation was established.

Given the cumbersome nature of the analysis, it was not possible to present excerpts from the transcripts of all the participants. Consequently, only particularly representative examples are presented below.

5.5.2.1 Condition 1: The incongruent word was identified

The analysis of the transcripts indicated that in the overwhelming majority of cases, readers were only able to identify the incongruent words when they also showed evidence of having established situation model elements in their mental representations of the relevant sections of the text. In general, this only occurred when they were able to process the language of the text efficiently and with minimal need for comprehension regulation (although see Section 5.5.2.3 for exceptions), thereby explaining why only the upper-intermediate readers were able routinely to identify the incongruent words. Furthermore, it was also found that situation model representations could be formed by readers drawing on their personal experience of the world, attitudes towards the information in the text, or specialist knowledge of one of the topics discussed in the article. It was also found that in some instances, the establishment of a strong

situation model, instead of allowing the reader to identify an incongruent word in the text, caused them to suppress it altogether in their mental representation in favour of a contextually appropriate alternative. Finally, although the establishment of a situation model was generally required for a reader to identify an incongruent word, there was one example to suggest that at times the establishment of a macrostructure textbase representation may also suffice.

5.5.2.2 Three sources of situation model representations

In describing the formation of situation models representations, Kintsch (1998) states that they may be established as a result of the activation of pre-existing knowledge of, or attitudes towards, the information expressed in the text. The analysis of the transcripts found evidence to support both of these conditions. An example of where a reader's pre-existing knowledge, drawn from her own personal experience of the world, resulted in the establishment of a situation model was seen when Participant 009, from the upper-intermediate group, read the opening passage of the article:

> (*title*) Cardiovascular Disease: A tale of good news and bad
>
> (*i*) Deaths from *cardiovascular disease (CVD), which includes heart attacks and high blood pressure*, first began to decrease in the 1960s in Western countries, and have continued to do so ever since. (*ii*) In the United States, for example, statistics show that death rates from **these brain diseases** fell by over 35 percent between 1980 and 1997

As can be seen from the excerpt from her transcripts shown in Figure 5.2, despite not knowing the meaning of the word *cardiovascular* in the title, she was able to swiftly regulate her comprehension of the word by drawing on her personal experience of going to the gym. Building on the establishment of this situation model representation, she was able to confirm that she was correct in her understanding when she read *heart attacks and high blood pressure* in sentence (*i*), and immediately identified the incongruency of the word *brain* when she read the target noun phrase shown in bold in sentence (*ii*).

Figure 5. 2: Excerpt from Participant 009 – example of a situation model containing information drawn from personal experience

P009: The title of this paper is called Cardiovascular [sounds out word syllabically] Disease, I think that's kind of disease, sounds like a horrible disease, I have no idea, cardio, something to do with your heart, with your breath, with all the body system

R: ok, why do you say that?

P009: Because I go to gym for cardio, one of the exercises is called, party cardio party

R: yeah, yeah, the one they dance to

P009: Yeah, yeah, it think to do with oxygen, with oxygen, not only do the weights, something good for your heart

A second example of a reader drawing on their pre-existing knowledge of the topic under discussion was found in the transcript of Participant 071 from the intermediate group. As can be seen from Figure 5.3, when she encountered the word *cardiovascular,* she immediately related it to the topic of her PhD research. That this allowed her to establish a situation model representation of the text is indicated by the fact that her subsequent rendition of the text's meaning included the subject specific technical term "induce", a word not used in the original text, and which thus must have been supplied from long-term memory.

Figure 5. 3: Excerpt from Participant 071 – example of situation model containing expert knowledge about the topic of the text

P071: Cardio [points to 'cardiovascular' in title], this disease I know this is a heart related disease， yeah, and it's a story about good and bad, it's show two sides.

R: How do you know that's a heart related disease？ You said， I know this is a heart related disease。How do you know that？

P071: Because my research is about the health, a little about the health

….

P071: This is the decrease [draws an arrow from cardiovascular disease to decrease] means get less, and have continued to do so ever since。It means it started, started in this time [points to 1960s in line 3], and it's always have this phenomenon until then and the result, or the bad thing is it can **induce** heart attack and high blood pressure

Finally, in addition to drawing on knowledge of the topic under discussion, there was also evidence to show that the activation of attitudes towards it could result in the establishment of situation model

representation, thereby allowing the reader to identify an incongruency. An example of this was seen when Participant 056, the highest proficiency reader in the whole sample, read the following section of text.

> So, what are the factors that could have brought about **these discouraging reductions** in the rates of heart disease?

As can be seen from Figure 5.4, when reading the target noun phrase shown in bold, Participant 056 experienced immediate processing disruption, based on her pre-existing, albeit commonplace, attitudes towards reductions in the rate of diseases. That this resulted in the establishment of a situation model representation is indicated by fact that she describes reduction in rates as "positive phenomenon", a phrase not used in the text, and which therefore must have been activated from long-term memory.

Figure 5. 4: Excerpt from Participant 056 – example of a situation model representaiton containing prior attitudes towards an idea in the text

> **P056**: **...** I am a little bit confused about discouraging, because **I think reduction in the rate of heart disease is something good for human beings, it is positive phenomenon**, but why here it mention discouraging... I think maybe it's an error or mistake that the writer has made...

5.5.2.3 Situation knowledge as compensation for processing difficulty

Although in general only the upper-intermediate readers were able to identify the incongruent words in the text, there were a small number of occasions where lower proficiency readers were also able to do so. Kintsch (1998) argues that when readers are unable to establish coherent textbase representations, they may be able to compensate for this by supplementing their representation of the texts with information in long-term memory about an idea in the text. Clear evidence for such an occurrence was found in the transcripts of Participant 032, the second lowest proficiency reader in the entire sample, when she read the following passage:

> (*i*) Here, rapid social development is introducing *the lifestyle patterns that we usually associate with CVD – a decrease in physical activity, an increase in smoking, and a change to a less healthy diet*. (*ii*) Unfortunately **harmless behaviours** are widely predicted to increase in the years to come.

Examination of her transcript, shown in Figure 5.5, reveals that she experienced considerable difficulty processing sentence (*i*) of the passage; however, she was able to establish a clear representation of the idea of an increase in smoking, a habit widely known not only to be harmful, but also causally related to cardiovascular disease. Perhaps in consequence, when she read sentence (*ii*), despite having read the word *harmless* aloud, and even underlined it, when she translated the sentence into Chinese to explain its meaning, she used the phrases "bad habits" and "harmful habits". This suggests that the activation of prior knowledge regarding smoking resulted in the establishment of a strong situation model representation of that section of the text, which in turn resulted in her suppressing (see following section for a fuller discussion of suppression) the incongruent word *harmless* in her mental representation of the text in favour of the contextually appropriate word "harmful".

Figure 5. 5: Excerpt from Participant 032 (utternaces originally made in Chinese shown in italics) – example of strong prior knowledge resulting in situation model formation, despite difficulty processing the language of the text

P032: Here, rapid *social development* is introducing, is introducing [indistinct while reading] *rapid social development is bringing lifestyle changes, because of this we often want*, it's ok, asso... [attempts 'associated' in sentence (*i*)] *I can't think what the word I am looking at is, we also want, I can't think what this word is* [points to 'associated' in sentence (*i*)] , *that is, this disease can also have that kind of* physical*, it's probably said like that, physical treatment, I don't know if it's said like this*, physical activity [indistinct while reading] increase, ***it leads to, that is, smoking can, can, smoking can, can increase you rate of this disease***, and a change into [indistinct while reading] , *as well as an unhealthy diet, you, probably you can use a kind of healthy, this kind of diet can reduce your* [indistinct]

... Unfortunately *there are still those kinds of* ***bad habits*** [underlines harmless in sentence (*ii*)], harmful, ***harmful habits*** and [indistinct while reading] predicted *guess* to *this word probably means guess* [circles predicted in sentence (*ii*)], *I remember, is it guess*, predicted *guess* [indistinct while reading], *that is those, its* ***these habits*** **[gesticulates loosely to all earlier text]** ***can possibly cause many people to catch cardiovascular disease*****.**

5.5.2.4 Further evidence of the establishment of a situation model representation resulting in an incongruent word being suppressed.

Participant 032 was not the only participant who exhibited evidence of having suppressed an incongruent word in favour of a contextually appropriate one as a result of having established a situation model bearing representation of the text. Suppression is a cognitive mechanism which "dampens the activation of contextually inappropriate meanings" (Gernsbacher & Faust, 1991, p. 246). Three of the occasions where there was evidence that a reader had suppressed the meaning of the incongruent word came when Participants 071, 060 and 032 read the target noun phrase *(these) harmless behaviours*, as described for Participant 032 above. In respect of this item, it may be that reading the root of the incongruent word, *harm*, following immediately after the readers had processed at least one of the three noun phrases *a decrease in physical activity, an increase in smoking,* and *a change to a less healthy diet* provided enough activation to the contextually appropriate form *harmful* for the suffix ~*less* to be suppressed in their discourse-level representation of the text. However, there was also evidence to suggest that even when an incongruent word differs substantially in form to the contextually appropriate alternative, it may still be suppressed if the reader has established a strong enough situation model of the text. A clear example of such a case was found in the transcripts of Participant 056 when she read the following passage:

> (i) *The surgical treatment of heart disease* will continue to benefit patients in countries where the expertise, equipment and other resources are available. (ii) By itself, however, **such simplistic treatment** is not an adequate response to the global CVD problem. (iii) **It** is so expensive that it is an enormous burden even for the wealthy nations that can afford to offer it at present. (iv) For developing nations, which often do not have the necessary financial, technological and human resources, **this approach is simply not possible**.

It may be instructive to note that prior to reading the incongruent noun phrase (shown in bold and underlined in sentence (*ii*), Participant 056 had seen the contextually congruent word *surgery*, or a closely related word, on three previous occasions, meaning it was likely to be highly activated in her memory:

> **First mention**: As a result, doctors can more easily treat conditions which are detected early, either by prescribing drugs to lower blood pressure or by **performing minor surgery**....

Second mention: Thanks to more high-tech equipment, better drugs and **more experienced surgeons**, many people now recover and go onto lead active lives after **operations** that were impossible just four or five decades ago....

Third mention: The **surgical treatment** of heart disease will continue to benefit patients

That Participant 056 may have suppressed the incongruent word *simplistic* in favour of the congruent word *surgical* is suggested by the way she read sentences (*iii*) and (*iv*) of the passage, as shown in the excerpts from her transcript given in Figure 5.6. In this passage, the idea of *surgical treatment* is first mentioned as *surgical treatment* in sentence (*i*), incongruently referred to as *such simplistic treatment* in sentence (*ii*), recast by the pronoun *it* in sentence (*iii*), and finally referred to as *this approach* in sentence (*iv*). Although Participant 056 did not react to the incongruent word when she read the target noun phrase in sentence (*ii*), when she read the coreferential terms *it* and *this approach* in the subsequent two sentences, she explicitly used the word *surgery* on both occasions to explain their meaning. This suggest that the reason that she did not react to the incongruency of the word *simplistic*, despite having identified all three of the other incongruencies in the text, was that she had developed a strong situation model representation of the idea of *surgery*, as indicated by her introduction of words such as *modern*, *money* and *experts* which were not included in the text. This situation model was so strong that it caused her to suppress the contextually incongruent word *simplistic* in her discourse-level representation of the text altogether.

Figure 5. 6: Excerpt from Participant 056 – example of suppression of an incongruent word

> **P056**: So, the surgical treatment of heart disease will continue to benefit patients in the countries where the expertise, equipment and other resources are available, yeah, so like, western countries they have money, they have experts, they have modern equipment, yeah the surgery right, it can help
>
> By itself, however, such simplistic treatment is not an adequate, adequate response to the global CVD problem, so although it works in western countries, but it is not a way out for the whole world, yeah, so what's the reason
>
> It is so expensive that it is an enormous burden even for the wealthy nations that can afford to offer it at present, because even for developed countries the surgery is quite expensive, so let alone, we can imply that, let alone even for developing countries it will be much more expensive, it will be a burden for them
>
> So, for developing nations, which often do not have the necessary financial, technological and human resources, this approach is simply not possible, so, surgery, this approach refers to surgery, surgery, it's too expensive, right, and also because of other reasons, they don't have enough techniques, they don't have enough professional doctors, that's what human resources refers to, so that's not possible in developing countries

5.5.2.5 Macrostructure representations

Although the predominant requirement for identification of the incongruent words was establishing a situation model, there was one example in which the establishment of a **macrostructure** representation of one of the major topics in the text was sufficient. This occurred in the case of Participant 029 (Figure 5.7) when she had established a sufficiently strong representation of the central topic of improvement in the treatment of heart disease developed in the first paragraph for her to experience the word *worsening* in paragraph 2 as incongruent.

Figure 5. 7: Excerpt from Participant 029 – example of macrostructure formation allowing reader to identify an incongruent word

> **P029:** Firstly, in spite of worsening techniques for treating heart related problems, heart disease is still the most common cause of death in most western countries, so this part see, so this is something I don't understand, its worsening, so it means like bad, so it means, despite bad techniques for treating heart related problems, so **because the first paragraph they talk about, like say the treatment has improved, but here it talks about worsening techniques. So, it doesn't make sense to me.**

5.5.2.6 Condition 2: The incongruent word was not identified

The analysis presented above argues that in order for a reader to have identified an incongruent word it was, in almost all cases, necessary for them to have established a situation model representation of the text. Similarly, the analysis of the transcripts of those participants who did not identify the incongruent words indicated that in every case there was no evidence of their having established a situation model, although they did on some occasions manage to establish more localised, text focused discourse representations. The analysis of the transcripts in which the readers did not identify the incongruent words highlighted three main reasons why they were unable to do so: *the incongruent word was unknown*; *the reader experienced difficulty processing the language of the text*; *the reader engaged* in *skim reading.*

5.5.2.7 The incongruent word was unknown to the reader

Perhaps rather prosaically, among the lower-intermediate readers the most common reason they did not identify the incongruent words in the text was simply that they did not know what the words meant. In this group, of the fifteen cases where the incongruent word was not identified, there were nine where the participant explicitly stated that they did not know what it meant. Instances where the reader did not know the meaning of the incongruent words were spread roughly equally across all four participants. Among the intermediate readers, although this was less common, there were still four instances where a participant declared the incongruent word to be either unknown or unfamiliar to them. Again, this was spread equally across all four participants in the group. Unsurprisingly, among the upper-intermediate readers unknown vocabulary was much less of an issue; although, here too one of the participants, 006, reported that one of the incongruent words was unfamiliar to her. Given that the word was unknown in these cases, and thus could not be experienced as incongruent, these examples shed little light on the question at hand.

5.5.2.8 Difficulties processing the language of the text

The second reason the readers did not identify the incongruent words was because they experienced difficulties processing the language of the text, as evidenced by the fact that they engaged in some kind of

comprehension regulation strategy. Where readers experienced difficulty processing the text, they were typically unable to establish a situation model representation (although see Section 5.5.2.3 for exceptions). The participants experienced two types of language processing difficulties. The first was with **unknown vocabulary**. The second was with **complex syntax**. As predicted by Halliday (1993a; 1993b), difficulties processing complex syntax related primarily to complex postmodified noun phrases; however, as will be shown below, the analysis of the transcripts also found evidence of a participant struggling to process a postmodified verb phrase.

5.5.2.9 Difficulty with unknown vocabulary

When readers encountered unknown vocabulary, this frequently resulted in them being unable to establish a situation model representation or to identify the incongruent words. A particularly extreme example was found in the transcript of Participant 025, from the lower-intermediate group, when he read the opening passage of the article:

> (*title*) Cardiovascular Disease: A tale of good news and bad
>
> (*i*) Deaths from *cardiovascular disease (CVD), which includes heart attacks and high blood pressure*, first began to decrease in the 1960s in Western countries, and have continued to do so ever since. (*ii*) In the United States, for example, statistics show that death rates from **these brain diseases** fell by over 35 percent between 1980 and 1997

As can be seen from the excerpt from his transcript shown in Figure 5.8, Participant 025 encountered multiple unknown words in the text. As a result, despite having used a number of comprehension regulation strategies, he was unable to establish even a stable, accurate propositional representation of this sentence, let alone a situation model representation. Perhaps unsurprisingly, given the difficulty he encountered reading the antecedent, when he encountered the incongruent target noun phrase *these brain diseases* in sentence (*ii*), he showed no sign of noticing that the word *brain* was incongruent with the idea of heart disease expressed in sentence (*i*), despite having established at least a partial propositional representation of that sentence.

Figure 5. 8: Excerpt from Participant 025 (utterances in italics originally made in Chinese) – first example of difficulty with unknown vocabulary

P025: ... death from, death *that* is die, *some kind of* die, *then*, which *a connective, this one after is a* which *attributive clause* [points at 'which' in sentence (*i*)], *then, this includes* high blood, is a kind of illness to, and this is and [points to 'and' in sentence (*i*)], so, this word [points to 'heart attack' in sentence (*i*)] is a kind of illness too.

R: Do you know this word [points to 'heart' in sentence (*i*)]?

P025: Heart, I think it's this is here [points to his heart], and high blood, I think.

R: Okay.

P025: So, this word [points to 'cardiovascular disease' in sentence (*i*)] *is a superordinate term for these two kinds of disease*

R: ... how do we know *this is a superordinate term*?

P025: *for example*, this which, there which *this word* [points to 'which' in sentence (*i*)] *afterwards is its attributive clause, this attributive clause, this is, this* from, *this is, this whole sentence's subject, then this attributive clause is modifying this subject* [points to 'cardiovascular disease']

R: So, what could we guess this word means [points to 'disease' in sentence (*i*)]?

P025: *this is the superordinate term for this kind of disease, it's also a technical term*

P025: ... pressure [reads 'pressure' at the start of the line as new word, rather than as part of high blood pressure] first began to /di:skri;s/ [attempts 'decrease'] in the one thousand nine hundred sixties in Western countries and have continued to do so ever since, *the meaning of this sentence is, this is probably a connective* [points to 'pressure'], the first, *just started* decrease, *decrease in 1960 in western countries*, and *this connective* [points to 'continued'], *I don't know what it means*, have *this is perfect tense, it's a (verb) tense sentence, it shows it's already complete, and during this period* (points to '1960s'].

P025: In the United States *this is in America*, for example, such as [indistinct] *I don't know this word* [points to 'statistics' in sentence (*ii*)] I don't know this meaning, but *this word* [points to 'statistics' in sentence (*ii*)] *is the subject of this sentence ... this is a verb* [points to 'show' in sentence (*ii*)], *this is followed by a* that *attributive clause, ... this is* death rates from this brain disease fell by, *this* death rate *is* the advantage of the death from this, *that is, this brain disease, so I infer this* [points at 'statistics' in sentence (*ii*)] *probably refers to an article or a report*

R: Why do you think that?

P025: *This is the subject* [points to 'statistics' in sentence (*ii*)] *the bit after this says... the bit after shows this brain disease's death rate, death rate, these years, exceeded 35% ... in 1980 and 1997, it also says, these clauses after, how can I put this, this bit after, this is* that *it's an attributive clause, this is the main clause* [points to 'statistics' in sentence (ii)], *this, that is according to the age*, (I) *can judge it is an article or a report, in order to have what can be considered rigorous data*

Although Participant 025 experienced particularly extreme difficulties with unknown vocabulary, it was found that even relative minor difficulty with unknown vocabulary could impede the establishment of a

situation model representation when this resulted in the need for comprehension regulation. An example of such minor difficulty was found in the transcript of Participant 013 when she read the same opening passage. As can be seen from the excerpt shown in Figure 5.9, Participant 013 regulated her comprehension of the unknown word *cardiovascular* in sentence (*i*) simply by noting that it was a kind of disease whose "name is CVD", and of the phrase *high blood pressure* by translating each word into Chinese to produce the Chinese equivalent. When she read sentence (*ii*), she engaged in a strategy of code switching, suggesting that she felt the need to rely on her first language to understand at least part of the sentence.

As can be seen from her transcripts, although she was able to establish accurate propositional representations of both sentences, Participant 013 showed no evidence of having established a situation model representation, or of having identified that the word brain in sentence (*ii*) was incongruent in this context. When explicitly asked what she thought *these brain diseases* referred to, she replied "brain, brain", while pointing to her head, and then moved immediately to the next sentence, thereby suggesting that she may not even have integrated the two sentences to establish a microstructure textbase representation. This suggest that even when readers are able to regulate their comprehension of unknown words well enough to establish accurate propositional representations of individual sentences, this does not necessarily allow them to establish discourse-level representations, or to identify incongruencies in a text.

Figure 5. 9: Excerpt from Participant 013 (utterances in italics originally made in Chinese) – second example of difficulty with unknown vocabulary

P013: ... he said death, I guess this is the name of some like *disease*, right, so it doesn't matter, just know his name is CVD [points to 'CVD')] ... and includes heart attacks and high blood pressure, like heart [points to her own heart] and high blood, I know the meaning... it's two kinds of illness, it's like your heart and high blood pressure, can I explain in Chinese, it's *high blood pressure* [points to each word as she translates]... so, *this disease, from, in western countries from 1960 started to decrease*, right, continued to do so ever since, and from now they always decreases the illness from this time, right?

... in the United States, in US, so here is an example [points to for example sentence (*ii*)] in US, statistics shows the death rate from this [indistinct while reading] so, *statistics, a statistical agency shows, because the death rate from this brain disease* [points to brain disease in sentence (*ii*)] *from, at this* [points at 1980 and 1997 in sentence (*ii*)], this time to this time, between this time, *decreased, decreased to thirty five* percent, right, just because I know the, I know the meaning of these words

Finally, evidence was found which suggested that even when readers encountered unknown vocabulary in one of the noun phrases in a coreferential pair, although they may be able to establish a local microstructure discourse representation, this is not necessarily enough for them to identify an incongruent word. For example, as can be seen from Figure 5.10, when Participant 033 read the opening passage of the article, she too needed to regulate her comprehension of the unknown word *cardiovascular*, which she had previously stated that she did not know. However, despite having been successful in her attempts at regulation, not having encountered any further difficulties reading the passage, as well as having displayed evidence of integrating the target noun phrase *these brain diseases* with the antecedent, she was still unable to identify the incongruency of the word *brain*. Notably, she did not display any evidence of having established a situation model representation. Hence, the evidence suggests that even when readers are able to regulate their comprehension of unknown words well enough to establish local microstructure discourse representations, this is not necessarily enough to allow them to identify incongruent words in a text.

Figure 5. 10: Excerpt from Participant 033 – third example of difficulty with unknown vocabulary

P 033: Deaths from [underlines 'heart' and 'high blood' in sentence (*i*)] oh God, I think I know its meaning in Chinese [points to 'cardiovascular']

R: Ok, which is?

P033: 三高 [**translation**: *the three highs* – used in Chinese to refer to hypertension, hyperglycaemia and hyperlipidaemia (e.g. high cholesterol)]

R: OK, could be. I don't know the word. What does it mean?

P033: High blood *high pressure*, *heart disease*, um, I don't, forget it... this kind of disease [points to 'disease' in sentence (*i*)] includes heart attacks and high blood, and these two [draws line linking 'heart attacks and high blood' in sentence (*i*) and labels them 'diseases'] are the two kind of diseases, right ? But this is the whole article [indicates whole text beneath], God! [resumes reading] first began to decrease, this kind of disease [points to 'disease' in sentence (*i*)] was fined, sorry, um, the people the population have this kind of disease [points to 'disease' in sentence (i)] decreased since 1960s, in western countries, right?

... in the United States, for example, in the UK, sorry, in the USA, maybe some richer, [indistinct], some studies were done, so there are some data about the death rate, **death rates from this brain disease fall over, the studies show that this this kind of disease [points to disease in sentence (*i*)]**, disease, hmm, decreased, decreased about 35% between these two years [points to 1980 and 1997].

5.5.2.10 Difficulty processing complex syntax

The second difficulty participants faced when processing the language of the text was with the kind of complex, postmodified noun phrases Halliday (1993b) identified as likely to cause difficulties for native Chinese speakers. A particularly clear example of such an occurrence was found when Participant 004, from the intermediate group, read the following passage:

> (*i*) Another important factor is that the diagnosis and treatment of *heart disease* has improved. (*ii*) These advances have made possible **the early diagnosis of *mild illnesses* that may cause serious problems later**. (*iii*) As a result, doctors can more easily treat conditions which are detected early, either by prescribing drugs to lower blood pressure or by performing minor surgery.

The target noun phrase in this instance, *mild illness*, was embedded in the kind of complex post-modified noun phrase that has been mentioned throughout this thesis. Inspection of the noun phrase underlined in sentence (*ii*) shows that the head noun in the phrase, *diagnosis*, is pre-modified by a determiner, *the*, and an adjective, *early*. In addition, it is also post-modified by a prepositional phrase, *of mild illness*, and a relative clause, *that may cause serious problems later*. It is instructive to note at this point that a simpler version of sentence (*ii*) could have read:

> These advances have made possible early diagnosis.

As can be seen from the excerpt shown in Figure 5.11, despite having established both a basic propositional and textbase representation of sentence (*ii*) and its relationship to the previous text, Participant 004's rendition of its meaning did not include the information in the relative clause at the end of the complex noun phrase highlighted at the end of sentence (*ii*). Consequently, when she read sentence (*iii*), she experienced insurmountable difficulties with discourse integration.

The confusion she experienced as she read this sentence was caused by the fact that she had correctly interpreted the relationship between sentences (*ii*) and (*iii*), as indicated by the connective *as a result*, as being one that could be summarised as 'a *good situation in sentence* (*ii*) *brings about a positive result in sentence* (*iii*)', with the positive result being that doctors can now more easily treat conditions. However, as

a result of not having correctly integrated the elements of the complex noun phrase in sentence (*ii*), she did not understand why the negative situation *cause problems later*, should have led to a positive outcome in sentence (*iii*). Put simply, this was because she interpreted the information in the relative clause as being self-standing, as opposed to being part of a larger semantic unit.

Figure 5. 11: Excerpt from Participant 004 – example of difficulty processing a complex noun phrase

> **P004**: As a result, doctors can more easily treat the conditions which are detected, as a result early [emphasises early], either by prescribing drugs to lower blood pressure or by performing minor surgery as a result doctors can more easily treat condition, treat conditions which are oh, as a result, so this doctors could prescribing drugs [pointing at 'prescribing drugs' in sentence (*iii*)] and minor surgery [pointing at 'minor surgery' in sentence (*iii*)] it said, oh why, oh, I'm confused, why this, I'm confused about this sentence [draws brackets around sentence (ii)]...
>
> **R**: why are you confused? What's the problem here?
>
> **P004**: Because it says as a result, so maybe it's, it's … and it can easily treat conditions so I guess before this sentence [indicates sentence (*iii*)] this, this, this words [indicates sentence (*ii*)] means a good, it's a good result, but then sentence, it says cause serious problems later, so it means something will become bad, but, as a result, I don't know I don't whether bad is, is causing something good, yeah, so it's a little confuse me now

This problem would not have occurred if the writer had used the simpler version of sentence (*ii*) described above. As such, this example provides a clear instance of where difficulty with processing a complex post modified noun phrases, of a type which does not exist in Chinese, impeded a reader from a establishing a strong enough sentence-level representation for subsequent discourse integration processing to occur. Consequently, her comprehension broke down, and she showed no evidence of having established a situation model representation of this section of the text, nor of having identified the word *mild* as being incongruent when used to describe heart disease, which had previously been described as causing death.

Although Halliday (1993b) only highlighted postmodified noun phrases as likely to cause comprehension difficulties for native Chinese speakers, there was evidence in the transcripts that post-modified verb

phrases can also cause comprehension breakdown. A clear example of this arose when Participant 071, from the intermediate group, read the following passage:

> (*i*) Here, rapid social development is introducing *the lifestyle patterns that we usually associate with CVD – a decrease in physical activity, an increase in smoking, and a change to a less healthy diet*. (*ii*) Unfortunately, ***these harmless behaviours*** are widely predicted to increase in the years to come**.**

As can be seen from Figure 5.12, when she read sentence (*i*) of this passage, Participant 071 was able to establish both a propositional and discourse-level representation of its meaning, clearly identifying CVD as the same disease mentioned at the start of the article. However, when she read sentence (*ii*), she quickly experienced processing difficulties when she arrived at the final verb phrase (relevant text underlined in standard font). When she first began reading the sentence, there is evidence to suggest that she did begin the process of integrating it with the information in sentence (*i*), as indicated by her claim that "they said again, this is harm...". When she arrived at the final verb phrase in the sentence though, she said that she was less sure of the sentence's meaning.

When asked why this was, she explicitly stated that there were "several parts" to the verb phrase, indicating the trouble was indeed with the complexity of the syntax. She then attempted to regulate her comprehension by parsing the sentence into its basic *subject-verb-object* components.

Figure 5. 12: Excerpt from Participant 071 – example of difficulty processing a complex verb phrase

> **P071**: The cause, is it makes people to have the new lifestyle pattern, the lifestyle pattern is like this [points to the three examples at the end of sentence (*i*)], and we usually associate with, so it means the rapid social development made people have this unhealthy lifestyle patterns, and as a result it make people have this disease.
>
> **R**: Which diseases is this, the CVD [points to CVD sentence (*i*)]?
>
> **P071**: CVD is here [points to CVD in first sentence of the article] **...** Unfortunately, these harmless behaviours are widely predicted to increase the years to come, yes, they said again this is harm, harmless, it's bad, is widely predicted to increase the years to come, I'm not sure, it's widely predicted to increase
>
> **R**: Which bit aren't you sure, what aren't you sure about?
>
> **P071**: Yeah, now I after second I, I know it means in the future it may increase, and what's the increase, what's the future, it's in the years to come, the upcoming several years.
>
> **R**: When you first read that you said you weren't sure, what stopped you?
>
> **P071**: Because I didn't find the, [indistinct] in the sentence there are several, several parts [draws diagram indicating syntactic structure of a sentence], this is for the *subject-verb-object* and here is the time, and here's the location and the part is like this is and so on.

However, the extent to which this attempt at regulation was successful is unclear, as her rendition of the meaning of the sentence was distinctly vague, and included a reference to location, which was not in fact mentioned. At this point, she moved on to the next sentence in the passage, showing no clear evidence of having established a propositional representation of the sentence as a whole, of having integrated the target noun phrase with its antecedent, nor of having identified the incongruity of the adjective *harmless* in the sentence. As such, despite having shown initial signs of integrating the target noun phrase in sentence (*ii*) with its antecedent in sentence (*i*), there is evidence to suggest that this process stalled as a result of difficulty with syntactic processing, with the consequence that she was unable to establish a situation model representation of the sentence, or to identify the incongruent word.

Taken together then, these two examples provide evidence to suggest that where readers experience difficulty processing complex grammatical structure, this can impede them from integrating adjacent sentences or establishing situation model bearing discourse-level representations.

5.5.2.11 Skim reading as detrimental to situation model formation or the identification of incongruent words

The third reason the incongruent words were not identified was when a reader engaged in skim reading. That this was seen in the transcript of Participant 029, the second highest level reader in the entire sample, is instructive of the need for a reader to establish a situation model representation if they are to identify incongruent words in a text. When engaged in skim reading, the reader processes the text quickly, paying minimal attention to detail, with a view simply to understanding the gist of what is being said. A clear example of such an approach was seen when Participant 029 read the opening passage of the text:

> (*title*) Cardiovascular Disease: A tale of good news and bad
>
> (*i*) Deaths from *cardiovascular disease (CVD), which includes heart attacks and high blood pressure*, first began to decrease in the 1960s in Western countries, and have continued to do so ever since. (*ii*) In the United States, for example, statistics show that death rates from **these brain diseases** fell by over 35 percent between 1980 and 1997

As can be seen from Figure 5.13, she encountered no difficulty processing the language of the text, and was consequently able to establish accurate propositional representations of both sentences (*i*) and (*ii*). However, she also explicitly stated when reading sentence (*ii*) that she would not usually remember details of the type given, as she did not think they were important. Perhaps as a consequence of this somewhat surface level approach to reading, she showed no evidence of having identified the incongruent word *brain*, or even of having integrated the target noun phrase with its antecedent at all. The fact that, despite having adopted a similar skim reading approach throughout, Participant 029 was able to integrate three of the four target noun phrases with their antecedents, suggest that such an approach to reading may be sufficient to establish at least microstructure textbase representations. Conversely, the fact that she was only able to identify one of the four incongruent words in the text suggests that the establish such textbase discourse representations alone is not sufficient to allow a reader to identify incongruencies as they read.

Figure 5. 13: Excerpt from Participant 029 – example of skim reading

P029: Ok. Deaths from this disease, which includes heart attacks and high blood pressure first began to decrease in the 1960s in western countries, ok, so it explains this one [points to cardiovascular in sentence (*i*)], and um, it tells you a timeline, and have continued to do so ever since, so basically, it just appears and keep affecting people's life,

P029: In the United States, for example, statistics show that death rate from these brain diseases fell by over 35 percent between 1980 and 1997, so I won't really remember these figures, if, because I don't think it's important.

R: So what, you won't remember them?

P029: I... won't, because I'm first, I'm not a doctor, so second, if I'm like doing the reading comprehension, I would first look at these questions [indicates anaphor resolution task] and then just...

R: But these questions are not really for reading comprehension, they're just for my research. So, just think about your understanding.

P029: So, basically shows that, it tells you a figure that the people suffer from brain disease fell down between this period [points to dates in sentence (*ii*)],

5.5.3 Conclusion

The analysis presented above set out to address the following research question:

RQ4. Why were the upper-intermediate readers able to identify the incongruent noun phrases while the intermediate and lower-intermediate readers were not?

The analysis found that while the ability to identify the incongruent words in the text was largely confined to the upper-intermediate readers, this was a trend not a rule. On a small number of occasions both intermediate and lower-intermediate readers were also able to identify an incongruent word. Furthermore, while the upper-intermediate readers were typically able to identify the incongruent words, Participant 029, the second highest proficiency reader in the sample, failed to identify three of the four, while even Participant 056, the highest proficiency reader, failed to identify one. This suggests that while there is some relationship between a reader's level of proficiency and their ability to identify incongruent words in a text, this relationship is not absolute.

What is more, the findings also revealed that while the ability to identify the incongruent words in the text was limited to a small subset of the sample, the ability to integrate the coreferential nouns, i.e. two words or phrases which refer to the same idea or entity, with their antecedents was not. With the exception of one participant, all of the readers were able to integrate at least 50% of the target nouns with their antecedents during initial reading, and all were able to do so during the anaphor resolution task in the second part of the interview. This suggests that the primary assumption which underpinned the design of this project, namely that the ability to identify incongruent words can be taken as a proxy for the ability to integrate a coreferential nouns its antecedent, is incorrect, at least in absolute terms.

Rather, the analysis suggested that the ability to identify incongruent words in a text depends on the extent to which a reader is able to establish a situation model representation of the meaning of the text, or at least a macrostructure representation of its main topics. It was found that there was a relationship between a reader's ability to process the language of the text efficiently and their ability to establish a situation model representation of its meaning. This explains the general relationship between proficiency and the ability to identify the incongruent words. However, here too this relationship was found not to be absolute. There were two occasions when readers who had experienced substantial difficulties processing

the language of the text showed evidence of having established such strong situation models that they suppressed the meaning of an incongruent word entirely in favour of the congruent meaning. In addition, it was also found that when one of the participants engaged in skim reading, even though they did not encounter difficulty with linguistic processing, this still only resulted in the establishment of relatively shallow, text-focused representations of the text that were insufficient to allow the incongruencies to be detected.

Finally, it was also found that although the establishment of a macrostructure representation of the text did at times allow a reader to identify an incongruency, this was not always the case. There were also instances where although the establishment of such a representation did allow for a coreferential noun to be integrated with the antecedent, the incongruent word was not identified. Furthermore, the findings also suggested that when readers were only able to establish propositional or even microstructure representations, they showed no evidence of having identified the incongruent word. Thus, the evidence suggests that the establishment of textbase discourse representations at the local level is insufficient to allow a reader to identify incongruous words in a text. Moreover, even at the more global level, such text focused representations are not always sufficient to make an incongruency salient to the reader. Rather, the predominant requirement for an incongruent word to be identified is the establishment of a situation model element in the reader's discourse-level representation of the text. Hence, the findings from the analysis reported in this chapter suggests that assumption 1 should be rewritten as:

> **Revised assumption**: A reader's ability to detect incongruent words in a text can be taken as a proxy for their ability to **establish situation model representations** of the meaning of that text

Discussion of these findings, their relationship to the existing literature on second-language reading comprehension, as well their implications for both theory and practice are provided in the following chapter.

Chapter 6: Discussion

6.1 Introduction: A brief summary of the study

The study reported on in this thesis set out to investigate the extent to which native-Chinese speaking readers of English language academic texts are able to integrate coreferential nouns phrases (i.e. two or more noun phrases which refer to the same idea or entity) with their antecedents in the text. The issue of noun phrase integration is of particular concern in the field of academic reading. Analysis of the linguistic construction of scientific texts by Halliday (1993a) has shown that much of the work of explaining and developing academic concepts is achieved through the use of highly complex coreferential noun phrases. Similarly, Francis (1994), explained how such coreferential noun phrases can be used as retrospective labels, which serve a vital cohesive function in academic texts by encapsulating, and frequently evaluating, the main idea in one section of text, such that it may form the ideational starting point for the next. Writers use these noun phrases to repeat ideas or concepts in a text so that information can be added to them as the discourse unfolds. Therefore, in order to understand academic texts, it is frequently necessary for readers to be able to recognise when a noun phrase is referring to a concept which has already been introduced. In psycholinguistic terms, the ability to integrate coreferential noun phrases with their antecedents in this way is called **anaphor resolution** (Kintsch & Rawson, 2005).

However, analysis by Halliday (1993b) suggests that, due to the considerable differences which exist between the Chinese and English languages, these types of noun phrases may be particularly difficult for native-Chinese speaking readers of English to understand. This is because the requirement in academic writing to pack a lot of information into individual noun phrases often means that, in English, the main noun in the phrase is subject to a considerable degree of **post-modification**. An example of such post-modification can clearly be seen in the following noun phrase (taken from Halliday, 1993a), in which that the main noun, *importance*, has been post-modified by no fewer than three prepositional phrases.

the **importance** (*i*) of **surface chemistry** (*ii*) in the **mechanical behaviour** (*iii*) of **brittle materials**

The grammar of Chinese, on the other hand, does not allow for post-modification of nouns at all. Rather, in Chinese all nouns must be **pre-modified** (see Section 2.2.2 for an example)**.** This is to say that all modification must come before the noun, as, for example, with adjectives in English. It may therefore be the case that native-Chinese readers struggle to integrate the components of complex post-modified English noun phrases well enough to establish their unified meaning. Given the important cohesive role these types of coreferential noun phrases frequently play in English academic writing, it may be that native-Chinese speakers are placed at a considerable disadvantage when reading, and attempting to learn from, academic texts.

That such a risk may exist is further suggested by a study by Pretorius (2005), which found that the ability to resolve anaphors in academic texts is highly correlated with academic outcomes at the university level. More worryingly still, her results also suggested that the types of complex coreferential noun phrases frequently found in academic writing were particularly difficult for L2 readers of English to resolve when compared with simpler anaphors, such as pronouns. Perhaps unsurprisingly, given the linguistic complexity of these noun phrases, she also found that both the ability to resolve anaphors, as well as academic outcomes, were positively correlated with L2 language proficiency. As such, given the centrality of reading in UK university education, it is vital to understand to what extent native-Chinese speaking readers are able to integrate coreferential noun phrases in English language academic texts.

6.2 A summary of the key assumptions, research questions and main findings of this study

The research design for this study was built on two key assumptions:

Assumption 1: A reader's ability to integrate a coreferential noun phrase with its antecedent, in order to form discourse-level representations, can be tested by investigating their ability to identify a word in a coreferential noun phrase which is incongruent with an idea expressed in the antecedent.

> **Assumption 2**: The presence of a determiner in a coreferential noun phrase ought to help a reader integrate that noun phrase with its antecedent, and so to identify any incongruent words that noun phrase may contain.

The first of these assumptions was based on a study by Walter (2007), which showed that lower-intermediate second-language readers of English are less able than upper-intermediate readers to detect anomalous words which contradict main points previously made in the text. Drawing on the Structure Building Framework (Gernsbacher, 1997), Walter argued this discrepancy exists because lower-intermediate readers are unable to access cognitive "structure building processes" (2007, p. 29) which allow skilled readers to integrate new information into their developing discourse-level representation of the text. Walter (2007) did not directly address the issue of noun phrase integration. However, given the centrality of coreferential nouns phrases in creating coherence in academic texts (Francis, 1994, Halliday, 1993a), it arguably follows that the ability to identify incongruencies embedded in such noun phrases may act as a proxy for the ability to integrate them with their antecedents to form discourse-level representations.

The second assumption was based on an analysis of cohesive devices in English by Halliday and Matthiessen (2014), which stated that grammatical determiners (e.g. *this*, *those*) serve as explicit cohesive markers in a text, indicating that the noun phrases they appear in are coreferential with an antecedent. Previous research (Van Silfhout et al., 2015) has found that another type of cohesive device, *connectives* (e.g. moreover, however), act as processing signals, indicating to the reader how the ideas in the text are logico-semantically related, and that this is beneficial for comprehension. Hence, it was reasoned that determiners may similarly act as processing signals, indicating to the reader that a noun phrase is coreferential, and so facilitating its integration with the antecedent.

In order to explore these issues, an eye-tracking experiment was conducted to address three research questions:

RQ1. To what extent are native-Chinese speaking readers able to identify words in coreferential noun phrases which are incongruent with an idea expressed in its antecedent?

RQ2. To what extent does the presence of a determiner in a coreferential noun phrase help native-Chinese speaking readers identify words in that noun phrase which are incongruent with an idea expressed in its antecedent?

RQ3. To what extent do native-Chinese speaking readers at different levels of reading proficiency differ (i) in their ability to identify incongruent words in coreferential noun phrases, or (ii) in the extent to which they are helped to identify incongruent words in such a noun phrase by the presence of a determiner?

Firstly, the initial results of this experiment suggested that, overall, the participants were able to identify the incongruent words in the text, as indicated by longer processing times of the main nouns in the target noun phrases. However, once the issue of reading proficiency was accounted for, this effect was only seen among the upper-intermediate readers in the sample. In order to further investigate this finding, multiple regression analyses were conducted to see which, if any, individual differences in terms of *vocabulary knowledge*, *grammar knowledge*, *working memory capacity in English* or *working memory capacity in Chinese* might predict the ability to identify incongruent words in a text. However, these analyses did not return any statistically significant results.

Secondly, it was also found that the presence of a determiner in a coreferential noun phrase did *not* help readers identify the incongruent word, or indeed have any effect on integration processing at all. As such, assumption 2 was shown to be incorrect. Rather, the evidence suggested that the presence of a determiner serves only to facilitate initial word access processing, and that this effect fades extremely rapidly. Hence, although the results suggested the ability to identify the incongruent words in the texts was confined to the upper-intermediate readers, they gave no insights into why this may have been, or what had allowed them to do so. In order to further investigate these issues, think-aloud interviews were conducted to shed light on one more question:

RQ4. Why were the upper-intermediate readers able to identify the incongruent noun phrases while the intermediate and lower-intermediate readers were not?

The answer to this question was complex. Initial analysis of the transcripts from the think-aloud interviews supported the finding that, in general, only the upper-intermediate readers were able to identify

the incongruent words. However, they also showed that although this was the general trend, it was not absolute. Specifically, one of the readers from the lower-intermediate group, and two from the intermediate group, also identified one incongruent word apiece. There were also five occasions where upper-intermediate readers did not identify an incongruent word, suggesting proficiency alone is not enough to ensure comprehension monitoring takes place.

Furthermore, the findings from the think-aloud interviews also showed that while only the upper-intermediate readers were able routinely to identify the incongruent words, the vast majority of the participants, at all proficiency levels, *were* able to integrate at least 50% of the coreferential nouns phrases they read with their antecedents. From the total of forty-eight target noun phrases read in the think-aloud interviews, in thirty-two instances there was clear evidence that the main noun had been integrated with the antecedent, five instances where no clear determination could be made, and only eleven where integration clearly did not occur. In contrast, it was found that there were only thirteen cases where a reader clearly identified an incongruent word. As such, although there were a number of reasons why the incongruent words were not detected, it was nonetheless found that there was little correspondence between the ability to integrate a coreferential noun and the ability to detect an incongruency in the text, at least in absolute terms. This suggests that Assumption 1 is not correct in all circumstances.

Rather, it was found that the key determinant of whether or not a reader identified the incongruent words was the type of mental representation they were able to establish of the preceding text. It was explained in Section 2.4 how Kintsch's (1998) Construction-Integration model of reading comprehension posits two main elements which may be present in a reader's discourse-level representation of a text: the *textbase* and the *situation model*. The evidence from the current study suggested that of these two, it was only the establishment of a situation model element which reliably allowed a reader to identify an incongruency in the text. With regards to the textbase, while on occasions the establishment of a macrostructural representation allowed a reader to identify an incongruency in the text, in at least one of

these there was also clear evidence to suggest the reader had also established a situation model element in their representations. Furthermore, there were also a number of occasions where the establishment of such a representation did not result in the incongruent word being noticed. Consequently, the evidence suggests that macrostructure formation alone does not reliably result in the ability to identify discourse-level incongruencies. There were no occasions when the establishment solely of a microstructure representation resulted in an incongruent word being detected. This led to a revision of Assumption 1:

> **Revised assumption**: A reader's ability to detect incongruent words in a text can be taken as a proxy for their ability to **establish situation model representations** of the meaning of that text

Collectively, these findings have raised a number of issues which are of relevance to both the theory and practice of reading in a second language at university level. The following discussion will address four specific themes: (*i*) *reading comprehension and learning from text*; (*ii*) *the role of individual differences in second-language reading comprehension*; (*iii*) *the role of grammatical determiners in processing coreferential noun phrase*; (*iv*) *comprehension monitoring, incongruency detection and the question of access*.

6.3 Reading comprehension and learning from text

The research reported in this thesis was motivated by concerns that many native-Chinese speaking students enrolled in English speaking universities may be unable to fully profit from their courses if, due to insufficient linguistic proficiency, they struggle to read the texts they encounter (Trenkic & Warmington, 2018). The results of the current study suggest this concern is well founded. The focus of the study was on how native-Chinese speakers process coreferential noun phrases in academic texts. This was for two reasons. Firstly, coreference is one of the major resources which allow readers to establish coherent discourse-level mental representations of a text (Kintsch & Rawson, 2005). Secondly, as already discussed, one of the primary linguistic features which characterises written academic discourse is the use of complex, heavily post-modified coreferential noun phrases (Francis, 1994; Halliday, 1993a).

Unlike in other areas of second-language reading instruction, within the field of English for Academic Purposes (EAP) the primary focus is on *reading to learn* as opposed to simply *learning to read* (Jordan, 2007). Consequently, students arguably need to achieve a higher level of reading comprehension than may be the case in other contexts of learning. The results from the current study suggests that for many lower-intermediate and intermediate level readers this level of comprehension may be difficult to attain. It was argued at the end of Chapter 5 that the assumption that the ability to identify incongruent words in a text can be taken as a proxy for the ability to integrate a coreferential noun phrase with its antecedent was shown to be false. Rather, the evidence suggests that the ability to identify incongruent words in a text can be taken as a proxy for the ability to establish a situation model bearing representation of the text. In this regard, the results from the current study support the findings from Ushiro et al., (2017) who found that the establishment of a situation model was required for second-language readers to identify incongruencies in narrative texts. This distinction is potentially of great relevance when considering to what extent students are able to learn from the texts they read. The evidence from both the eye-tracking experiment and the think-aloud interviews suggested that only the upper-intermediate readers were routinely able to identify the incongruent words in the text, or to establish situation model representations. As previously explained, Kintsch's (1998) Construction-Integration model of reading comprehension argues that readers may establish two discourse-level elements in their representation of a text: the *textbase* and the *situation model*. The textbase comprises the information directly encoded in the language of the text, while the situation model is that part of the representation which the reader already possesses in long-term memory once activated by, and integrated with, the information in the text. Although it is possible to establish a representation composed solely of textbase elements, or primarily of a situation model, a complete representation of the text must contain *both* elements. Importantly for the field of EAP reading comprehension, Kintsch (1998, p. 295) argues that "learning from text... requires the formation of a situation model", where learning is defined as "the ability to use the information required from [a] text productively in novel environments" (p. 290). The prerequisite for such learning, he argues, is that "the

information in the text be integrated with the reader's prior knowledge and become a part of it, so that it can support comprehension and problem solving in new situations" (p. 290). Support for this view was provided by Raudszus et al. (2019), who found that the establishment of a situation model uniquely predicted for the reading comprehension in both first and second-language 4th grade children. Their definition for what constituted the establishment of a situation model was a "mental model of a text [that] consists of concepts and the relationships between them (Broek & Helder, 2017; Rapp et al., 2007; Zwaan, 2016)" (p. 108). This emphasis on establishing a mental representation of not only the concepts contained in an expository text, but also the relationships between them, arguably constitutes one of the bedrocks of learning. If, as the findings from the current study suggest, only readers at an upper-intermediate or higher level of reading proficiency are routinely able to establish complete, situation model bearing representations of the texts they read, it may be that readers below this level of proficiency will struggle to fully acquire information presented to them solely in written form.

That this may be the case is further supported by Pretorius (2005), who found a significant statistical relationship between the ability to resolve coreferential noun phrases in academic texts and levels of academic achievement. It is important to note that Pretorius' study did not investigate the issue of anomaly detection, focusing only on the integration of coreferential anaphoric noun phrases. The results of the current study have indicated that these two issues, although frequently related, are not the same thing. The evidence suggested that the ability to integrate the coreferential nouns with their antecedents to form a textbase representation did not necessarily result in the ability to identify the incongruent words. However, it was also found to be the case that the ability to identify an incongruent word *did* depend on the ability to integrate the coreferential noun phrase to form at least a local microstructural textbase representation. That is to say, the establishment of a textbase was a necessary, but not a *sufficient* condition for an incongruent word to be identified. A situation model was also required. Hence, it may be that the relationship Pretorius found between the ability to integrate coreferential noun phrases and her participants' learning outcomes

resulted from the higher-proficiency readers in her sample being able to establish complete representations of the texts read, composed not only of strong textbase representations of the text, but, in part consequence, situation model elements as well. That this may have been the case is supported by the fact that Pretorius (2005) also found positive correlations between both the ability to integrate coreferential noun phrase as well as academic outcomes on the one hand, and second-language linguistic proficiency on the other. The analysis of the think-aloud interviews from the current study found that the major impediment to situation model formation was difficulties processing the language of the text, thereby adding support to the view that "problems in the lower-level processes of lexical access and syntactic integration [can] lead to poorer higher-level integration" (Raudszus et al., 2019, p. 107). As such, it may be that the relationship which Pretorius (2005) found between the ability to integrate coreferential noun phrases, academic outcomes and L2 linguistic proficiency may reflect the role of linguistic proficiency as the mediating factor between the ability to establish mental representations of a text and the ability to learn from those texts.

6.3.1 Possible compensatory mechanisms that might scaffold learning

Although the findings from the current study suggest it may only be upper-intermediate readers who have the ability to routinely establish situation models of, and thus to learn from, academic texts, there was also evidence to suggest that the situation is not beyond remedy for lower-proficiency learners. Firstly, Kintsch (1998) argues that although readers must establish both textbase and situation model elements in their representation of a text in order to comprehend it completely, where they have strong prior knowledge of the topic being discussed, readers may be able to overcome difficulties establishing a coherent textbase representation by relying on prior knowledge. Evidence to support this claim was found in the current study. On two occasions, despite having experienced considerable difficulty with linguistic processing of the antecedent text, lower-proficiency readers showed evidence of having used relevant background knowledge of the topic under discussion as a compensatory mechanism (Bernhardt, 2010; Stanovich, 1980) to establish

strong situation models. In both instances, the evidence suggested that the situation model they established was strong enough not only to allow them to integrate the coreferential noun in the target noun phrase, but also to suppress (Gernsbacher & Faust, 1991) the contextually incongruent word in their final representation of the text's meaning at that point. On the one hand, it may be argued that a reliance on prior knowledge to overcome linguistic difficulties may be of little help in the early stages of a student's academic career, where they are likely to be confronted with many new ideas. However, as they proceed through their courses, it is also likely that their stock of knowledge about the subjects they read about will grow. This means that the ability of lower proficiency readers to compensate for difficulties establishing coherent textbase representations by means of establishing situation model representations is likely to grow in tandem (Smith et al., 2021)

Furthermore, although Walter (2004; 2007) concluded on the basis of her findings that lower-intermediate readers are able to establish sentence-level but not discourse-level representations of the texts they read, the findings from the current study do not support such a conclusion, at least in its entirety. Rather, it was found that readers at all levels of proficiency showed some ability to establish at least localised microstructure representation of the text read, and at times even lower-intermediate readers displayed evidence of having established more global macrostructure representations. This finding may have important implications for learning, at least in the long term. Although Kintsch (1998) argues that learning *per se* can only occur when a reader has established a situation model representation, it is important to note that he draws a clear distinction between *learning* and *text memorisation*. As stated above, his definition of learning requires that a reader be able to use the information they acquire productively to solve problems in novel situations. As such, he sets the bar for learning quite high. Text memorisation, on the other hand, he defines as "the ability to reproduce the text verbatim, in paraphrase or by summarising it", arguing that this "may be achieved on the basis of only superficial understanding" (p. 290). Such text memorisation, he states, is possible purely from the establishment of textbase representations. While it may be argued that simple

memorisation of a text is inferior to actual learning, within Kintsch's strict definition, and perhaps most of all in the context of university education, it is undoubtedly preferable to no learning at all. Given that much study at UK university is made up of a combination of independent reading and teacher-led seminar discussions, in many circumstances it may be that the ability to establish memorable, even if incomplete, textbase representations of the texts they read may provide second- language readers with at least some foundation for future teacher supported learning.

6.4 Individual differences

Much research in second-language reading comprehension has focused on the role of individual differences in terms of vocabulary knowledge, grammar knowledge as well as working memory capacity (e.g. Jeon & Yamashita, 2014; Joh & Plakans, 2017; Yamashita & Shiotsu, 2017; Zhang, 2012). The following discussion will address the ways these three issues impacted upon the ability of the participants in this study to identify the incongruent words in the texts read.

6.4.1 Vocabulary knowledge

Although the results for the vocabulary test used in the present study correlated strongly (r = .7) with reading proficiency, as measured by IELTS reading scores, suggesting the test used was valid, they were not found to be predictive of the participants' ability to identify the incongruent words in the text. There are several reasons why this may have been. Firstly, the Schmitt et al. (2001) vocabulary test measured breath of knowledge, i.e. the number of words the student knows. However, research into second-language reading has suggested that in addition to breadth of vocabulary knowledge, it is also important to consider depth of knowledge (Nation, 2008; Proctor et al., 2012; Raudszus et al., 2019), i.e. how much a student knows about each word in terms of such issues as morphological form, collocational behaviour and contextually dependent variations in meaning (Koda, 2005; Nation, 2013) . As such, given that the ability to identify the incongruent words in the text depended principally on the ability to establish a complete, sophisticated

mental representation of the meanings it contained, it could be that this required the readers to have considerable depth of knowledge of the vocabulary it contained.

Secondly, when analysing the features which are most likely to cause difficulty to native-Chinese speakers when reading academic texts, Halliday (1993b) identified both lexical density in general, as well as a preponderance of low frequency, subject-specific technical words as potential obstacles. As such, it may be of relevance to note that the Schmitt et al (2001) test was designed to test for knowledge of general vocabulary only, albeit with one section testing for knowledge of Coxhead's (2000) Academic Word List. Consequently, it would not have been capable of measuring the participants' knowledge of the subject-specific vocabulary which was contained in the texts used in the present study. That this may have been the case is to some extent supported by the findings from the think-aloud interviews. As previously stated, the major impediment to establishing the situation model representations needed for the readers to identify an incongruency in the text was difficulty with linguistic processing. Analysis of the transcripts indicated that such difficulties often occurred as a result of participants encountering unknown or unfamiliar words. Importantly, even when they were able to successfully regulate their comprehension of the unknown items, this often resulted in the establishment of weak, text-focused representations which did not allow the reader to identify the incongruent words in the target noun phrases. The use of top-down, context-based compensatory strategies as an aide to dealing with unknown vocabulary in a text has been widely promoted within the fields of second-language vocabulary acquisition in general (Nation, 2013) and in second-language reading in particular (Grabe, 2009). However, the results from the current study suggest that although such strategies may be sufficient to overcome local comprehension difficulties well enough to establish propositional and even microstructure textbase representations, they may not be adequate to allow readers to establish complete, situation model bearing representations of academic texts of the type required for deep learning to occur.

6.4.2 Grammar knowledge

Similarly to the vocabulary test, despite correlating with the participants' IELTS reading scores (*r* = .6), the scores from the grammar test was not predictive of their ability to identify the incongruent words in the text. Once again, there are several reasons why this could have been. The test used to measure the participants' grammar knowledge was based on DeKeyser's (2000) Grammaticality Judgement Test, and tested knowledge in five categories: *past tense*, *progressive aspect*, *articles*, *pronominalisation*, and *word order*. As explained in Section 3.6, during pilot testing the results of reliability testing resulted in the removal of eight items from the grammar test. However, constraints of time did not allow for the resulting test to be examined for reliability again. The fact that the grammar test correlated strongly with the participants' IELTS reading scores suggest that it was indeed a valid measure of grammar knowledge. However, inspection of the five categories it contained reveals that it was also rather a general test of grammatical knowledge, while the analysis of the think-aloud interviews suggested the participants' ability to identify the incongruent words was impaired by one main grammatical issue; complex post-modified noun or verb phrases. Knowledge of this area of grammar was not included in the test. Hence, it may be that the grammaticality judgement test correlated highly with the participants' IELTS scores because the higher proficiency readers also had a stronger grasp of general grammar, irrespective of their level of reading proficiency. However, because it did not explore the readers' knowledge of, and thus ability to process, complex post-modified grammatical structures, it was unable to predict their ability to identify the incongruent words in the text. The effect of sentence complexity on readability has been the source of controversy within second-language reading comprehension studies for some time (Koda, 2005). However, the current study lends weight to Halliday's (1993b) suggestion that native-Chinese readers may struggle to process the kind of complex, post-modified noun phrases frequently used in academic texts. Furthermore, the analysis of the think-aloud transcripts suggest that this issue may even extend to post-modified verb phrases, thereby further adding to their difficulties.

6.4.3 Working memory capacity

Walter (2007) reported strong statistical relationships between working memory capacity and the ability to detect anomalies in a text. The results from the current study did not support this finding. The participants' scores on the working memory test did not predict their ability to identify the incongruent noun phrases. There may be several reasons why this was. Firstly, Walter used a word span test to measure working memory capacity. By contrast, the current study used a digit span test. This was done because the participants in the study were spread across a wide range of reading proficiency levels, and concerns have been raised that word span tests may not provide a measurement of working memory which is entirely separate from reading comprehension (Kintsch, 1998; Oakhill et al., 2005). Consequently, it was hoped that the use of a digit span test would provide a measure which was more comparable across reading proficiency levels. However, it may be that if a word span test had been used, the results would have been different.

The second difference between Walter (2007) and the current study is that in testing the relationship between working memory capacity and the ability to identify incongruent words, Walter separated her participants by proficiency level. The results suggested that the lower-intermediate readers had lower L2 working memory capacity than the upper-intermediate readers, and that this may have affected their ability to identify the incongruent words in the text they read. However, in the current study the participants were not separated by proficiency level in the regression analysis. This was because the sample was not large enough to maintain a sufficient degree of statistical power if they were split, and there was no way to cross reference the results with the findings from the think-aloud study, as there was when the sample was split by proficiency regarding the ability to identify the incongruencies in the text. If proficiency had been factored for in terms of the impact of working memory the results may have been different (Brunfaut et al., 2021). In addition to Walter (2007), several studies on second-language reading comprehension have found differences between the role of working memory in the reading comprehension of higher and lower proficiency readers (Joh & Plakans, 2017; Juffs & Harrington, 2011; Peng et al., 2018; Walter, 2004). Of

particular relevance to the current study and its claims about the importance of situation model representations are findings that the role of working memory (WM) may be directly related to the amount of prior knowledge the reader has about the topic being read. For example, Joh and Plakans (2017, p. 115) argued that their results:

> support the premise that the contribution of WM to L2 reading comprehension varies as a function of readers' topic knowledge. More specifically, WM seems to facilitate L2 reading comprehension *only* when a reader has enough knowledge of topic-related vocabulary in a given text [emphasis added]

Similarly, a study by Alptekin and Erçetin (2011) found that readers with both high and low working memory capacity performed similarly on what they termed *literal* comprehension, but that those with lower working memory capacity performed significantly worse in terms of *inferential* comprehension. Interestingly, in defining what they meant by literal and inferential comprehension, Alptekin and Erçetin (2011) also made direct reference to Kintsch's (1998) Construction-Integration model. Literal comprehension they defined in relation to textbase representations, arguing that to establish such a text-focused representation, the reader only needs to comprehend what was literally meant. Inferential comprehension, on the other hand, they defined relative to situation model representations, i.e. those which contain information from the reader's prior knowledge. In making this claim, they cite Kintsch's (1998) argument that in order to make elaborative or bridging inferences the reader must fill in what is implied but unstated in the text with information they already hold in long-term memory. Given that in the current study, with three notable exceptions, only the upper-intermediate readers were able to identify the incongruent words in the text as a result of establishing situation model bearing representations of the text, it may be that this resulted at least in part from relative differences in their levels of working memory capacity.

Such a conclusion would fit to some extent with the Construction-Integration model, although it may be that nuance needs to be added to what is meant by working memory capacity. Drawing on a paper by Ericsson and Kintsch (1995), Kintsch (1998) argues that the establishment of situation model representations

involves what he calls *long-term working memory*, which refers to the reader's ability to activate and integrate information in long-term memory as cued by the words in the text. Notably, the evidence from the current study suggested it was the establishment of situation model representations that allowed a reader to identify the incongruent words in the text. Hence, it may be that the use of a digit span test of short-term working memory capacity was not measuring the appropriate cognitive ability. Had a test of long-term working memory capacity been used, it may have yielded different results.

However, another potential explanation for why working memory capacity was not predictive of the participants' ability to identify the incongruent words in the text is simply that working memory does not play a determinant role in such an ability. Johns et al (2015, p.15) argue that "poor comprehenders do not *de novo* employ a qualitatively different memory mechanism than that used by good comprehenders", and that differences in comprehension ability depend more on "differences in representational quality rather than on retrieval speed". It is important to note that by representational quality, Johns et al. refer specifically and only to representations of lexical knowledge, rather than to representations more broadly. However, it may be the case that the quality of the mental representations that readers are able to establish of the information in a text has an impact over and above their working memory capacity. It was argued above in Section 6.4.1 that a test measuring depth of vocabulary knowledge rather than simply breadth may have been better able to predict a reader's ability to identify incongruent words in a text. In Johns et al.'s (2015) terminology, depth of vocabulary knowledge is arguably synonymous with quality of representation. If it is the case that higher quality lexical representations are an important determinant of successful reading comprehension, rather than simply differences in memory retrieval, it may equally be the case that the quality of the resulting discourse-level representations could also result in improved comprehension, and so the ability to identify incongruent words in a text.

6.5 The role of grammatical determiners in processing coreferential noun phrases

The second key assumption which underpinned the research reported in this thesis was that the presence of a determiner in a coreferential noun phrase, as an explicit linguistic marker of cohesion, would aid the reader in integrating that noun phrase with its antecedent. This assumption was based on research in both first and second-language reading studies (e.g. Degand & Sanders, 2002; Ozono & Ito, 2003; Van Silfhout et al., 2015), which has suggested that connectives, words such as *because* or *therefore*, act as processing instructions. The role of such processing instructions is to indicate to the reader how ideas in a text are logico-semantically related, thereby aiding integration processing. It was reasoned that if connectives, as a form of specific cohesive marker, aided integration, then the presence of a determiner, as another explicit marker, should perform likewise. However, this prediction was shown not to be correct. Not only did the presence of a determiner not help the readers to identify the incongruent words in the texts read, as suggested by the lack of interaction between the determiner and congruency factors, the evidence also suggested that in fact they played almost no role in integration processing whatsoever. Analysis of the eye-tracking data suggested that the determiner only affected initial word access processing, as indicated by the fact that significant effects were only found in the first fixation duration, and faded extremely rapidly thereafter.

A re-examination of the theory which gave rise to this assumption may cast light on why this was. As explained in Section 2.1, researchers in both linguistics (Ariel, 2001) and psycholinguistics (Sanders & Gernsbacher, 2004, p. 85) have argued that anaphors, as explicit linguistic markers of coreference:

> are [processing] instructions to connect incoming information with already mentioned referents, and the referent nodes can be more or less accessible [in working memory]. As a result, it takes more or less processing time to understand anaphors

The initial assumption that a determiner in a coreferential noun phrase would aid integration processing was based on this view of accessibility. If the determiner indicates to the reader that the information in the noun phrase is coreferential with an earlier section of text, it was reasoned, this ought to help the reader

connect the new information to that already established in their mental representation of the text. However, the results suggest that although determiners do act as processing instructions, they perform a different role than that played by connectives. The fact that the presence of a determiner aids initial word access processing rather than later integration processing suggests that their role may fall more in the *construction* than the *integration* phase of discourse processing. According to Kintsch's (1998) account of discourse processing, the initial construction phase is governed by rules which "are weak, dumb and do not discriminate between what is contextually appropriate and what is not" (p. 95). During this phase, which occurs when a reader first fixates a word, all possible meanings of the word are activated. Kintsch suggests that this early phase of processing is by nature chaotic, and that order is not imposed on the reader's mental representation of the text until a secondary round of integration processing occurs. During this secondary phase, the contextually relevant meaning of the word acquires enhanced activation and irrelevant meanings are quickly suppressed (Gernsbacher, 1989), as the word is integrated with information which was previously activated when the reader processed earlier sections of text. Importantly, Kintsch (1998) argues that during the construction phrase, word access works in the same way regardless of whether or not a word is repeated in a text. For example, he cites evidence which suggests that when readers process repeated homographs, words which share the same spelling but have different meanings (e.g. river **bank** and investment **bank**) in a text, all possible meanings of the repeated word are activated each time it is fixated. It is not until the ambiguous word is integrated with other words which are active in memory, such as **tree** in the first case, or **money** in the second, that the contextually appropriate meaning is established, and the inappropriate meaning suppressed (Duffy et al., 1988).

However, Kintsch (1998) goes on to state that although repeated words do not enjoy an initial processing advantage over non-repeated words, there *is* an important difference in the way they are processed. Specifically, he suggests that "part of the meaning construction process for the repeated words involves retrieving the context in which the word appeared earlier and incorporating it into the present context" (p.

145). In discussing the specific case of anaphors, as a type of coreferential noun phrase, he cites evidence which suggests that "explicit anaphora are rapidly identified with their prior reference, reactivating in the process, at least temporarily, other related concepts from the prior context [so that] in the construction of their meaning the context of their prior appearance is integrated" (pp. 146-147).

This question of context may be of particular importance when processing the kind of coreferential noun phrases which are frequently found in academic writing. As explained in Section 2.2 much of the work of creating coherence in scientific texts is done by the use of complex coreferential noun phrases which allow the writer to repeat and recycle ideas as the discourse unfolds. One of the features which typifies much academic writing in such circumstances is the use of **shell nouns** (Gray, 2010). Shell nouns are a broad class of abstract nouns, including such words as *issue, problem, solution* or *advantage,* which depend for a large part of their meaning on their surrounding context (Halliday & Hasan, 1976). This surrounding context is in part provided by the preceding discourse (Kintsch, 1998). However, in the case of complex noun phrases, part of the semantic context is also provided by postmodification of the main noun in the phrase, by the use of prepositional phrases or relative clauses. While not all of the target nouns in the current study were shell nouns, in a strict sense, in the majority of instances they did need context to add specificity to their meaning. For example, although words like *technique*, *disease* or *symptom* may locate the reader in a particular semantic field, there are nonetheless a wide range of possible specific referents to which each of these words could refer. Part of the specificity required to fully construct the meaning of such words is provided by the preceding discourse. Hence, it may be that the reason the presence of a determiner in the target noun phrases resulted in faster initial word access processing of the target nouns is that determiners act as processing instructions which indicate to the reader what specific prior context needed to be reactivated during the construction process. Once this job is done, the role of the determiner, as an activation cue, comes to an end. The task of integration may then be cued by the processing of other linguistic items in the text, such as prepositions or explicit lexical terms acting as modifiers. That this may be the case is supported

by the fact that only the upper-intermediate readers displayed evidence of having derived benefit from the presence of a determiner in the target noun phrases. Moreover, they were also the only group that displayed evidence of having routinely established complete, situation model bearing representations of the text. If the assumption that the role of a determiner is to improve access to the initial context of use of a coreferential noun phrase, it is arguable at least that it would be necessary for the reader to have established a stable representation of that context in memory before the benefit of the determiner could be derived.

6.6 Comprehension monitoring, incongruency detection and the question of access

Finally, although the major focus of this project was not on comprehension monitoring, the Anomaly Detection Paradigm it adopted was based on a model frequently used to investigate comprehension monitoring in both first and second-language readers (Hessel et al., 2021; Morrison, 2004). Consequently, it was found that some of the results of the current study may shed fresh light on some of issues surrounding the use of error detection as a research tool used to investigate comprehension monitoring during online reading. Furthermore, they also lend weight to Walter's (2007) contention that second-language reading comprehension should be viewed more as a question of accessing fundamental cognitive comprehension processes through the second-language than of transferring reading strategies from the first.

The use of the error detection paradigm to investigate comprehension monitoring rests on a simple premise. The researcher embeds incongruent items in a text, and if the reader detects them, they are assumed to have successfully monitored their comprehension; conversely, if they fail to detect them, it is assumed that they did not (Morrison, 2004). Generally speaking, comprehension monitoring is defined as a metacognitive reading strategy, and much research has found a relationship between comprehension monitoring and successful reading comprehension (Grabe, 2009; Hessel et al., 2021; Morrison, 2004; Perfetti & Stafura, 2014) . However, it is important to note that the direction of causality between a reader's ability to monitor their comprehension and their level of reading comprehension remains unclear (Oakhill et al.,

2005). The results of the current study may shed some light on this issue, at least in so far as what allows readers to identify incongruencies in a text.

Studies which have investigated the ability of adult second-language readers (Morrison, 2004) and of EAL children (Hessel et al., 2021) to detect errors in a text have found that language background itself does not necessarily play the defining role. Specifically, Morrison (2004) found that regardless of whether the participants in her study were reading in their first or second language, they were only able to detect discourse-level errors, i.e. words which contradicted an idea previously stated in the text, at an average rate of 45.22% and 41.98% respectively. This difference was not found to be statistically significant. Furthermore, when reading in both their first and second languages, there were strong correlations between the participants' level of reading proficiency and their ability to detect errors. This suggests that the ability to detect discourse-level incongruencies in a text, of the type used in the current study, may be determined chiefly by the reader's level of proficiency in the language they are reading in, rather than whether that language is their first or their second. Similarly, when comparing the ability of native English speaking children and EAL children to detect contradictions in a text, Hessel et al. (2021, p. 170) concluded that while in both sets of students there was a clear relationship between reading proficiency and the ability to detect incongruencies, "differences in online reading between bilingual and monolingual samples are better understood as differences in vocabulary knowledge rather than a reflection of categorical differences between first and second-language speakers".

The current study in many ways supported both these sets of findings. It was found that there was a clear difference between the ability of the upper-intermediate readers and the two sets of lower proficiency readers to identify the incongruent words in the text, thereby supporting the link between reading proficiency and anomaly detection. Furthermore, although the results of the vocabulary test did not predict the participants' ability to identify the embedded incongruencies, the analysis of the transcripts from the think-aloud interviews did show that when readers experienced difficulties processing the language of the

text, they were typically, if not always, unable to identify the incongruent words. This supports Hessel et al's (2021) contention that vocabulary knowledge may be a key determinant in this regard. However, the current study goes beyond these previous conclusions to establish potential mechanisms as to *why* these relationships exist.

It has been argued here that the key determinant of whether or not a reader is able to identify an incongruent word in a text is the extent to which they are able to process the language of the text efficiently and to establish a situation model bearing representations of its meaning, a conclusion shared by Ushiro et al. (2017). If this is the case, it arguably follows that that a reader's ability to monitor their comprehension well enough to detect discourse-level incongruencies is a *result* rather than a cause of proficient reading comprehension. Previous studies which used an error detection paradigm to investigate comprehension monitoring potentially lend weight to such a conclusion. For example, in a study of comprehension monitoring in first language children, Oakhill et al. (2005) state that:

> Studies have found that, when untrained in inconsistency detection, poor comprehenders and young children are particularly poor at [detecting] internal consistency [i.e. between ideas directly stated in the text], which requires an individual to hold a representation of the text in his/her mind and to evaluate the level of consistency between that representation and incoming information. Of course, external inconsistencies (between the text and prior knowledge) also require the comparison of a representation of the text with one of prior knowledge, but such comparisons may be easier, at least for children, because a representation of prior knowledge would be expected to be more stable than one of information only recently acquired from a text.

In this interpretation, it is assumed that external inconsistencies are easier to detect than internal ones because the reader may have more stable representations of prior knowledge than for recently acquired information. However, if the determinant factor which allows a reader to identify an incongruency is whether or not they have been able to establish a situation model representation of its meaning, it may be that internal inconsistencies are harder to identify than external ones because there is an extra cognitive step involved. In the case of external inconsistencies, the reader is only required to establish a situation

model representation of one section of the text, which they can then compare, consciously or otherwise, with relevant prior knowledge. However, in the case of internal inconsistencies, they are required to establish at least *two* situation model representations, such that the incongruent information in the second representation may be noticed as contrasting with information already established in the first. The analysis of the think-aloud data from the current study suggested that when readers encounter difficulty processing the language of the text, although they may still be able to establish a textbase representation, they can be impeded from activating relevant prior knowledge in order to establish a situation model representation (Raudszus et al., 2019). In such a case, the evidence suggests, error detection is unlikely to succeed. As such, the potential to fail to identify internal inconsistencies in a text is at least double that for external inconsistencies, as there two occasions where the reader may prove unable to establish the requisite situation model representation.

The contention that the ability to detect errors in a text is dependent on the ability to establish a situation model of its reading is also supported by Rubman and Waters (2000), albeit with a slight reinterpretation of their results. In their study, the authors trialled an intervention designed to improve the ability of first language children classified as weak comprehenders to detect errors in a text. Rather than simply being asked to read the text, one group of participants were asked to construct a physical story board of the events in the story using plastic cut outs. The results suggested that when asked to construct a physical representation of the story whilst reading, the participants' ability to detect errors embedded in the text was significantly improved. Rubman and Waters concluded that the improvement resulted from the fact that the requirement to build a story board encouraged the children to focus more on the meaning of the text rather than simply on decoding its linguistic surface structure. However, more than this they also suggested that the use of the story board enabled the readers firstly to be more successfully in encoding the key propositions in the text, and secondly "forced" them to integrate those propositions together to build

their comprehension of the story (Rubman & Waters,2000, p. 510). From this, they concluded that their results:

> support the hypothesis that proposition integration is the key process in effective comprehension monitoring and that techniques that enhance that process are likely to improve children's comprehension. (p. 510)

Interestingly, this conclusion contrasts with those from the current study in one subtle yet important respect. Specifically, the findings from the think-aloud interviews indicated that although only the upper-intermediate readers were routinely able to identify the incongruent words in the text, readers at all proficiency levels displayed clear evidence of having integrated information in adjacent sentences to establish microstructure textbase representations. As such, this suggests that while proposition integration and the resulting establishment of textbase representations may be a necessary condition for incongruency detection to occur, it is not a sufficient condition simply in itself. What is needed is a situation model representation. Consequently, in Rubman and Waters' study, it may be that having the participants construct a story board whilst reading resulted in them visualising the story in 'real world' terms, and this helped them to establish a clear situation model representation of its meaning as they processed the language of the text. It may be that it was the establishment of this situation model that improved the ability to notice errors in the text, rather than simply that they were better able to integrate the propositions it contained, as Rubman and Waters originally concluded.

6.6.1 The question of access to fundamental cognitive processes as the key driver of second-language reading comprehension

The extent to which Rubman and Waters' (2010) findings may apply to those from the current study is of course open to question. Not only were their participants first language children, they were also reading simplistic stories rather than complex academic texts. However, it may be of relevance to note that previous studies of comprehension monitoring in first and second-language readers, adults and children as well as of narrative versus expository texts have reported similar results (Hessel et al., 2021; Morrison, 2004). This

suggests that comprehension monitoring, as defined as the ability to detect inconsistencies in a text, may be a fundamental aspect of the cognitive process of reading comprehension. This idea is not new. Although Walter (2007) did not frame her study in terms of comprehension monitoring, as explained in Section 2.4, she did use an Anomaly Detection Paradigm to investigate her participants' ability to integrate information in successive sections of text to establish a coherent mental representation of the texts they read. The results of her study conformed to predictions made using the Structure Building Framework (Gernsbacher, 1997), a psycholinguistic theory of reading comprehension, which posits that the ability to establish coherent mental representations of the meaning of text rests upon amodal, language independent cognitive faculties which allow humans to establish coherent mental representations of the world around them from sensory input (Gernsbacher et al., 1990). From this, Walter (2007) suggested that rather than conceptualising the relationship between reading comprehension ability in a first and second language as being one of the conscious transfer of reading strategies, as has been traditional in the ESL/EFL literature (Grabe & Stoller, 2015) , it may be more accurate to consider the relationship as one of *access* of fundamental cognitive faculties, through the second language.

In this regard, Walter's views align with, but also bring some much needed clarity to the Linguistic Threshold Hypothesis (see Section 2.3.2) by explaining *why* exactly such a threshold may exist. Specifically, if it is the case that readers access fundamental comprehension processing through the second language, it stands to reason that they will need to reach a certain level of linguistic competence before this can be achieved.

Although Walter conceptualised access in terms of Gernsbacher's Structure Building Framework, the evidence from the current study suggests that this view may be equally applicable to the process of reading comprehension as proposed by Kintsch's (1998) Construction-Integration model. Firstly, data from both the eye-tracking experiment and the think-aloud interviews conducted for this study found that only the upper-intermediate readers were routinely able to identify the incongruent words in the text. Furthermore, as

previously explained, this ability resulted from the ability to establish situation model representations of the text. Analysis of the think-aloud data revealed that in order to do this, readers typically needed to have been able to process the linguistic surface structure of the text efficiently and with minimum need for comprehension regulation. This suggest that the ability to establish situation model representations was indeed accessed through the second language, once the readers had attained a sufficient level of linguistic proficiency. Moreover, even in the rare instances when lower-proficiency readers were are able establish situation model elements in their representation of the text and so identify an incongruency, this still resulted from them successfully processing at least one noun phrase in the text about which they held information in long-term memory. As such, although they were not able to access fundamental comprehension processing as frequently as the higher-proficiency readers, when they did so, they also did so through the second language.

The second piece of evidence which supports the view that second-language comprehension monitoring depends on the reader accessing fundamental cognitive faculties is the speed at which the upper-intermediate readers experienced processing disruption when they encountered an incongruent word, namely within 367 milliseconds. That processing disruption should occur so rapidly arguably suggests that it results from a highly automated response to difficulty integrating conflicting information into the reader's existing mental representation of the text. Although the present data do not provide an explanation for why disruption should occur at such speed, one potential cause may be provided by Graesser & Mello (2012), who examined the moment to moment emotions readers experience as they process a text. One of the emotions they highlighted was that of confusion, and argued that this results from the reader experiencing "cognitive disequilibrium [which] arises from obstacles to goals, contradictions, anomalous information, uncertainty, text cohesion gaps, and obvious gaps in knowledge" (Graesser & Mello, 2012, p. 239). Cognitive disequilibrium, also known as cognitive dissonance, results from a fundamental drive in human beings to maintain internal psychological consistency (Festinger, 1957). As such, it may be that when attempting to

integrate contradictory information into the reader's already established representation of the text, the consequent loss of coherence in their mental representations results in them rapidly experiencing an uncomfortable sensation of cognitive dissonance, a sensation which manifests within hundreds of milliseconds. If so, it may be that the attempts to reduce this dissonance by returning the reader's internal cognitive state to equilibrium are the psychological driver for the implementation of comprehension strategies such as rereading, at least in the first instance. Hence it is possible that what we refer to as comprehension monitoring in reading is the ability to access the cognitive processes whereby human beings monitor and maintain the internal coherence of their mental representations, at least in so far as anomaly detection is concerned.

Chapter 7: Limitations to the study

As with all research, this study was subject to a number of limitations which must be borne in mind when considering the conclusions presented in Chapter 8. These limitations relate to the measures of individual differences included in the study, the size of the sample, as well as to problems inherent to conducting mixed methods and cross-disciplinary research.

7.1 Tests of individual difference

With regards to the measures of individual differences, only the participants' vocabulary knowledge, grammar knowledge and working memory capacity were included as predictors. These three were included because vocabulary and grammar knowledge have consistently been found to be the strongest predictors of second-language reading comprehension (Bernhardt, 2010; Jeon & Yamashita, 2014; Joh & Plakans, 2017; Yamashita & Shiotsu, 2017; Zhang, 2012), while working memory capacity has been frequently argued to be a foundational component of both first and second reading comprehension (Alptekin & Erçetin, 2011; Brunfaut et al., 2021; Joh & Plakans, 2017; Kintsch, 1998; Peng et al., 2018; Perfetti & Stafura, 2014), as well as being associated with the ability to detect anomalies in a text (Oakhill et al., 2005; Walter, 2004, 2007). However, it is important to note that while these three elements have been found to be important predictors of reading comprehension, they are by no means an exhaustive list of *all* the predictors that have been identified as relevant to reading. Studies of both first and second-language reading have identified a host of other aspects which impact up on reading proficiency, including such issues as word decoding (Woore, 2009), rapid automatised naming (Cho & Chiu, 2015) or phonological awareness (Brennan & Booth, 2015), to name but a few. In the current study, tests of these issues were not omitted because they were thought unimportant, but simply because there was not time available in already lengthy data gathering sessions to administer the tests. Similarly, as argued in Section 3.6, even those tests which were used to measure the participants' vocabulary and grammar knowledge, as well as their working memory capacity each had their own limitations. Hence, although, as argued above, all the tests of individual differences used

in this study were reliable, valid instruments, the data they were able to collect was at least partially incomplete, and this may have affected the results.

7.2 Sample size

As argued in Section 3.4, while the sample used for the eye-tracking experiment included 82 participants, and was thus large enough to meet the demands of the Central Limit Theorem (Field, 2013), once this sample was split into three sub-samples to form three proficiency groups, the number of participants in each of three group was no longer large enough that the assumptions of the Central Limit Theorem could apply. It was in part for this reason that the results from the eye-tracking experiment were triangulated with the results of the think-aloud analysis. However, although it was found that the results from these two analyses did support each other, it is important to bear in mind that both the think-aloud analyses and the second round of analyses of the data from eye-tracking experiment were conducted on data obtained from a relatively small number of participants, and in consequence their generalisability cannot be guaranteed. All the findings and conclusions reached should be considered with that caveat in mind.

7.3 Mixed method and cross-disciplinary research

As explained in Section 3.3, the research reported on in this thesis used a mixed methods design. Furthermore, this study was also cross-disciplinary in nature, drawing on work from the fields of Second-Language Education, Applied Linguistics, and Psycholinguistics. Although the use of a cross-disciplinary approach to researching reading comprehension can bring valuable insights (Ariel, 2001a; T. J. Sanders & Gernsbacher, 2004), inevitably it comes with a cost. Just as concerns have been raised that the use of mixed methodologies can result in researchers failing to engage in sufficient depth with the various theoretical assumptions and limitations of the research methods they employ (Hesse-Biber, 2018; Mark, 2015; Mutch, 2009), the adoption of a cross-disciplinary approach, and the resulting need to cover a broad sway of literature, means that it is not possible to review each field in as much depth as would be the case if the study were grounded in a single discipline. While it is firmly believed that the literature reviewed for the

current study covered all relevant aspects and presents an up-to-date picture of the current state of knowledge in each field, it was necessary to be selective about which papers to include, meaning the review was not exhaustive.

Chapter 8: Conclusions and implications

The research reported in this thesis was motivated by a dual sense of optimism twinged with concern over the explosive growth in the international market in education, and in particular the rapid and ongoing increase in the number of non-native speaking students enrolling in UK universities. When the introduction to this thesis was first written back in 2018, the UK Council For International Student Affairs reported that Chinese students accounted for almost one third of all international students in the UK (ukcisa.org.uk, 2018). According to the Higher Education Statistics Agency, from 2018/19 to 2019/20 the number of Chinese students in the UK increased by a further 20%, bringing the total cohort to 101, 855 (https://www.hesa.ac.uk/data-and-analysis/students/where-from). It seems this is a trend that is set to continue. Maintaining the UK's strong position in the international education market is now a stated aim of the UK government, who aim to increase the value of British 'educational exports' to £35 billion by 2030, and increase the number of international students coming to the UK to 600, 000 over the same period (Hubble & Bolton, 2020). As stated in the introduction to this thesis, if the growth in the UK's share in the international education market is to be pursued in a way that is sustainable, responsible and ethical, it is vital that we provide those students with the linguistic support and resources they need to fulfil their maximum potential on the courses they choose. This means not only optimising the language support that students are given after they have begun their courses, but also striving to continually improve pre-sessional English language foundation courses taken before enrolment, as well as the various international English language testing programs used to distinguish between those applicants who have the requisite language skills to succeed, and those who do not. The findings from the current study have identified a number of potential problems which lower proficiency native-speaking Chinese readers of academic English may face when attempting to learn from English texts. However, as will be argued below, they may also be able to make some contribution to towards developing solutions to these problems. In the most practical sense, some of the findings from this study may contribute towards the further development of the practice of both

language teaching and language testing, as well as more institutional concerns. More theoretically, they may also contribute to the knowledge base surrounding second-language reading comprehension more generally. The following sections will discuss the implications of these study for each of these areas in turn. A number of recommendations will also be made for future research that may further investigate some of the issues developed in this thesis.

8.1 Implications for the *teaching* of reading comprehension for academic purposes

The following section discusses the implications of this study for teaching second-language reading comprehension.

8.1.1 Situation model formation and metacognitive strategy

It was argued in Section 6.3 that in curricula aimed at developing English for Academic Purposes (EAP) the focus of reading instruction must be on helping the students develop the skills necessary to *learn* from the texts they read. The results from the current study lend weight to Kintsch's (1998) contention that in order to learn from texts students must form situation model representations of the meanings of those texts (see Section 6.3 for a full discussion of this point). This argument has also found recent support in the work of Raudszus et al. (2019) who demonstrated that the ability to establish situation model representations of expository texts makes a unique contribution to the reading comprehension of both first and second-language children (see Section 2.4 for a full review of their study). Given that the current study was conducted with adult learners, this strongly suggests that this is a general facet of reading comprehension. Consequently, if native-Chinese speaking readers are to be helped to prepare for study in an EMI setting, or supported once they have started their courses, it is important that teaching time be devoted to helping them develop situation model representations of the texts they read. It is beyond the scope of the current thesis to suggest specific approaches that could be taken towards developing such an ability. However, a considerable amount of research has consistently demonstrated the utility of metacognitive strategy training in improving reading outcomes in second-language readers, and it may well be the case that teaching such

readers about the need to establish situation model representations as a metacognitive strategy would pay dividends. Developing such an approach could potentially offer a fruitful path for future reading strategy research.

8.1.2 Specific instruction on the grammatical construction and cohesive function of complex coreferential noun phrases in academic writing

One of the specific aims of the current thesis was to respond to calls from researchers such as Ariel, (2001) and Sanders and Gernsbacher (2004) who have encouraged research into reading comprehension to bring together work from both text linguistics and psycholinguistics so that we may gain a better understanding of how readers cognitively engage with the texts they encounter as lexico-grammatically *constructed* linguistic artefacts. In the field of second-language reading education such a need is perhaps more urgent still (Koda, 2005). In responding to this need, the current study drew heavily on the work of Halliday (1993a, 1993b) who analysed the lexico-grammatical construction of scientific texts written in English, as well as the difficulties that native-Chinese speaking readers may face as a consequence of the substantial differences which exist between the grammar of these two languages, most particularly in respect of pre and post-modification of nouns (see Section 2.2 for a full description of this issue). One of the areas Halliday noted as likely to cause particular comprehension difficulties for native-Chinese speaking readers of English is with the types of complex coreferential noun phrases that have been the main focus of the current thesis. The findings presented above suggest that Halliday's concerns were justified. The analysis of the think-aloud interviews (see Section 5.5.2.10) uncovered multiple occasions where a participant struggled to integrate the components of a complex noun phrase as a consequence of not realising the information in either a prepositional phrase or relative clause were modifying the main noun in the phrase. The inability to fully comprehend complex coreferential noun phrases of this type raises two potentially serious issues. Firstly, as Halliday (1993a) explains (see Section 2.2.1 for details), complex noun phrases in academic writing are often used to condense a lot of information into compact grammatical units so that writers can develop complex ideas in short sections of text. If native-Chinese readers find themselves unable

to process noun phrases of this type due to their grammatical construction this is likely to seriously impede their ability to comprehend, and thus learn from the texts they read. Furthermore, as Francis (1994) explains (see Section 2.2 for details), complex coreferential noun phrases in academic writing are frequently used as retrospective labels, serving to encapsulate and evaluate the information in previous sections of text, ranging from a single paragraph to a full chapter, as they develop their arguments throughout a text. One of the key features of such retrospective labelling is the frequent the use of semantically underspecified shell nouns such as *problem*, *issue* of *solution* (Gray, 2010). Understanding exactly what is meant by such words in their specific context of use often requires the reader to process the postmodifying elements of noun phrase they appear in. Consequently, if native-Chinese readers struggle to process these types of noun phrases, it is possible that they will not only experience difficulties processing individual sentences, but also to follow the overall chain of coherence used by academic writers to advance their arguments over extended sections of text.

Therefore, given the importance of these types of complex coreferential noun phrases in the construction of academic texts, it is important that time be spent teaching second-language readers in general, and native-Chinese speaking readers in particular, how they are grammatically constructed, and how they can be used to structure academic writing. Such lessons would arguably not only pay dividends in improving our students' academic reading comprehension abilities, but also help to improve their academic writing too.

8.1.3 Teaching about cohesive devices

The design for the current project was in part inspired by work in both first and second-language reading comprehension which has found that teaching students about conjunctions, a form of explicit cohesive device, brings comprehension benefit (see Section 2.5.1 for details). This is because words like *because*, *however* and *therefore* specifically indicate how the ideas in a text are logico-semantically related. It was reasoned from this, that the presence of a determiner in a coreferential noun phrase ought to similarly aid comprehension, by indicating to the reader that it is anaphoric in nature. However, the results of the study

were somewhat mixed in this regard. While it was found that determiners did bring some processing benefit, this was limited to upper-intermediate and above readers. Furthermore, the fact that the effect from the determiner was only seen in the early eye-tracking measures, suggests that determiners only have an impact on initial word access processing rather than on integration processing. If this is the case, the processing of determiners is most likely a highly automatised process. In consequence, whether or not these types of cohesive devices merit classroom time is open to question, beyond teaching about their potential role as part of complex coreferential noun phrases. Time devoted to teaching the various cohesive devices may be better spent on those which are more salient, and thus liable to be noticed whilst reading.

8.1.4 Teaching vocabulary and grammar

As regards teaching vocabulary and grammar, the findings from the present study were again somewhat mixed. As noted in Section 4.10, despite correlating highly with the participants' IELTS reading scores, suggesting they were valid measures at least to some degree, the scores from the vocabulary and grammar tests were not predictive of the participants' ability to identify the incongruent words in the eye-tracking experiment. However, the findings from the think-aloud analysis did indicate that the primary impediment to establishing a situation model representation of the text read was difficulties processing the language of the text. Importantly, even when the participants were able to successfully regulate their comprehension in the face of such linguistic difficulties, this frequently only resulted in them being able to establish textbase representations, rather than the situation model bearing representations needed for deep learning. In consequence, the results of this study suggest that one of the key requirements of second-language reading comprehension courses in general, and EAP course in particular, must be to help students develop their knowledge and command of the vocabulary and grammar of English.

8.2 Implications for the *testing* of reading comprehension for academic purposes

One of the key assumption which underpinned the study reported in in this thesis was that a reader's ability to integrate a coreferential noun phrase with its antecedent can be tested by investigating their ability

to identify an incongruent word in a coreferential noun phrase which contradicts an idea expressed in the antecedent noun phrase. This was shown to be incorrect. In light of findings from the study, this assumption was revised, to state that:

> A reader's ability to identify words in coreferential noun phrases which are incongruent with an idea expressed in the antecedent can be taken as a proxy for their ability to establish situation model elements in their discourse-level representation of the meaning of the text.

In the EAP context, when judging a student's level of reading proficiency, we are interested in measuring their capacity to *learn* from a text (Jordan, 2007). Given that the ability to learn from text depends to a large degree on being able to establish a situation model representation of a text (Kintsch, 1998; Raudszus et al., 2019), it is arguably the case that tests of reading comprehension designed to assess a learner's readiness to study at university level should include exercises specifically designed to test their capacity to establish situation model representations. The results from the current study suggest that one way to do this may be to embed incongruencies in text in order to test the reader's power to identify them. On the one hand, it may be argued that such a test would be somewhat inauthentic, given that academic texts do not typically contain clear contradictions of the type that were used in the current study. However, one of the key skills which differentiates academic reading from reading in other areas is the need to read critically, weighing and evaluating a writer's argument, often over lengthy sections of text. In this regard, it is not enough for readers at the university level simply to know what *was* written in a text, as demonstrated by the ability to answer comprehension questions. Introducing anomaly detection tests into advanced test of reading for academic purposes could be used to test for the ability to establish the types of situation model needed for readers to be able to perform these advanced academic skills.

8.3 Institutional implications

The findings from the research presented in this thesis indicate that in order for readers to learn from the texts they encounter, at least to the level of nuance and profundity required at university level, it is necessary for them to establish situation model elements in their mental representations of those texts.

However, the results also indicate that this ability may not become routinely available to native-Chinese speaking readers of English until they attain at least an upper-intermediate level of proficiency. While this study focused exclusively on native-Chinese speakers, and thus cannot make definite claims about students from other linguistic backgrounds, it seems plausible at least that these findings may apply to other non-native English readers. It was noted at the start of this chapter that the UK government has a stated aim of continuing to grow the international component of the university sector's student base. If UK universities are to continue this expansion, whilst maintaining the reputation for quality education they currently enjoy, there is arguably both an ethical and a *practical* imperative for adjustments to be made to the way international students are supported during their time at university. It is beyond the scope of the current thesis to propose detailed suggestions on how such amendments could be made. However, put briefly, these fall under two major categories: (*i*) improved in-sessional reading support; (*ii*) enhanced training for lecturers and other teaching staff on how best to support international students to learn from their texts. Firstly, it is important that training be provided to students on a regular and ongoing basis that helps them approach the task of reading with the specific intention of establishing situation model representations of the texts they read, rather than simply interacting with them at the linguistic level. Ancillary to this point, with regards to staff training, it was noted in Section 6.3.1. that although the establishment of situation model representations may only be routinely possible for students with more advanced reading skills, students at less advanced levels may be able to establish textbase representations, allowing them to recall information from a text, even though they may not fully comprehend it. With this in mind, it would be useful for lecturers and other teaching staff who lead seminar discussions to be provided with training on how best to guide lower proficiency international students to build upon these lower-order textbase representations in order to integrate the information they contain into a 'real world' context, with a view to helping them establish situation model representations. Within the EMI teaching profession, the maxim 'every lesson is a language lesson' has become commonplace, and is used to stress the importance of every part of the curriculum, whether specifically language focused or otherwise, taking into account the linguistic needs of the students,

so that they may continuously develop their language skills. If universities are to continue to expand the number of international students on their courses, it is vital that this ethos now be adopted in domestic universities. As this will, in many cases, represent a new teaching challenge for lecturers used to working with native- English speaking students, it is important that they now be given the training and support they need to develop this extra skill set.

8.4 Implications for the theory of second-language reading comprehension

The findings from this study also make some contribution towards the theory of second-language reading comprehension. It was explained in Section 2.4 how Walter (2007) challenged the long dominant metaphor of reading comprehension skills being transferred from the first language to the second. Rather, she suggested, it may be the case that amodal cognitive comprehension processing is *accessed* through the second language, once the reader attains a certain level of proficiency. By so doing, Walter brought some much needed clarity to the Linguistic Threshold Hypothesis, by positing a reason as to exactly what that threshold may allow for. The results from the current study have lent additional weight to the claim that second-language reading comprehension should be seen as a matter of access rather than transfer. Furthermore, while Walter (2007) conceptualised reading comprehension through Gernsbacher's (1997) Structure Building Framework, in the current research, reading comprehension was seen through the lens of Kintsch's (1998) Construction-Integration Model of reading comprehension. Kintsch's model argues that when processing a text, readers establish mental representations contain a number of elements, including sentence-level propositional representations, textbase representations and situation model representations (see Section 2.4 for a description of the Construction-Integration model). The findings from the current study suggested that the main impediment to the establishment of these various representations is difficulty processing the language of the text, thereby bringing yet more weight to the Linguistic Threshold Hypothesis. However, once the participants had reached a certain level of linguistic proficiency, albeit typically quite a high level, they showed evidence of having established complete representations,

containing both textbase and situation model elements. This suggests that they had been able to process the language of the text read, integrate its various propositions to establish higher order representations of their combined meaning, as well as to activate and integrate relevant prior knowledge held in long-term memory. That is to say, they had been able to access these various comprehension processes through the second language.

Furthermore, in addition to accessing the processes which allowed certain of the participants to establish complete mental representations of the text, there was also evidence to suggest that these participants had also been able to access the fundamental cognitive processes which underpin the ability for comprehension monitoring, at least in so far as anomaly detection is concerned. The eye movement data revealed that the upper-intermediate readers reacted to the presence of an incongruent word in a text within hundreds of milliseconds, a speed clearly too fast to be under conscious control. Rather, it may be the case that reaction times of this speed resulted from the readers experiencing dissonance when they attempted to integrate an incongruent idea into their developing mental representation of the text. That is, they were able to access, through their second language, the fundamental processes whereby human beings maintain cognitive equilibrium by preserving consistency in their mental representations of the world (Festinger, 1957). The fact that this ability was, for the most part confined to the upper-intermediate readers, who all had strong linguistic proficiency, suggests that this too is an instance of second-language readers being able to access fundamental cognitive functions through their second language only once they have passed a certain linguistic threshold.

8.5 Avenues for future research

As with all cutting edge research, the findings from the current study raise almost as many questions as they answer. Furthermore, constraints of time and resources inevitably meant that the data gathering process was subject to a number of limitations. With these two factors in mind, the following section will set out a number of possible directions for future research, based on the findings and limitations reported in

this thesis. These are presented as they relate to eye-tracking, the use of think-aloud interviews, and the combination of these two methodologies in mixed methods research designs.

8.5.1 Future avenues for eye-tracking research and the measurement of individual differences

Although the findings from this research suggested that differences do exist between second-language readers at different levels of proficiency in terms of their ability (i) to make use of grammatical determiners (e.g. *this*/*those*) as an aid to processing or (ii) to identify incongruencies in a text, the regression analyses did not detect a relationship between this ability and the participants' level of vocabulary knowledge, grammar knowledge or their working memory capacity. It was suggested in Section 6.4.1. that this may have resulted from deficiencies in the tests used to measure these various facets of individual difference. Specifically, the Schmidt et al. (2001) vocabulary test used only measured breadth of vocabulary knowledge, omitting depth. Given the role both of these types of vocabulary knowledge have been found to play in proficient reading (Nation, 2013), it would be worthwhile conducting research which measured both breadth and depth of knowledge, and synthesised the results into a more holistic measure, perhaps through the use of structural equation modelling, as a first step to generating more accurate predictor variables. In a similar vein, it was also suggested in Section 6.4.1 that the lack of a relationship between the participants' vocabulary scores and their behaviour in relation to the determiner or congruency factors may have resulted from the fact that the Schmidt et al. (2001) test measured knowledge of general, rather than subject specific, vocabulary. In order to overcome this potential limitation, a vocabulary test could be developed that is composed of items drawn from the subject on which the stimulus articles focus; for example, in the case of the current project, medical affairs. In this way, researchers would be able to draw a closer link between the participants' ability to establish situation model elements in their mental representations of the texts read, and their knowledge, or lack thereof, of the specific vocabulary contained within those texts. Such an approach, taken in conjunction with the combination of tests measuring both breadth and depth of vocabulary knowledge, would arguably further enhance the development of accurate predictor variables.

Secondly, it is important to note that the use of the participants' IELTS reading scores as a measurement of proficiency, while fulfilling the desired role in the current study, was not without its problems. This was not least because the way the IELTS score is calculated meant that is was not suitable for inclusion in rigorous statistical analysis (see Section 4.7. for a full discussion of this point). As a result, rather than being included as either an additional factor in the ANOVAs conducted, or as a covariate, the IELTS grades were used to divide the participants into groups which corresponded to the types of proficiency groupings they would encounter in an educational setting. However, whatever this approach provided in terms of allowing the results from this study to reflect 'real world' educational environments, it undoubtedly detracted in terms of the precision of the statistical tests conducted. Consequently, it would be useful for future research projects to repeat the experiments described in this thesis, but using a standardised reading test which provides a more precise metric of the participants' level of reading proficiency.

A second limitation of the eye-tracking results here reported regarding the effect of proficiency on the participants' performance was that once the sample was divided into three proficiency groups, the numbers in each group fell below the conventional threshold of at least thirty participants for the Central Limit Theorem to apply (Field, 2013), albeit only by one or two participants in each group. Consequently, in addition to using a more rigorous measure of reading proficiency, future studies should also be conducted with larger samples. This would allow for a higher degree of statistical confidence to be reached regarding the relationship between proficiency and (i) the ability to make use of grammatical determiners as an aid to processing, and (ii) the ability to detect incongruencies in a text.

8.5.2 Avenues for further research using think-aloud interviews

It was noted in Section 5.4. that although only the upper-intermediate readers were able to routinely identify the incongruent words in the text, the majority of participants at all levels were able to integrate the coreferential noun phrases in the majority of instances. The think-aloud interviews conducted in the current studies were designed to gather data on a number of issues, including the type of comprehension difficulties

the participants faced as they read, the regulatory strategies they attempted to use in order to overcome them, as well as how they believed the sentences in the text to be related to each other. Although the analysis was able to provide insight into the types of mental representations the participants were able to establish as they read, these were gleaned by analysing what the participants said as they explained how they thought the sentences in the text read related to each other. It was found that the clearest indication of whether or not the participants were able to establish situation model representations of the meaning of the text, or only propositional or textbase ones, were provided when they spoke in their native Chinese. This was because the process of translation gave a clearer indication of not only what they believed the text to mean, but also of what they were able to infer when they did not understand a word or phrase it contained. For example, it was found that although Participant 032 stated during reading that she did not know the word *surgeon*, after several attempts to regulate her comprehension of the relevant section of the text, she produced a rendition of its meaning in Chinese which included the word *doctor* where the original word *surgeon* had been used. Although this clearly represents a simplification of the original meaning, it was concluded from this that since she did not know the word *surgeon*, the idea of *doctor* must have been accessed through the establishment of a situation model representation of the text. In order to build on this finding, it may prove fruitful for subsequent research into how second-language readers establish situation model representations of a text to conduct think aloud interviews in which the participants are instructed to explain exactly what they believe a text to mean in their first language.

8.5.3 Avenues for future research using a mixed-methods research design

As noted above, the results from the analysis reported in this thesis found that although only the upper-intermediate participants were able to routinely identify the incongruent words in a text, in general the participants were all able to integrate the coreferential noun phrases with their antecedents. It is important to acknowledge that this result was contrary to initial expectations. At the outset of this project, it was assumed the ability to identify an incongruent word in a coreferential noun phrase could be taken as a proxy

for the ability to integrate that noun phrase with its antecedent. That is to say, these two abilities were taken to be almost co-dependent. While the results from the eye-tracking study were able to demonstrate that the upper-intermediate readers were able to detect the incongruent words in the texts more readily than the lower-proficiency readers, as indicated by increased gazed durations in the second pass reading times, they did not provide any indication of why this was. It was only through the subsequent use of think-aloud interviews that the lack of relationship between the ability to integrate coreferential noun phrase and the ability to identify the incongruencies in a text came to light. Similarly, it was only through the analysis of the think-aloud transcripts that an alternative explanation for the ability to identify such incongruencies, namely through the establishment of situation model representations, was derived. Had the research conducted for this study been restricted to the use of eye-tracking alone, these findings would not have been possible, and an important truth potentially lost. In recent years, there have been growing calls for the use of mixed-method research designs in educational research in general (see Section 3.2.1. for a full discussion of this issue), and for the specific combination of eye-tracking and introspective interview techniques in reading research in particular (Godfroid & Spino, 2015). While no recommendations for specific mixed-methods research grow out of this project, the results reported add substantial weight to the case that these types of designs can, if used properly, provide a powerful analytical tool to researchers working in the field of reading comprehension, allowing for considerably greater insights to be reached than is possible through the use of single method designs.

References

Abbott, M. L. (2006). ESL reading strategies: Differences in Arabic and Mandarin speaker test performance. *Language Learning*, *56*(4), 633–670. https://doi.org/10.1111/j.1467-9922.2006.00391.x

Alderson, C. J. (1984). Reading in a foreign language: a reading problem or a lanaguage problem. In C. J. Alderson & A. H. Urquhart (Eds.), *Reading in a foreign language* (pp. 1–27). Longman.

Alptekin, C. E. M., & Erçetin, G. (2010). Research in Reading The role of L1 and L2 working memory in literal and inferential comprehension in L2 reading. *Journal of Research in Reading*, *33*(2), 206–219. https://doi.org/10.1111/j.1467-9817.2009.01412.x

Alptekin, C. E. M., & Erçetin, G. (2011). *The Effects of Working Memory Capacity and Content Familiarity on Literal and Effects of Working Memory Capacity and Content Familiarity on Literal and Inferential Comprehension in L2 Reading*. *45*(2), 235–266.

Anderson, N. J. (1991). Individual Differences in Strategy Use in Second Language Reading and Testing. *The Modern Language Journal*, *75*(4), 460–472.

Ariel, M. (2001a). Accessibility Marking. *Text Representation: Linguistic and Psycholinguistic Aspects*, *6950*(September), 29–87. https://doi.org/10.1207/s15326950dp3702

Ariel, M. (2001b). Accessibility theory: An overview. In T. J. M. Sanders, J. Schilperoord, & W. Spooren (Eds.), *Text Representation: Linguistic and Psycholinguistic Aspects* (pp. 29–87). John Benjamins Publishing Company.

Armstrong, A., Bulkes, N., & Tanner, D. (2018). Quantificational cues modulate the processing of English subject-verb agreement by native Chinese speakers. *Studies in Second Language Acquisition*, *40*, 731–754. https://doi.org/10.1017/S0272263118000013

Aviad-Levitzky, T., Laufer, B., & Goldstein, Z. (2019). The New Computer Adaptive Test of Size and Strength (CATSS): Development and Validation. *Language Assessment Quarterly*, *16*(3), 345–368. https://doi.org/10.1080/15434303.2019.1649409

Bax, S. (2013). The cognitive processing of candidates during reading tests: Evidence from eye-tracking. *Language Testing*, *30*(4), 441–465. https://doi.org/10.1177/0265532212473244

Bernhardt, E. (2010). *Understanding Advanced Second-Language Reading*. Routledge.

Bernhardt, E. B., & Kamil, M. L. (1995). Interpreting Relationships between LI andL2 Reading : Consolidating the Linguistic Threshold and the Linguistic Interdependence Hypotheses. *Applied Linguistics*, *16*(1), 15–34.

Block, E. (1986). The Comprehension Strategies of Second Language Readers. *Quarterly, Tesol*, *20*(3), 463–494.

Bosch, L. J. Van Den, Segers, E., & Verhoeven, L. (2018). Online processing of causal relations in beginning first and second language readers. *Learning and Individual Differences*, *61*, 59–67. https://doi.org/10.1016/j.lindif.2017.11.007

Bott, O., & Gattnar, A. (2015). The cross-linguistic processing of aspect – an eyetracking study on the time course of aspectual interpretation in Russian and German. *Language, Cognition and Neuroscience*, *30*(7), 877–898. https://doi.org/10.1080/23273798.2015.1029499

Bowles, M. A. (2010). *The think-aloud contoversy in second language research (Second language acquisition*

research series) (1st ed.). Routledge.

Braze, D., Shankweiler, D., Ni, W., & Palumbo, L. C. (2002). Readers' eye movements distinguish anomalies of form and content. *Journal of Psycholinguistic Research*, *31*(1), 25–44. https://doi.org/10.1023/A:1014324220455

Brennan, C., & Booth, J. R. (2015). Large grain instruction and phonological awareness skill influence rime sensitivity, processing speed, and early decoding skill in adult L2 learners. *Reading and Writing*, *28*, 917–938. https://doi.org/10.1007/s11145-015-9555-2

Brevik, L. M. (2019). Explicit reading strategy instruction or daily use of strategies? Studying the teaching of reading comprehension through naturalistic classroom observation in English L2. *Reading and Writing*, *32*(9), 2281–2310. https://doi.org/10.1007/s11145-019-09951-w

Broek, P. Van Den, & Helder, A. (2017). Cognitive Processes in Discourse Comprehension: Passive Processes, Reader-Initiated Processes, and Evolving Mental Representations. *Discourse Processes*, *54*(5–6), 360–372. https://doi.org/10.1080/0163853X.2017.1306677

Brunfaut, T., Kormos, J., Michel, M., & Ratajczak, M. (2021). Testing young foreign language learners ' reading comprehension : Exploring the effects of working memory , grade level , and reading task. *Language Testing*, *38*(3), 356–377. https://doi.org/10.1177/0265532221991480

Brunfaut, T., & McCray, G. (2015). Looking into test-takers' cognitive processes while completing reading tasks: A mixed-method eye-tracking and stimulated recall study. In *AGAGs Research Reports Online*.

Bultena, S., Bekkering, H., Danielmeier, C., & Lemhöfer, K. (2020). The Role of Conflicting Representations and Uncertainty in Internal Error Detection During L2 Learning. *Language Learning*, *70*(2), 75–103. https://doi.org/10.1111/lang.12401

Cameron, C. A., & Lee, K. (1999). Emergent use of English grammatical morphemes by Chinese-speaking children. *IRAL - International Review of Applied Linguistics in Language Teaching*, *37*(1), 43–58.

Carter, B. T., & Luke, S. G. (2018). Individuals ' Eye Movements in Reading are Highly Consistent Across Time and Trial. *Journal of Experimental Psychology*, *44*(3), 482–492.

Casanave, C. P. (1988). Comprehension Monitoring in ESL Reading : A Neglected Essential. *TESOL Quarterly*, *22*(2), 283–302.

Cheng, J., & Matthews, J. (2018). *The relationship between three measures of L2 vocabulary knowledge and L2 listening and reading*. https://doi.org/10.1177/0265532216676851

Cho, J. R., & Chiu, M. M. (2015). Rapid naming in relation to reading and writing in Korean (Hangul), Chinese (Hanja) and English among Korean children: A 1-year longitudinal study. *Journal of Research in Reading*, *38*(4), 387–404. https://doi.org/10.1111/1467-9817.12020

Clarke, M. A. (1980). The Short Circuit Hypothesis of ESL Reading- or When Language Competence Interferes with Reading Performance. *The Modern Language Journal*, *64*(2), 203–209.

Coady, J. (1979). A psycholinguistic model of the ESL reader. In R. Mackay, B. Barkman, & R. . Jordan (Eds.), *Reading in a second langauge* (pp. 5–12).

Conklin, K., & Pellicer-Sánchez, A. (2016). Using eye-tracking in applied linguistics and second language research. *Second Language Research*, *32*(3), 453–467. https://doi.org/10.1177/0267658316637401

Cox, B., & Forshaw, J. (2010). *Why Does E=mc2 (And why should we care?)* (ed.). Da Capo; UK.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*(2), 213–238.

Coxhead, A. (2008). Phraseology and English for academic purposes: Challenges and opportunities. In F. Meunier & S. Granger (Eds.), *Phraseology in foreign language learning and teaching* (pp. 149–162). John Benjamins Publishing.

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and Conducting Mixed Methods Reserach* (3rd ed.). Sage Publications.

Creswell, J. W., & Tashakkori, A. (2007). Differing perspectives on mixed research methods. *Journal of Mixed Research Methods*, *1*(4), 303–308.

Croninger, R. G., & Valli, L. (2009). Mixing it up about methods. *Educational Researcher*, *38*(7), 541–545. https://doi.org/10.3102/0013189X09348787

Cummins, J. (1981). Four misconceptions about language proficiency in bilingual education. *Nabe Journal*, *5*(3), 31–45.

Cummins, J. (1998). Immersion education for the millennium: What have we learned from 30 years of research on second language immersion? In M. R. Childs & R. M. Bostwick (Eds.), *Learning through two languages: Research and practice. Second Katoh Gakuen International Symposium on Immersion and Bilingual Education* (pp. 34–47). Katoh Gakuen.

Degand, L., & Sanders, T. E. D. (2002). The impact of relational markers on expository text comprehension in L1 and L2. *Reading and Writing*, *15*(7), 739–757.

DeKeyser, R. M. (2000). the Robustness of Critical Period Effects in Second Language Acquisition. *Studies in Second Language Acquisition*, *22*(4), 499–533. https://doi.org/10.1017/s0272263100004022

Dewey, J. (1929). *Experience and nature*. George Allen & unwin.

Dörnyei, Z. (2007). *Reserach methods in applied linguistics*. Oxford University Press.

Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical Ambiguity and Fixation Times in Reading. *Journal of Memory and Language*, *27*, 429–446.

Ericsson, K.A. Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*(2), 211–245.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, *87*(3), 215–251.

Erlich, K., & Rayner, K. (1983). Pronoun Assignment and Semantic Integration during Reading : Eye Movements and Immediacy of Processing. *Journal of Verbal Learning and Verbal Behavior*, *22*, 75–87.

Fesel, S. S., Segers, E., Clariana, R. B., & Verhoeven, L. (2015). Quality of children's knowledge representations in digital text comprehension: Evidence from pathfinder networks. *Computers in Human Behavior*, *48*, 135–146. https://doi.org/10.1016/j.chb.2015.01.014

Festinger, L. (1957). *A theory of cognitive dissonance (Vol. 2)*. Stanford University Press.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). Sage Publications.

Fox, E. (2009). The Role of Reader Characteristics in Processing and Learning From Informational Text. *Review of Educational Research*, *79*(1), 197–261. https://doi.org/10.3102/0034654308324654

Fox, M. C., Ericsson, K. A., & Best, R. (2011). Do Procedures for Verbal Reporting of Thinking Have to Be Reactive ? A Meta-Analysis and Recommendations for Best Reporting Methods. *Psychological Bulletin*, *137*(2), 316–344. https://doi.org/10.1037/a0021663

Francis, G. (1994). Labelling discourse: an aspect of nominal-group lexical cohesion. In M. Coulthard (Ed.), *Advances in Written Text Analysis* (pp. 83–101). Routledge.

Frenck-Mestre, C. (2005). *Eye-movement recording as a tool for studying syntactic processing in a second language : a review of methodologies and experimental findings*. *2*, 175–198.

Fukkink, R. G., Hulstijn, J., & Simis, A. (2005). Does training in second-language word recognition skills affect reading comprehension? An experimental study. *Modern Language Journal*, *89*(1), 54–75. https://doi.org/10.1111/j.0026-7902.2005.00265.x

García-Rodicio, H., & Sánchez, E. (2014). Does the detection of misunderstanding lead to its revision? *Metacognition and Learning*, *9*, 265–286. https://doi.org/10.1007/s11409-014-9116-0

Gernsbacher, M. A. N. N. (1989). Mechanisms that improve referential access. *Cognition*, *32*, 99–156.

Gernsbacher, M.A, & Faust, M. (1995). *Interference and inhibition in cognition*. Academic Press.

Gernsbacher, Morton Ann. (1997). Two decades of structure building. *Discourse Processes*, *23*(3), 265–304. https://doi.org/10.1080/01638539709544994

Gernsbacher, Morton Ann, & Faust, M. E. (1991). The mechanism of suppression. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(2), 245–262. https://doi.org/10.1037/0278-7393.17.2.245

Gernsbacher, Morton Ann, Varner, K. R., & Faust, M. E. (1990). Investigating differences in general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(3), 430–445.

Geva, E. (1992). The Role of Conjunctions in L2 Text Comprehension. *TESOL Quarterly*, *26*(4), 731. https://doi.org/10.2307/3586871

Ghavamnia, M., Ketabi, S., & Tavakoli, M. (2013). L2 reading strategies used by Iranian EFL learners : A think-aloud study. *Reading Psychology*, *34*, 355–378. https://doi.org/10.1080/02702711.2011.640097

Godfroid, A., & Spino, L. A. (2015). Reconceptualizing reactivity of think-alouds and eye tracking : Absence of evidence is not evidence of absence. *Language Learning*, *65*(4), 896–928. https://doi.org/10.1111/lang.12136

Goodman, K. (1973). Psycholinguistic universals of the reading process. In F. Smith (Ed.), *Psycholinguistics and Reading* (pp. 21–29). Holt, Reinhart and Winston.

Gough, P. ., & Turner, W. . (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, *7*(1), 6–10.

Grabe, W. (2009). *Reading in a second language Moving from theory to practice*. Cambridge University Press.

Grabe, W., & Stoller, F. L. (2015). *Teaching and Researching Reading* (2nd ed.). Routledge.

Graesser, A. C., & Mello, S. D. (2012). Moment-To-Moment Emotions During Reading. *The Reading Teacher*, *66*(3), 238–242. https://doi.org/10.1002/TRTR.01121

Gray, B. (2010). On the use of demonstrative pronouns and determiners as cohesive devices : A focus on sentence-initial this / these in academic prose. *Journal of English for Academic Purposes*, *9*(3), 167–183. https://doi.org/10.1016/j.jeap.2009.11.003

Greene, J. C., Carcelli, V. J., & Graham, W. F. (1989). Toward a Conceptual Framework for Mixed-Method

Evaluation Designs Jennifer C . Greene ; Valerie J . Caracelli ; Wendy F . Graham. *Educationl Evaluation and Policy Analysis*, *11*(3), 255–274.

Gu, Y. (2014). To code or not to code : Dilemmas in analysing think-aloud protocols in learning strategies research. *System*, *43*, 74–81. https://doi.org/10.1016/j.system.2013.12.011

Güss, C. D. (2018). What is going through your mind ? Thinking aloud as a method in cross-cultural psychology. *Frontiers in Psychology*, *9*(August), 1–11. https://doi.org/10.3389/fpsyg.2018.01292

Halliday, M. A. K. (1993a). Some grammatical problems in scientific English. In M. A. K. Halliday & J. R. Martin (Eds.), *Writing science: Literacy and discursive power*. The Falmer Press.

Halliday, M. A. K. (1993b). The analysis of scientific texts in English and Chinese. In M. A. K. Halliday & J. R. Martin (Eds.), *Writing science: Literacy and discursive power* (pp. 124–132). The Falmer Press.

Halliday, M. A. K., & Matthiessen, C. (2014). *Halliday's Introduction to Functional Grammar*. Routledge.

Halliday, M., & Hassan, R. (1976). *Cohesion in English*. Longman.

Hesse-Biber, S. (2018). Toward an Understanding of a Qualitatively Driven Mixed Methods Data Collection and Analysis: Moving Toward a Theoretically Centered Mixed Methods Praxis. *The SAGE Handbook of Qualitative Data Collection*, *2017*, 545–563. https://doi.org/10.4135/9781526416070.n35

Hessel, A. K., Nation, K., & Murphy, V. A. (2021). Comprehension Monitoring during Reading: An Eye-tracking Study with Children Learning English as an Additional Language. *Scientific Studies of Reading*, *25*(2), 159–178. https://doi.org/10.1080/10888438.2020.1740227

Hoey, M. (2001). *Textual interaction: An introduction to written discourse analysis*. Psychology Press.

Holstein, J. A., & Gubrium, J. F. (1995). *The Active Interview*. Sage Publications Inc.

Hu, J., & Gao, X. (Andy). (2017). Using think-aloud protocol in self-regulated reading research. *Educational Research Review*, *22*, 181–193. https://doi.org/10.1016/j.edurev.2017.09.004

Hubble, S., & Bolton, P. (2020). International and EU students in higher education in the UK. *Commons Library*, *42*(4), 1–22.

Irwin, D. E. (1998). Lexical Processing during Saccadic Eye Movements. *Cognitive Psychology*, *36*, 1–27.

Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language Learning*, *64*(1), 160–212. https://doi.org/10.1111/lang.12034

Joh, J., & Plakans, L. (2017). Working memory in L2 reading comprehension: The influence of prior knowledge. *System*, *70*, 107–120. https://doi.org/10.1016/j.system.2017.07.007

Johns, C. L., Matsuki, K., & Van Dyke, J. A. (2015). Poor readers' retrieval mechanism: Efficient access is not dependent on reading skill. *Frontiers in Psychology*, *6*(OCT), 1–20. https://doi.org/10.3389/fpsyg.2015.01552

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational Researcher*, *33*(7), 14–26. https://doi.org/10.3102/0013189X033007014

Jordan, R. R. (2007). *English for academic purposes: A guide and resource book for teachers*. Ernst Klett Sprachen.

Joseph, H., Wonnacott, E., & Nation, K. (2021). Online inference making and comprehension monitoring in children during reading: Evidence from eye movements. *Quarterly Journal of Experimental Psychology*,

74(7), 1202–1224. https://doi.org/10.1177/1747021821999007

Juffs, A., & Harrington, M. (2011). State-of-the-Art Article Aspects of working memory in L2 learning. *Language Teaching*, *44*(2), 137–166. https://doi.org/10.1017/S0261444810000509

Just, M. A., & Carpenter, P. A. (1980). A Theory of Reading: From Eye Fixations to Comprehension. *Psychological Review*, *87*(4), 329–354.

Kardash, C. M., & Howell, K. L. (2000). Effects of epistemological beliefs and topic-specific beliefs on undergraduates' cognitive and strategic processing of dual-positional text. *Journal of Educational Psychology*, *92*(3), 524–535. https://doi.org/10.10371/13022-0663.92.3.524

Kim, K., & Clariana, R. B. (2015). Knowledge Structure Measures of Reader's Situation Models Across Languages: Translation Engenders Richer Structure. *Technology, Knowledge and Learning*, *20*, 249–268. https://doi.org/10.1007/s10758-015-9246-8

Kinnunen, R., & Vauras, M. (1995). Comprehension monitoring and the level of comprehension in high- and low-achieving primary school children's reading. *Learning and Instruction*, *5*, 143–165. https://doi.org/10.1016/0959-4752(95)00009-R

Kintsch, W. (1998). *Comprehension A paradigm for cognition*. Cambridge University Press. https://doi.org/10.1016/s0378-2166(99)00090-9

Kintsch, W., & Rawson, K. A. (2005). Comprehension. In C. E. Snowling, M.J. & Hulme (Ed.), *The science of reading: A handbook* (pp. 312–335). Blackwell Publishing.

Knagg, J. (2013). EMI within a Global Context – Towards a British Council Perspective. In *British Council Regional Policy Dialogues 2013-14*. British Council. https://www.britishcouncil.org/sites/default/files/bc_regionalpolicydialogues_baja.pdf

Koda, K. (1993). Transferred L 1 Strategies and L2 Syntactic Structure in L2 Sentence Comprehension. *The Modern Language Journal*, *77*(4), 490–500.

Koda, K. (2005). *Insights into second language reading: A cross-linguistic approach*. Cambridge University Press.

Kormos, J., & Sáfár, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition*, *11*(2), 261–271. https://doi.org/10.1017/S1366728908003416

Kucan, L., & Beck, I. L. (1997). Thinking aloud and reading comprehension research: Inquiry, instruction, and social interaction. *Review of Educational Research*, *67*(3), 271–299.

Leeser, M. J. (2007). Learner-Based Factors in L2 Reading Comprehension and Processing Grammatical Form: Topic Familiarity and Working Memory. *Langauge Learning*, *57*(2), 229–270.

Leow, R. P., Grey, S., Marijuan, S., & Moorman, C. (2014). Concurrent data elicitation procedures , processes , and the early stages of L2 learning : A critical overview. *Second Language Research*, *30*(2), 111–127. https://doi.org/10.1177/0267658313511979

Leow, R. P., & Morgan-Short, K. (2004). To think aloud or not to think aloud: The Issue of Reactivity in SLA Research Methodology. *Studies in Second Language Acquisition*, *26*(1), 35–57. https://doi.org/10.1017/s0272263104261022

Li, P., & Clariana, R. B. (2019). Reading comprehension in L1 and L2: An integrative approach. *Journal of*

Neurolinguistics, *50*, 94–105. https://doi.org/10.1016/j.jneuroling.2018.03.005

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production : A meta-analysis. *Psychonomic Bulletin and Review*, *21*(4), 861–883. https://doi.org/10.3758/s13423-013-0565-2

Linderholm, T., & van den Broek, P. (2002). The Effects of Reading Purpose and Working Memory Capacity on the Processing of Expository Text. *Journal of Educational Psychology*, *94*(4), 778–784. https://doi.org/10.1037//0022-0663.94.4.778

Mak, W. M., & Sanders, T. J. M. (2013). The role of causality in discourse processing: Effects of expectation and coherence relations. *Language and Cognitive Processes*, *28*(9), 1414–1437. https://doi.org/10.1080/01690965.2012.708423

Mann, S. (2011). A critical review of qualitative interviews in applied linguistics. *Applied Linguistics*, *32*(1), 6–24. https://doi.org/10.1093/applin/amq043

Mark, M. M. (2015). Mixed and Multimethods in Predominantly Quantitative Studies, Especially Experiments and Quasi-Experiments. In *The oxford handbook of multimethod and mixed methods research inquiry* (pp. 21–41).

McEntire, J., & Williams, J. (2013). *Making Connections Level 2 Student's Book: Skills and Strategies for Academic Reading* (2nd ed.). Cambridge University Press.

Melby-Lervåg, M., & Lervåg, A. (2011). Cross-linguistic transfer of oral language, decoding, phonological awareness and reading comprehension: A meta-analysis of the correlational evidence. *Journal of Research in Reading*, *34*(1), 114–135. https://doi.org/10.1111/j.1467-9817.2010.01477.x

Miller, B. W. (2015). Using Reading Times and Eye-Movements to Measure Cognitive Engagement Using Reading Times and Eye-Movements to Measure Cognitive Engagement. *Educational Psychologist*, *50*(1), 31–42. https://doi.org/10.1080/00461520.2015.1004068

Moore, P. J., & Scevak, J. J. (1997). Learning from texts and visual aids: A developmental perspective. *Journal of Research in Reading*, *20*(3), 205–223.

Morrison, L. (2004). Comprehension monitoring in first and second language reading. *Canadian Modern Language Review*, *61*(1), 77–106. https://doi.org/10.3138/cmlr.61.1.77

Mutch, C. (2009). Mixed method research : methodological eclecticism or muddled thinking ? *Journal of Educational Leadership, Policy and Practice*, *24*(2), 18–30.

Nation, I. S. (2008). *Teaching ESL/EFL reading and writing*. Routledge.

Nation, Ian SP. (2013). *Learning vocabulary in another language*. Cambridge University Press.

Neuman, L. W. (2014). *Social Research Methods: Qualitative and Quantitative Approaches* (7th ed.). Pearson Education Limited.

Oakhill, J., Hartt, J., & Samols, D. (2005). Levels of comprehension monitoring and working memory in good and poor comprehenders. *Reading and Writing*, *18*(7–9), 657–686. https://doi.org/10.1007/s11145-005-3355-z

Ozono, S., & Ito, H. (2003). Logical connectives as catalysts for interactive reading. *System*, *31*(2), 283–297.

Peng, P., Wang, C., Li, S., Dardick, W., Barnes, M., Wang, W., Swanson, H. L., & Tao, S. (2018). A Meta-Analysis on the Relation Between Reading and Working Memory. *Psychological Bulletin*, *144*(1), 48–76.

Perfetti, C., & Stafura, J. (2014). Word Knowledge in a Theory of Reading Comprehension Reading Comprehension. *Scientific Studies of Reading*, *18*(1), 22–37. https://doi.org/10.1080/10888438.2013.827687

Pretorius, E. J. (2005). English as a second language learner differences in anaphoric resolution: Reading to learn in the academic context. *Applied Psycholinguistics*, *26*(4), 521–539. https://doi.org/10.1017/S0142716405050289

Proctor, C. P., Silverman, R. D., Harring, J. R., & Montecillo, C. (2012). The role of vocabulary depth in predicting reading comprehension among English monolingual and Spanish-English bilingual children in elementary school. *Reading and Writing*, *25*(7), 1635–1664. https://doi.org/10.1007/s11145-011-9336-5

Rai, M. K., Loschky, L. C., Harris, R. J., Peck, N. R., & Cook, L. G. (2011). Effects of Stress and Working Memory Capacity on Foreign Language Readers' Inferential Processing During Comprehension. *Language Learning*, *61*(1), 187–218. https://doi.org/10.1111/j.1467-9922.2010.00592.x

Rapp, D. N., van den Broek, P., McMaster, K. L., Kendeou, P., & Espin, C. A. (2007). Higher-Order Comprehension Processes in Struggling Readers: A Perspective for Research and Intervention. *Scientific Studies of Reading*, *11*(4), 289–312. https://doi.org/10.1080/10888430701530417

Raudszus, H., Segers, E., & Verhoeven, L. (2019). Situation model building ability uniquely predicts first and second language reading comprehension. *Journal of Neurolinguistics*, *50*(October 2018), 106–119. https://doi.org/10.1016/j.jneuroling.2018.11.003

Rayner, K. (1975). The Perceptual Span and Peripheral Cues in Reading. *Cognitive Psychology*, *7*, 65–81.

Rayner, K. (1998). Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychological Bulletin*, *124*(3), 372–422. https://doi.org/10.1037/0033-2909.124.3.372

Rayner, K. (2009). Eye movements and attention in reading , scene perception , and visual search. *Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506. https://doi.org/10.1080/17470210902816461

Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye Movements as Reflections of Comprehension Processes in Reading. *Scientific Studies of Reading*, *10*(3), 241–255.

Rayner, K., Juhasz, B. J., Warren, T., & Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning Memory and Cognition*, *30*(6), 1290–1301. https://doi.org/10.1037/0278-7393.30.6.1290

Reichle, E. D., Reineberg, A. E., & Schooler, J. W. (2010). Eye Movements During Mindless Reading. *Psychological Science*, *21*(9), 1300–1310. https://doi.org/10.1177/0956797610378686

Rubman, C. N., & Waters, H. S. (2000). A , B seeing: The role of constructive processes in children's comprehension monitoring. *Journal of Educational Psychology*, *92*(3), 503–514.

Russel, B. (1946). *History of Western Philosophy* (Kindle). Routledge.

Sanders, T. J., & Gernsbacher, M. A. (2004). Accessibility in Text and Discourse Processing Accessibility in Text and Discourse Processing. *Discourse Processes*, *37*(2), 79–89. https://doi.org/10.1207/s15326950dp3702

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, *18*(1), 55–88. https://doi.org/10.1177/026553220101800103

Schoepp, K. (2018). Predictive validity of the IELTS in an English as a medium of instruction environment. *Higher Education Quarterly*, *72*(4), 271–285. https://doi.org/10.1111/hequ.12163

Seliger, H. W. (1983). The language learner as linguist: Of metaphors and realities. *Applied Linguistics*, *4*(3), 179–191.

Sheridan, H., & Reingold, E. M. (2012). The time course of contextual influences during lexical ambiguity resolution: Evidence from distributional analyses of fixation durations. *Memory & Cognition*, *40*(7), 1122–1131. https://doi.org/10.3758/s13421-012-0216-2

Smith, R., Snow, P., Serry, T., & Hammond, L. (2021). The Role of Background Knowledge in Reading Comprehension: A Critical Review. *Reading Psychology*, *42*(3), 214–240. https://doi.org/10.1080/02702711.2021.1888348

Sparks, R. L. (2015). Language Deficits in Poor L2 Comprehenders: The Simple View. *Foreign Language Annals*, *48*(4), 635–658. https://doi.org/10.1111/flan.12163

SR Research. (2018). *EyeLink ® Data Viewer User ' s Manual. Document Version 1.8.221* (3.2.1).

Stanovich, K. E. (1980a). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*, *16*(1), 32–71.

Stanovich, K. E. (1980b). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Reserach Quarterly*, *16*(1), 32–71.

Staub, A., & Rayner, K. (2007). *Eye movements and on-line comprehension processes* (pp. 327–342).

Stepanov, A., Andreetta, S., Stateva, P., Zawiszewski, A., & Laka, I. (2020). Anomaly detection in the processing of complex syntax by early L2 learners. *Second Language Research*, *36*(3), 371–397. https://doi.org/10.1177/0267658319827065

Stratman, J. F., & Hamp-Lyons, L. (1994). Reactivity in concurrent think-aloud protocols: Issues for research. *Speaking about Writing: Reflections on Research Methodology*, *8*, 89–111.

Tadros, A. (1994). Predictive categories in expository text. In M. Coulthard (Ed.), *Advances in Written Text Analysis* (pp. 69–82). Rout.

Taillefer, G. (1996). L2 Reading Ability : Further Insight into the Short-circuit Hypothesis. *The Modern Language Journal*, *80*(4), 461–477.

Tashakkori, A., & Teddlie, C. (2003). Preface. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of Mixed Methods in Social & Behavioral Research* (pp. ix–xv). Sage Publications Inc.

Thompson, G. (2004). *Introducing Functional Grammar*. Hodder Education.

Trenkic, D., & Warmington, M. (2018). Language and literacy skills of home and international university students: How different are they, and does it matter? *Bilingualism: Language and Cognition*, *22*(2), 349–365. https://doi.org/10.1017/S136672891700075X

Tsai, Y. R., Ernst, C., & Talley, P. C. (2010). L1 and L2 strategy use in reading comprehension of Chinese EFL readers. *Reading Psychology*, *31*(2), 1–29. https://doi.org/10.1080/02702710802412081

Ushiro, Y., Hosoda, M., Mori, Y., Kimura, Y., Tanaka, N., Nakagawa, H., & Amagai, Y. (2017). Supporting the Maintenance of Global Coherence With Situational Instruction: Evidence From Eye Movements During EFL Reading. *JACET Journal*, *61*, 61.

Ushiro, Y., Ogiso, T., Nahatame, S., Hosoda, M., Kamimura, K., Sasaki, Y., Komuro, R., Okada, R., & Aoki, S. (2020). EFL Readers' Understanding of Protagonist, Temporal, and Spatial Links in Narrative: Evidence From Eye Tracking. *ARELE: Annual Review of English Language Education in Japan*, *31*, 97–112.

van der Schoot, M., Reijntjes, A., & van Lieshout, E. C. D. M. (2012). How do children deal with inconsistencies in text? An eye fixation and self-paced reading study in good and poor reading comprehenders. *Reading and Writing*, *25*(7), 1665–1690. https://doi.org/10.1007/s11145-011-9337-4

Van Silfhout, G., Evers-Vermeul, J., & Sanders, T. (2015). Connectives as processing signals: How students benefit in processing narrative and expository texts. *Discourse Processes*, *52*(1), 47–76. https://doi.org/https://doi.org/10.1080/0163853X.2014.905237

Varma, S., & Janssen, A. (2019). The structure of situation models as revealed by anaphor resolution. *Language Sciences*, *72*, 104–115. https://doi.org/10.1016/j.langsci.2018.09.002

von der Malsburg, T., & Angele, B. (2017). False positives and other statistical errors in standard analyses of eye movements in reading. *Journal of Memory and Language*, *94*, 119–133. https://doi.org/10.1016/j.jml.2016.10.003

Wade, S. E. (1990). Using think alouds to assess comprehension. *The Reading Teacher*, *43*(7), 442–451.

Walter, C. (2004). Transfer of reading comprehension skills to L2 is linked to mental representations of text and to L2 working memory. *Applied Linguistics*, *25*(3), 315–339. https://doi.org/10.1093/applin/25.3.315

Walter, C. (2007). First- to second-language reading comprehension: Not transfer, but access. *International Journal of Applied Linguistics*, *17*(1), 14–37. https://doi.org/10.1111/j.1473-4192.2007.00131.x

Weir, C., Hawkey, R., Green, A., & Devi, S. (2012). The cognitive processes underlying the academic reading construct as measured by IELTS. In L. Taylor & C. J. Weir (Eds.), *IELTS collected papers, 2* (pp. 212–269). Cambridge University Press.

White, C., Schramm, K., & Chamot, A. . (2007). Research methods in strategy research: Re-examining the toolbox. *Language Learner Strategies*, *30*, 93–116.

Wineburg, S. . (1991). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology*, *83*(1), 73–87.

Winter, E. (1994). Clause relations as information structure: two basic text structures in English. In M. Coulthard (Ed.), *Advances in Written Text Analysis* (pp. 46–68). Routledge.

Woore, R. (2009). Beginners' progress in decoding L2 French: Some longitudinal evidence from English modern foreign languages classrooms. *Language Learning Journal*, *37*(1), 3–18. https://doi.org/10.1080/09571730902717398

Wu, J. (2014). Shifts of cohesive devices in English-Chinese translation. *Theory and Practice in Language Studies*, *4*(8), 1659–1665.

Yamashita, J., & Shiotsu, T. (2017). *Comprehension and Knowledge Components That Predict L2 Reading : A Latent-Trait Approach*. *February 2015*, 43–67. https://doi.org/10.1093/applin/amu079

Yang, J.-S. (2014). A Contrastive Study of Cohesion in English and Chinese. *International Journal of English Linguistics*, *4*(6), 1923–8703.

Yapp, D., de Graaff, R., & van den Bergh, H. (2021). Effects of reading strategy instruction in English as a

second language on students' academic reading comprehension. *Language Teaching Research*, 1–24. https://doi.org/10.1177/1362168820985236

Zhang, D. (2012). Vocabulary and Grammar Knowledge in Second Language Reading Comprehension: A Structural Equation Modeling Study. *Modern Language Journal*, *96*(4), 558–575. https://doi.org/10.1111/j.1540-4781.2012.01398.x

Zhou, H., Rossi, S., & Chen, B. (2017). Effects of working memory capacity and tasks in processing L2 complex sentence: Evidence from Chinese-english bilinguals. *Frontiers in Psychology*, *8*, 1–10. https://doi.org/10.3389/fpsyg.2017.00595

Zhou, Y., & McBride, C. (2018). The same or different: An investigation of cognitive and metalinguistic correlates of Chinese word reading for native and non-native Chinese speaking children. *Bilingualism: Language and Cognition*, *21*(4), 765–781. https://doi.org/10.1017/S1366728915000279

Zhou, Y., McBride, C., Leung, J. S. M., Wang, Y., Joshi, M., & Farver, J. (2018). Chinese and English reading-related skills in L1 and L2 Chinese-speaking children in Hong Kong. *Language, Cognition and Neuroscience*, *33*(3), 300–312. https://doi.org/10.1080/23273798.2017.1342848

Zwaan, R. A. (2016). Situation models , mental simulations , and abstract concepts in discourse comprehension. *Psychonomic Bulletin & Review*, *23*, 1028–1034. https://doi.org/10.3758/s13423-015-0864-x

Appendix A: The four texts use as stimuli in the eye-tracking experiment (master copies – all experimental factors marked)

Text 1: Cardiovascular Disease – A tale of good news and bad

Deaths from cardiovascular disease (CVD), which includes heart attacks and high blood pressure, first began to decrease in the 1960s in Western countries and have continued to do so ever since. In the United States, for example, statistics show that death rates from *[slot_1]* **THESE** [heart/brain] diseases fell by over 35 percent between 1980 and 1997. Similar decreases have occurred in other Western countries. So, what are the factors that could have brought about *[slot_2]* **THESE** [encouraging/discouraging] reductions in the rates of heart disease? One factor is that well informed people are willing to change to healthier lifestyles in order to reduce their risk of (CVD). Another important factor is that the diagnosis and treatment of heart disease has improved. These advances have made possible the early diagnosis of *[slot_3]* **THESE** [dangerous/mild] illnesses that may cause serious problems later. As a result, doctors can more easily treat conditions which are detected early, either by prescribing drugs to lower blood pressure or by performing minor surgery. What's more, for the many patients who in the past have already suffered *[slot_4]* **SUCH** [health/language] problems the news is also good. Thanks to more high-tech equipment, better drugs and more experienced surgeons, many people now recover and go onto lead active lives after operations that were impossible just four or five decades ago

The news about cardiovascular disease, however, is not all positive. Firstly, in spite of *[slot_5]* **THESE** [improving/worsening] techniques for treating heart related problems, heart disease is still the most common cause of death in most Western countries. The statistics for the year 2000 in the United States, for example, show that it was responsible for more than 30 percent of deaths in that year, and was the single greatest cause of death overall. The second piece of bad news is that CVD, which was once considered a problem exclusively in wealthier countries, is now becoming a major health problem in developing countries too. Here, rapid social development is introducing the lifestyle patterns that we usually associate with CVD – a decrease in physical activity, an increase in smoking, and a change to a less healthy diet. Unfortunately *[slot_6]* **THESE** [risky/harmless] behaviours are widely predicted to increase in the years to come. a result, the World Health Organisation predicts that CVD will be the leading cause of death in the developing world within ten years. So, what can be done to fight [slot_7] THIS [fatal/minor] disease in the years to come?

The surgical treatment of heart disease will continue to benefit patients in countries where the expertise, equipment and other resources are available. By itself, however, *[slot_8]* **SUCH** [surgical/simplistic] treatment is not an adequate response to the global CVD problem. It is so expensive that it is an enormous burden even for the wealthy nations that can afford to offer it at present. For developing nations, which often do not have the necessary financial, technological and human resources, this approach is simply not possible. It seems more affordable solutions will have to be found.

Text 2: Heart disease and changing attitudes

In the last decades of the twentieth century, medical researchers showed that heart diseases, including high blood pressure or heart attacks are often associated with certain danger factors, such as smoking tobacco, poor diet or lack of exercise. Doctors and other health experts began to emphasize that we can reduce the risk of developing [slot_1] **SUCH** [health/musical] difficulties by making small changes to our daily habits. As a result, people now understand that there is a connection between heart disease and lifestyle, and that we can significantly improve our chances of living longer, healthier, altogether better lives by substantially reducing or even completely eliminating [slot_2] **THESE** [negative/positive] behaviours from our daily routines. Encouragingly, it seems [slot_3] **THIS** [growing/decreasing] awareness of how our lifestyle choices impact on our health is now starting to influence people's behaviour.

For example, in the past, people tended to think it was sufficient to have access to doctors on whose advice they could rely. Now, people understand that merely seeking out [slot_4] **SUCH** [medical/military] expertise when illness or injury occurs is not enough. As a result, societal attitudes about who is primarily responsible for maintaining and improving public health have undergone significant improvement in recent years. Furthermore [slot_5] **THIS** [encouraging/annoying] improvement in how people promote their own well-being can also be seen in some other behavioural changes that have occurred since the 1970s. For instance, in the United States today, many smokers have broken the habit, and fewer have taken it up. As a result, the percentage of smokers is far fewer than in the 1960s and 1970.

Similarly, many people are becoming aware of the benefits of gentle exercise, like walking, or swimming. Importantly, [slot_6] **THESE** [moderate/extreme] activities are considered to be the most beneficial for promoting a healthy heart. The health benefits from all these changes in attitude and behaviour are clear. Since the 1950s, the number of people dying from heart disease has fallen sharply. It is undoubtedly true that a partial explanation for [slot_7] **THESE** [improving/worsening] outcomes is that better diagnosis or treatment is helping people avoid or survive heart attacks. However, health experts have no doubt that in

recent years people have started to adopt more health-conscious living habits, and it is [slot_8] **THESE** [healthier/unhealthier] lifestyles which have led to much of the improvement in public health.

Text 3: Medicine and genetic (DNA) research: promise and problems

Since the 1970s, medicine has been benefiting from genetic research. Specifically, research in this field has allowed scientists to identify the genes associated with some serious genetic diseases. In recent years, *[slot_1]* **THESE** [exciting/worrying] advances have helped researchers develop tests that can identify and warn people who are at risk from developing *[slot_2]* **SUCH** [dangerous/pleasing] disorders at some time during their lives. Increasingly, doctors are very excited by *[slot_3]* **THESE** [innovative/ ancient] discoveries, as they have offered new ways to prevent some serious genetic disorders, such as Cystic Fibrosis. Most simply, doctors can now inform the people who are unfortunate enough to carry the genes which cause *[slot_4]* **SUCH** [harmful/harmless] conditions about the risk of passing them to their children.

Consequently, people with genes that cause fatal diseases like Huntington's disease, a condition which leads to severe loss of muscle control and almost total memory loss, often decide not to have children, as they are afraid they will suffer *[slot_5]* **THESE** [serious/minor] symptoms if the gene is passed down to them. In recent years, developments in the field of genetic medicine have led to an increase in funding for research in this area. Excitingly, *[slot_6]* **THIS** [growing/decreasing] interest has led to the development of new ways to treat many genetic illnesses. For example, in the year 2000, three babies who suffered from a rare disease they had inherited from their parents were treated using *[slot_7]* **SUCH** [genetic/traditional] techniques in one of the first non-experimental trials of these methods. Ten months later, to their parents' delight, the children had completely recovered.

However, as well as *[slot_8]* **THESE** [encouraging/upsetting] outcomes, there have also been disappointments. For example**,** researchers predicted they would soon be able to cure serious diseases, like cancer. However, they ignored the fact that earlier successful experiments had only been done using small numbers of patients and had only been used to treat relatively simple genetic conditions. As a result, despite a considerable amount of research funding being poured into investigating genetic solutions for some of the most serious illnesses currently facing society, so far, no new treatments have become available for general use in hospitals.

Text 4: New approaches to health and wellness

Caroline Silva is pregnant for the second time and requires weekly visits to a clinic for check-ups so that the doctors can monitor her and her baby's health. However, she runs her own business, and finds it difficult to take time away to get to the hospital. Another patient, Zhang Bao, has cardiovascular disease; his heart is weak, and he has extremely high blood pressure. As a result of suffering from *[slot_1]* **THESE** [dangerous/valuable] symptoms he requires medication, as well as frequent blood tests and heart examinations to monitor his cardiovascular health. Unfortunately, there is no hospital near his home, so he has to take time off from work to go to a hospital in another town. For *[slot_2]* **THESE** [lucky/unlucky] patients the currently available medical care is both inconvenient and expensive, and not as safe as their doctors would like.

In the coming years however, their difficult situation is likely to change significantly for the better. As with so many developments in the medical professions, *[slot_3]* **THIS** [welcome/unwelcome] improvement in the way medical care is provided is being made possible by new technology, known as mobile health care, or 'mHealth'. Mobile health care is defined as all forms of health care that take advantage of mobile devices, like telephones. The development of mHealth solutions has become possible due to a combination of technological advances: smaller, more powerful sensors; increases in cloud-based computer power and storage; and wireless internet. There are a wide variety of different types of mHealth solutions; however, the recent rapid development of *[slot_4]* **THESE** [advanced/ancient] technologies has become available because of a single device – the smart phone. This healthcare revolution has been made possible by *[slot_5]* **THESE** [amazing/useless] devices which are capable of providing and transmitting data about people's health wirelessly. Before recently *[slot_6]* **THIS** [medical/financial] information could only be gathered in doctors' offices or hospitals.

Although there are many roles for mHealth, it is most widely used in the monitoring and management of health problems, particularly those associated with chronic, long term diseases. Monitoring comes in two primary forms and operates in two directions. In one form, patients wear sensors on their bodies, for example on a watch or skin patch. The power of *[slot_7]* **THESE** [wearable/breathable] machines lies in the fact that they can measure and record information about a wide range of bodily functions, such as body temperature, breathing and blood pressure and send it safely, quickly and accurately to a hospital to be analysed. For example, they can be worn by pregnant women and used to transmit data about their

babies' heart rate and other health details continuously and wirelessly to a laboratory or doctor. As an important result of *[slot_8]* **THIS** [reliable/unreliable] transmission of data as soon as problems of any kind start to develop, the relevant medical services can respond immediately, in a number of ways. For example, *[slot_9]* **SUCH** [immediate/delayed] response could involve adjusting medication or even admitting patients to a hospital. This continuous monitoring reduces the need for patients to come in for frequent tests, and can even help avoid the development of serious health problems, by providing an early warning. Perhaps not surprisingly, the sale of these mobile sensors has grown dramatically in recent years.

However, while *[slot_10]* **THESE** [powerful/terrible] sensors send information from the patient directly to the doctor, a second form of mHealth solution works by sending information from expert doctors directly to less skilled health care providers, to allow them to care for their patients better. These advances in mHealth services are allowing health-care workers to provide care to patients who do not have easy access to heart disease specialists in hospitals. In China, for instance, there is

increasing concern about cardiovascular disease, which kills around three million people a year, partly because local doctors do not have the expertise to diagnosis problems and respond appropriately. Many of *[slot_11]* **THESE** [unnecessary/necessary] fatalities could be prevented by regular heart monitoring, a service which is difficult to provide to patients who live outside of large cities. Today nurses and local health care workers in China are responding to *[slot_12]* **THESE** [significant/insignificant] challenges by using a handheld device to collect patient data. It records thirty seconds of data about the patient's heart and transmits it to a central facility in Beijing. There, doctors and technicians analyse the data and then provide an immediate diagnosis and recommend a treatment plan. The program is already having a major impact on the annual heart attack rate in China, and over the coming years could save a considerable number of lives.

Appendix B: The vocabulary test (Schmidt et al, 2001)

The 2000 word level 1 copy

2 event	___6_ end or highest point
3 motor	___3_ this moves a car
4 pity	___1_ thing made to be like another
5 profit	
6 tip	

1 accident	
2 debt	_____ loud deep sound
3 fortune	_____ something you must pay
4 pride	_____ having a high opinion of yourself
5 roar	
6 thread	

1 coffee	
2 disease	_____ money for work
3 justice	_____ a piece of clothing
4 skirt	_____ using the law in the right way
5 stage	
6 wage	

1 clerk	
2 frame	_____ a drink
3 noise	_____ office worker
4 respect	_____ unwanted sound
5 theater	
6 wine	

1 dozen	
2 empire	_____ chance
3 gift	_____ twelve
4 opportunity	_____ money paid to the government
5 relief	
6 tax	

1 admire	
2 complain	_____ make wider or longer
3 fix	_____ bring in for the first time
4 hire	_____ have a high opinion of someone
5 introduce	
6 stretch	

1 arrange	
2 develop	_____ grow
3 lean	_____ put in order
4 owe	_____ like more than something else
5 prefer	
6 seize	

1 blame	
2 elect	_____ make
3 jump	_____ choose by voting
4 manufacture	_____ become like water
5 melt	
6 threaten	

1 ancient	
2 curious	_____ not easy
3 difficult	_____ very old
4 entire	_____ related to God
5 holy	
6 social	

1 bitter	
2 independent	_____ beautiful
3 lovely	_____ small
4 merry	_____ liked by many people
5 popular	
6 slight	

The 3000 word level

1 bull	
2 champion	_____ formal and serious manner
3 dignity	_____ winner of a sporting event
4 hell	_____ building where valuable objects are shown
5 museum	
6 solution	

1 blanket	
2 contest	_____ holiday
3 generation	_____ good quality
4 merit	_____ wool covering used on beds
5 plot	
6 vacation	

1 comment	
2 gown	_____ long formal dress
3 import	_____ goods from a foreign country
4 nerve	
5 pasture	_____ part of the body which carries feeling
6 tradition	

1 administration
2 angel _____ group of animals
3 frost _____ spirit who serves God
4 herd _____ managing business and
5 fort affairs
6 pond

1 atmosphere
2 counsel _____ advice
3 factor _____ a place covered with
grass
4 hen _____ female chicken
5 lawn
6 muscle

1 abandon
2 dwell _____ live in a place
3 oblige _____ follow in order to catch
4 pursue _____ leave something
5 quote permanently
6 resolve

1 assemble
2 attach _____ look closely
3 peer _____ stop doing something
4 quit _____ cry out loudly in fear
5 scream
6 toss

1 drift
2 endure _____ suffer patiently
3 grasp _____ join wool threads
together
4 knit _____ hold firmly with your
hands
5 register
6 tumble

1 brilliant
2 distinct _____ thin
3 magic _____ steady
4 naked _____ without clothes
5 slender
6 stable

1 aware
2 blank _____ usual
3 desperate _____ best or most important
4 normal _____ knowing what is
happening
5 striking
6 supreme

The 5000 word level

1 analysis
2 curb _____ eagerness
3 gravel _____ loan to buy a house
4 mortgage _____ small stones mixed with
5 scar sand
6 zeal

1 cavalry
2 eve _____ small hill
3 ham _____ day or night before a
4 mound holiday
5 steak _____ soldiers who fight from
6 switch horses

1 circus
2 jungle _____ musical instrument
3 nomination _____ seat without a back or
4 sermon arms
5 stool _____ speech given by a priest
in
6 trumpet a church

1 artillery
2 creed _____ a kind of tree
3 hydrogen _____ system of belief
4 maple _____ large gun on wheels
5 pork
6 streak

1 chart
2 forge _____ map
3 mansion _____ large beautiful house
4 outfit _____ place where metals are
5 sample made and shaped
6 volunteer

1 contemplate
2 extract _____ think about deeply
3 gamble _____ bring back to health
4 launch _____ make someone angry
5 provoke
6 revive

1 demonstrate
2 embarrass _____ have a rest
3 heave _____ break suddenly into small
4 obscure pieces
5 relax _____ make someone feel shy
or
6 shatter nervous

1 correspond
2 embroider _____ exchange letters
3 lurk _____ hide and wait for
someone
4 penetrate _____ feel angry about
something
5 prescribe
6 resent

1 decent
2 frail _____ weak
3 harsh _____ concerning a city
4 incredible _____ difficult to believe
5 municipal
6 specific

1 adequate
2 internal _____ enough
3 mature _____ fully grown
4 profound _____ alone away from other
5 solitary things
6 tragic

The 10,000 word level

1 alabaster
2 chandelier _____ small barrel
3 dogma _____ soft white stone
4 keg _____ tool for shaping wood
5 rasp
6 tentacle

1 benevolence
2 convoy _____ kindness
3 lien _____ set of musical notes
4 octave _____ speed control for an
5 stint engine
6 throttle

1 bourgeois
2 brocade _____ middle class people
3 consonant _____ row or level of something
4 prelude _____ cloth with a pattern or gold
5 stupor or silver threads
6 tier

1 alcove
2 impetus _____ priest
3 maggot _____ release from prison early
4 parole _____ medicine to put on
wounds
5 salve
6 vicar

1 alkali
2 banter _____ light joking talk
3 coop _____ a rank of British nobility
4 mosaic _____ picture made of small
pieces
5 stealth of glass or stone
6 viscount

1 dissipate
2 flaunt _____ steal
3 impede _____ scatter or vanish
4 loot _____ twist the body about
5 squirm uncomfortably
6 vie

1 contaminate
2 cringe _____ write carelessly
3 immerse _____ move back because of
fear
4 peek _____ put something under water
5 relay
6 scrawl

1 blurt
2 dabble _____ walk in a proud way
3 dent _____ kill by squeezing
someone's
4 pacify throat
5 strangle _____ say suddenly without
6 swagger thinking

1 illicit
2 lewd _____ immense
3 mammoth _____ against the law
4 slick _____ wanting revenge
5 temporal
6 vindictive

1 indolent
2 nocturnal _____ lazy
3 obsolete _____ no longer used
4 torrid _____ clever and tricky
5 translucent
6 wily

Academic vocabulary

1 area
2 contract _____ written agreement
3 definition _____ way of doing something
4 evidence _____ reason for believing
5 method something is or is not true
6 role

1 debate
2 exposure _____ plan
3 integration _____ choice
4 option _____ joining something into a
5 scheme whole
6 stability

1 access
2 gender _____ male or female
3 implementation _____ study of the mind
4 license _____ entrance or way in
5 orientation
6 psychology

1 accumulation
2 edition _____ collecting things over time
3 guarantee _____ promise to repair a broken
4 media product
5 motivation _____ feeling a strong reason or
6 phenomenon need to do something

1 adult
2 exploitation _____ end
3 infrastructure _____ machine used to move
4 schedule people or goods
5 termination _____ list of things to do at
6 vehicle certain times

1 alter
2 coincide _____ change
3 deny _____ say something is not true
4 devote _____ describe clearly and
exactly
5 release
6 specify

1 correspond
2 diminish _____ keep
3 emerge _____ match or be in agreement
4 highlight with
5 invoke _____ give special attention
6 retain to something

1 bond
2 channel _____ make smaller
3 estimate _____ guess the number or size
4 identify of something
5 mediate _____ recognizing and naming
6 minimize a person or thing

1 explicit
2 final _____ last
3 negative _____ stiff
4 professional _____ meaning `no' or `not'
5 rigid
6 sole

1 abstract
2 adjacent _____ next to
3 controversial _____ added to
4 global _____ concerning the whole
world
5 neutral
6 supplementary

Appendix C: The grammaticality judgement test

Instructions:

Read each of the sentences and decide if you think it is correct or incorrect. Remember, this is a grammar test – you only need to look for problems with the grammar, not the vocabulary, punctuation or logic of the sentences. There is only ever one mistake in a sentence.

If you think a sentence is ***correct***, please circle the ✓ symbol on the right-hand side of the page. If you feel a sentence is ***incorrect***, please circle the ✗ on the right- hand side of the page and change the part of the sentence you feel is incorrect. For some answers, you will only need to change one word. For others, you may need to rewrite the sentence. The important thing is to make it clear what you feel needs to be changed. Please see the examples below:

Examples:

a. When the young boy from the village arrived, the snake bit him on the leg.

✓ ✗

b. The police tried to arrested the man, but he escaped.

arrest

✓ ✗

c. The bread the woman ate.

The woman ate the bread ✓ ✗

d. The man from England came.

came ✓ ✗

Time taken:

Score:

1. **The beauty is something that lasts forever.**

_____ ✓ ✗

2. **The man who came to work in the bar last week was so unlucky to cut himself on a piece of glass on his first day!**

_____ ✓ ✗

3. **Last night the old lady who had been in the hospital since she fell down the stairs at her daughter's house die in her sleep.**

_____ ✓ ✗

4. **Bob has trying to fix Jim's car in his garage for nearly two days.**

_____ ✓ ✗

5. **Mary looked at the flowers but didn't buy until much later that day.**

_____ ✓ ✗

6. **A policeman gived Alan a ticket for speeding in the centre of town when he had only had his license for a week.**

_____ ✓ ✗

7. **The violet is a wonderful colour to paint a child's bedroom walls if he or she is feeling unhappy about something and you want to make the room more cheerful.**

✓ ✗

_____

8. **Peter's wife did not have any money on him when Peter arrived from the station.**

_____ ✓ ✗

9. **The boy who caught the ball almost dropped it last time.**

_____ ✓ ✗

10. **The little boy has been write to the policeman who helped him get home for several years.**

_____ ✓ ✗

11. The books fell off the shelves when the regular train came roaring past the couple's window for the second time in only twenty minutes.

_____ ✓ ✗

12. The woman who had just arrived from Liverpool the policeman asked a question.

_____ ✓ ✗

13. John sang regularly for the church choir before his operation, even though he was always a heavy smoker.

_____ ✓ ✗

14. John took a sweater along, but didn't put it on as the weather was hotter than expected.

_____ ✓ ✗

15. The dinner the man who came yesterday burned.

_____ ✓ ✗

16. Sandy, the girl who is visiting from Germany, filled a glass of water to take to bed with her.

_____ ✓ ✗

17. The boy is helping the man build a house.

_____ ✓ ✗

18. The man who used to live with us has been lying regularly to his father about what happened for several years.

_____ ✓ ✗

19. Mrs. Johnson went to the library for the twenty-first time today.

_____ ✓ ✗

20. The children have been playing in the garden till dark recently.

_____

✓ ✗

21. Linda for his birthday a cake baked John.

✓ ✗

22. Lucy fell from a ladder at work, but despite hurting himself quite badly, did not break any bones.

✓ ✗

Appendix D: The English and Chinese working memory capacity tests

Participant number: _____

DIGIT SPAN

INSTRUCTIONS FOR DIGITS FORWARD:
Say "I am going to say some numbers. Listen carefully, and when I am through say them right after me." Read the digits at a rate of one per second. Administer two trials of each item, even if the first trial is passed. Discontinue after failure on two trials of any item. Score 1 point for each trial passed.
(**Chinese version**)

DIGITS FORWARD					Score
SPAN	Trial 1	Pass-Fail	Trial 2	Pass-Fail	2,1, or 0
2	3-7		5-2		
3	6-2-9		4-7-2		
4	5-4-1-7		8-3-9-6		
5	2-6-9-3-5		5-9-4-7-3		
6	9-1-8-4-2-7		8-3-5-1-4-6		
7	1-4-8-5-2-7-3		2-8-1-6-9-5-4		
8	5-2-1-9-3-7-4-8		2-9-6-1-3-8-5-7		
9	4-9-6-5-1-3-7-2-8		2-7-5-1-3-8-6-9-4		
Total Forward: Max=16					

INSTRUCTIONS FOR DIGITS BACKWARD:
Say "Now I am going to say some more numbers, but this time when I stop I want you to say them backwards. For example, if I say 2-8-3, what would you say?" Pause for the subject to respond. ----If the subject responds correctly, say "That's right" and proceed to Trial 1 of item 1.

Say "Now listen to these numbers, and remember you are to say them backwards." ----If the subject fails the example, say "No, I said 2-8-3, so to say them backwards you would need to say 3-8-2. Now try these numbers. Remember you are to say them backwards. Ready? 1-5-8." Give no help on this second example or any of the trials that follow. Whether the subject passes or fails with the second example, proceed to trial1 of Items. Read the digits at a rate of one per second. Administer two trials of each item, even if the first trial is passed. Discontinue after failure on two trials of any item. Score 1 point for each trial passed.

DIGITS BACKWARD					Score
SPAN	Trial 1	Pass-Fail	Trial 2	Pass-Fail	2,1, or 0
2	8-3		4-9		
3	5-9-7		3-2-8		
4	1-8-2-5		2-7-9-4		
5	6-2-9-5-3		1-9-6-3-8		
6	7-1-5-3-9-4		2-6-9-7-3-8		
7	4-9-1-5-2-8-3		6-1-9-4-8-2-7		
8	3-5-8-1-7-4-9-6		5-2-6-1-9-4-8-3		
Total Backward: Max=14					

Overall total:
Max=30 _____

Digit Span Forward=_____(highest number of digits for which 1 trial was passed)
Digit Span Backward=_____(highest number of digits for which 1 trial was passed)

(English version)

DIGITS FORWARD					Score
SPAN	Trial 1	Pass-Fail	Trial 2	Pass-Fail	2,1, or 0
2	4-9		1-6		
3	3-8-4		8-3-5		
4	6-1-5-7		2-9-5-3		
5	3-9-4-7-5		8-6-1-7-3		
6	8-1-9-2-4-7		7-1-5-8-3-6		
7	7-2-8-1-3-5-9		8-3-6-1-9-7-2		
8	6-3-1-9-2-7-5-8		7-5-1-4-9-2-6-3		
9	6-9-4-5-1-7-3-2-8		5-9-8-2-6-3-7-4-1		
Total Forward: Max=16					

DIGITS BACKWARD					Score
SPAN	Trial 1	Pass-Fail	Trial 2	Pass-Fail	2,1, or 0
2	7-3		9-2		
3	6-4-9		3-8-5		
4	4-7-2-9		5-2-6-4		
5	9-2-8-5-7		6-1-4-9-5		
6	8-1-5-3-4-7		6-1-7-4-3-9		
7	7-9-2-5-4-8-1		3-8-6-9-4-7-2		
8	6-5-9-1-7-4-8-2		8-6-9-3-1-7-2-5		
Total Backward: Max=14					

Overall total:
Max=30 _____

Digit Span Forward=_____(highest number of digits for which 1 trial was passed)
Digit Span Backward=_____(highest number of digits for which 1 trial was passed)

Appendix E: Data viewer four stage cleaning process

Data Viewer's four-stage cleaning works as an iterative process. In **stage 1**, "Data Viewer checks whether each fixation's duration is shorter or equal to the Stage 1 Duration Threshold"(SR Research, 2018, p. 48). In this experiment, the threshold was set at 80ms, as fixations below this are unlikely to be sufficient to extract introduction (Rayner, 1998).In addition, "the software further checks the duration and distance of the fixations immediately before and after the current fixation. The fixation will be merged to one of its neighbouring fixations if the neighbour's duration is longer than the threshold value and its distance along the x-axis … from the current fixation is shorter or equal to the Stage 1 Distance Threshold" (SR Research, 2018, p. 48), which in this case was set at 0.5 degree, in accordance with the predetermined setting. **Stage 2** of the process operates in the same way as stage 1, except that a shorter duration threshold and a longer distance threshold are used. Once again, the pre-set values in Data Viewer were used for both the duration (40ms) and distance (1.25^{o}) thresholds. In **stage 3** of the process, "Data Viewer searches for interest areas that include at least three fixations shorter than the Stage 3 Duration Threshold value and no fixations longer than the Duration Threshold (in this case set at 140ms). In such cases the shorter fixations are merged into a single fixation" (SR Research, 2018, p. 48). In **stage 4** of the process, Data Viewer deletes all fixations below a minimum threshold, which in this case was set at 80ms and above a maximum threshold, in this case 1400ms.

Appendix F: Logo 10 data tables for descriptive statistics for all four eye-tracking measures

Descriptive statistics and Kolmogorov-Smirnov test for first fixation duration in all four experimental conditions after Log 10 transformation

			K-S test	
	M (ms)	*SD*	*D*(82)	*Sig.*
Determiner-Congruent condition	2.38	0.07	.054	.200
Determiner- Incongruent condition	2.38	0.08	.073	.200
No Determiner-Congruent condition	2.41	0.07	.059	.200
No Determiner-Incongruent condition	2.41	0.08	.065	.200

Descriptive statistics and Kolmogorov-Smirnov test for gaze duration in all four experimental conditions after Log 10 transformation

			K-S test	
	M	*SD*	*D*(82)	*Sig.*
Determiner-Congruent condition	2.57	0.12	.069	.200
Determiner- Incongruent condition	2.55	0.12	.071	.200
No Determiner-Congruent condition	2.58	0.11	.045	.200
No Determiner-Incongruent condition	2.58	0.13	.055	.200

Descriptive statistics and Kolmogorov-Smirnov test for selective regression path duration in all four experimental conditions after Log 10 transformation

			K-S test	
	M (ms)	*SD*	*D*(82)	*Sig.*
Determiner-Congruent condition	2.63	0.12	.055	.200
Determiner- Incongruent condition	2.64	0.15	.065	.200
No Determiner-Congruent condition	2.63	0.12	.057	.200
No Determiner-Incongruent condition	2.67	0.12	.068	.200

Descriptive statistics and Kolmogorov-Smirnov test for second pass time in all four experimental conditions after Log 10 transformation

			K-S test	
	M (ms)	*SD*	*D*(82)	*Sig.*
Determiner-Congruent condition	2.79	0.16	.052	.200
Determiner- Incongruent condition	2.82	0.19	.054	.200
No Determiner-Congruent condition	2.81	0.16	.074	.200
No Determiner-Incongruent condition	2.86	0.17	.055	.200

Appendix G: Original excerpts from the eye-tracking interviews plus the translation

Translated excerpt 1

Original:

P032: What's more 另外, for the many patients who in the past have already [indistinct while reads] 另外一些病人他们，过去的一些病人已经 suffered language problems 已经忍受了这种语言的问题 the news is good [indistinct] suffered a language problem, the news,也是好的 [indistinct]

Translation:

P032: What's more, *in addition*, for the many patients who in the past have already [indistinct while reads], *some additional patients, they, some past patients already,* suffered language problems, *already suffered this kind of language problem,* the news is good [indistinct], suffered a language problem, the news *is also good*

Translated excerpt 2

Original:

P025: ... death from, death 就是 die, 什么的 die, 然后 which 连接词这个后面是 which 的从句，然后这个是包括 high blood, is a kind of illness to, and this is and [points to and], so, this word [points to heart attack] is a kind of illness too.

R: Do you know this word [points to 'heart' in sentence (i)]?

P025: Heart, I think it's this is here [points to his heart], and high blood, I think.

R: okay.

P025: So, this word [points to 'cardiovascular disease' in sentence (i)] 就是这两种疾病的一个统称

R: ... how do we know 这是一个统称 [translation: this is a superordinate term]?

P025: 比如说 this which, there which 这个单词这后面是它的从句，这个从句，这个是，这个 from 这个是，这句话 [points to 'cardiovascular disease' in sentence (i)]是整个句子的主语，然后这个从句是修饰这个主语的 [points to 'cardiovascular disease' in sentence (i)].

R: So, what could we guess this word means [points to 'disease' in sentence (i)]?

P025: 这是这种类型的疾病的统称，也是一个专有名词.

Translation:

P025: ... death from, death *that* is die, *some kind of* die, *then,* which *a connective, this one after is a* which *attributive clause*， *then*, *this includes* high blood, is a kind of illness to, and this is and [points to and], so, this word [points to heart attack] is a kind of illness too.

R: Do you know this word [points to 'heart' in sentence (i)]?

P025: Heart, I think it's this is here [points to his heart], and high blood, I think.

R: okay.

P025: So, this word [points to 'cardiovascular disease' in sentence (i)] *is a superordinate term for these two kinds of disease*

R: ... how do we know *this is a superordinate term*?

P025: *for example*, this which, there which *this word* [points to 'which' in sentence (i)] *afterwards is its attributive clause, this attributive clause, this is, this* from, *this is, this whole sentence's subject, then this attributive clause is modifying this subject* [points to 'cardiovascular disease']

R: So, what could we guess this word means [points to 'disease' in sentence (i)]?

P025: *this is the superordinate term for this kind of disease, it's also a technical term*

Translated excerpt 3:

Original:

P025: In the United States, 这是在美国, for example such as, [indistinct] 这个单词 [points to 'statistics' in sentence (ii)] I don't know this meaning，but 这个单词 [points to 'statistics' in sentence (ii)] 是这句话的主语，然后这是动词 [points to 'show' in sentence (ii)]，这是后面接的是 that 从句，然后这个是 death rates from this brain disease fell by 这个 death rate 就是，the advantage of the death from this 就是这个大脑的疾病的,所以说可以推断出来，它这个 [points at 'statistics' in sentence (ii)] 指的应该是一篇文章或者一个报道。

R: Why do you think that?

P025: 这是主语嘛 [points to 'statistics' in sentence (ii)]，它的后面就是说, 然后, 后面展示了大脑的这个疾病的死亡率, 死亡率，这些年，在超过 35%，然后在 1980 和 1997 年, 也就是说这后面的这些句子，怎么说呢，这后面这些是 that 后面就是从句嘛，这个是主句 [points to 'statistics' in sentence (ii)]，这个就是以岁数 [indistinct] 判断它，就是一篇文章或者说一个报道，才能说这些算是严格的数据吧.

Translation:

P025: In the United States *this is in America*, for example, such as [indistinct] *I don't know this word* [points to 'statistics' in sentence (*ii*)] I don't know this meaning, but *this word* [points to 'statistics' in sentence (*ii*)] *is the subject of this sentence ... this is a verb* [points to 'show' in sentence (*ii*)], *this is followed by a* that *attributive clause, ... this is* death rates from this brain disease fell by, *this* death rate *is* the advantage of the death from this, *that is, this brain disease, so I infer this* [points at 'statistics' in sentence (*ii*)] *probably refers to an article or a report*

R: Why do you think that?

P025: *This is the subject* [points to 'statistics' in sentence (ii)] *the bit after this says... the bit after shows this brain disease's death rate, death rate, these years, exceeded 35% ... in 1980 and 1997, it also says, these clauses after, how can I put this, this bit after, this is* that *it's an attributive clause, this is the main clause*

[points to 'statistics' in sentence (ii)], *this, that is according to the age*, (I) *can judge it is an article or a report, in order to have what can be considered rigorous data*

Translated excerpt 4

Original:

P033: ...but he said death, I guess this is the name of some like 疾病 [translation: disease]， right, so it doesn't matter, just know his name is CVD [points to CVD in sentence (i)] ... ...and includes heart attacks and high blood pressure, like heart [points to her own heart] and high blood, I know the meaning... it's two kinds of illness, it' like your heart and high blood pressure, can I explain in Chinese, it's 高血压 [translation: high blood pressure].

Translation:

P033: ...but he said death, I guess this is the name of some like *disease*， right, so it doesn't matter, just know his name is CVD [points to CVD in sentence (i)] ... and includes heart attacks and high blood pressure, like heart [points to her own heart] and high blood, I know the meaning... it's two kinds of illness, it' like your heart and high blood pressure, can I explain in Chinese, it's *high blood pressure*.

Translated excerpt 5

Original:

P033: so 这个疾病，从，在西方国家从一九六零年开始减少, [**translation**: '*so'*, this disease, from, in western countries from 1960 started to decrease], right, continued to do so ever since, and from now they always decreases the illness from this time, right?

Translation

P033: so, *this disease, from, in western countries from 1960 started to decrease*, right, continued to do so ever since, and from now they always decreases the illness from this time, right?

Translated excerpt 6

Original:

P033: ... in the United States, in US, so here is an example [points to for example sentence (ii)] in US, statistics shows the death rate from this [indistinct while reading] so, 统计，一个统计机构显示，因为这个脑部疾病的 [points to brain disease in sentence (ii)] 死亡率从，，在这个[points at 1980 and 1997 in sentence (ii)] this time to this time， between this time 下降了，下降到了三十五 percent, right， just because I know the, I know the meaning of these words.

Translation:

P033 ... in the United States, in US, so here is an example [points to for example sentence (*ii*)] in US, statistics shows the death rate from this [indistinct while reading] so, *statistics, a statistical agency shows, because the death rate from this brain disease* [points to brain disease in sentence (*ii*)] *from, at this* [points at 1980 and 1997 in sentence (*ii*)], this time to this time， between this time, *decreased, decreased to thirty five* percent, right， just because I know the, I know the meaning of these words

Translated excerpt 7

Original:

P071: this is for the 主谓宾

Translation:

P071: this is for the *subject-verb-object*

Translated excerpt 8

Original:

P071: Yeah， because I... search about the temperature and it is related with the 血管 [translation: blood vessels] 跟心脏跳动 [translation: heartbeat]，so this is about the 血管 [translation: blood vessels].

Translation:

P071: Yeah， because I... search about the temperature and it is related with the *blood vessels, with the heartbeat*, so this is about the *blood vessels*

Translated excerpt 9

Original:

P032: ...很多人, 心血管儿疾病，包括心脏 [pauses on 'attacks' in sentence (i)] 这个后面的是高血压 [points to 'high blood pressure' in line sentence (i)] 是吗， high blood pressure 高血压, 前面这个 [points to 'heart attacks' in sentence (i)] 心脏，这个我说不…这个应该是个专业的词，然后， first began to [indistinct as she reads] 在西方国家第一个人是开始得这个疾病的时候一九六零年， 哦， 不，是六十年代上一世纪六十年代，然后继续， continue to do so ever since，就继续，持续这个问题

Translation:

P032: ...*many people cardiovascular disease includes heart* [pauses on 'attacks' in sentence (*i*)], *this latter one is high blood pressure* [points to 'high blood pressure' in line sentence (*i*)], *right*, high blood pressure, *high blood pressure, this first one* [points to 'heart attacks' in sentence (*i*)] *heart, this I don't, this one is probably a technical term, next*, first began to [indistinct as she reads], *in western countries the person started to catch this disease was in 1960, oh, no, it's the 1960s, last century, the 1960s, then, continue,* continue to do so ever since, continue, *continue this problem*]

Translated excerpt 10

Original:

P: 同样的这种疾病 have 发生在，同样的这种疾病也发生在其他的西方的国家

Translation:

P032*:similar diseases of this kind* have *occurred in, similar diseases of this kind also occur in other Western countries*

Translated excerpt 11

Original:

P032: ... suffered 是忍受的意思，对吗？忍受语言的问题，这个消息还是个好消息，不知道我是不是 [sounds like xiang pian le – check (30:34)] 它是说这些过去有一些病人他们患了心血管疾病之后它会有语言的障碍嘛，就是，他恢复了以后它会说话也许跟之前不一样嘛 ...因为在中国，好多人得心血管疾病他们恢复了以后都，走，那个身体不协调走路的时候有这样，[indistinct] 半个身子都不 [indistinct] 有的人说话也说不清楚，就是很模糊， not clear

Translation:

P032: ... suffered *means suffered, right? Suffered language problems, this new is still good news, I don't know if I have got this wrong, (but) are they saying that these patients in the past who had cardiovascular disease had some language obstructions afterwards, that is, after the recovered they couldn't speak the same as before, because in China, for many people, after they recover from cardiovascular disease, walking, their body is not consistent when they are walking, like, half their body can't [indistinct], there are other people who aren't clear when they speak, they sound blurry*, not clear

Translated excerpt 12

Original:

P004: Deaths from this one CVD [points at 'CVD' in sentence (i)] , [indistinct while reading], first began to disease [misreads 'decrease'] in the 1960s in Western countries, and have continued to, to [indistinct while reading] oh, this sentence is just like what is the CVD include and it's like [indistinct] and the high, maybe, the highlight word is heart attack and high blood, blood pressure [points to 'heart attack and high blood pressure' in sentence (i)] and it's the time what it begins and it's the time it says what it continues so it's the 现状

Translation:

P004: Deaths from this one CVD [points at 'CVD' in sentence (i)] , [indistinct while reading], first began to disease [misreads 'decrease'] in the 1960s in Western countries, and have continued to, to [indistinct while reading] oh, this sentence is just like what is the CVD include and it's like [indistinct] and the high, maybe, the highlight word is heart attack and high blood, blood pressure [points to 'heart attack and high blood pressure' in sentence (i)] and it's the time what it begins and it's the time it says what it continues so it's the *current situation*

Translated excerpt 13

Original:

P004: High blood 高压, 心脏病. [translation: high pressure, heart disease] Um, I don't, forget it... this kind of disease [points to 'disease' in sentence (i)] includes heart attacks and high blood, and these two [draws line linking 'heart attacks and high blood' in sentence (i) and labels them 'diseases'] are the two kind of diseases, right ? But this is the whole article [indicates whole text beneath], God! [resumes reading] first began to decrease, this kind of disease [points to 'disease' in sentence (i)] was fined, sorry, um, the people the

population have this kind of disease [points to ‘disease’ in sentence (i)] decreased since 1960s, in western countries, right?

Translation:

P004: High blood, *high pressure, heart disease*, um, I don't, forget it... this kind of disease [points to ‘disease’ in sentence (i)] includes heart attacks and high blood, and these two [draws line linking ‘heart attacks and high blood’ in sentence (i) and labels them ‘diseases’] are the two kind of diseases, right ? But this is the whole article [indicates whole text beneath], God! [resumes reading] first began to decrease, this kind of disease [points to ‘disease’ in sentence (i)] was fined, sorry, um, the people the population have this kind of disease [points to ‘disease’ in sentence (i)] decreased since 1960s, in western countries, right?

Translated excerpt 14

Original:

P060: Here, rapid social development is introducing the lifestyle patterns that we usually associate with CVD – a decrease in physical activity, an increase in smoking, and a change to a less healthy diet. Here, a rapid social development 就是快速的这个，高速的社会发展，引进了，正在引进这种 lifestyle that we usually associate with CVD 然后经常和这个心血管疾病联系在一起, a decrease in physical activity, and an increase in smoking, a decrease in physical activity, 就是身体，运动活动的发展, 怎么翻译呢，就是 about the body activity, 有一个减少, 然后，他的减少，然后 smoking，吸烟人数的增多，还有这个 a change to a less healthy diet，就是这个生活习惯越来越不好，不好了，less, because of the less healthy diet

Translation:

P060: Here, rapid social development is introducing the lifestyle patterns that we usually associate with CVD – a decrease in physical activity, an increase in smoking, and a change to a less healthy diet. Here, a rapid social development, *that is, this kind of rapid, rapid social development, brings about, is bringing about this kind of* lifestyle that we usually associate with CVD ... *often linked together with this cardiovascular disease*, a decrease in physical activity, and an increase in smoking, a decrease in physical activity, *that is body, the development of sporting activities, how can I translate this, that is,* about the body activity, *had a decrease, then, its decrease, then*, smoking, *the number of people smoking increased, as well as this* a change to a less healthy diet, *that is this lifestyle habit is getting worse and worse, worse*, less, because of the less healthy diet’

Translated excerpt 15

Original:

P032: Here, rapid 社会的发展 is introducing, is introducing, [indistinct while reading] 社会的高度发展也带来了生活方式的改变，因此我们经常要 , it's ok, asso...[attempts associated] 眼前词想不起来了，我们也要，这词我想不起来了[points to associated] 就是, 这个病也可以有那种 physical，应该怎么说，物理治疗，不知道是这么说的, physical activity [indistinct while reading] increase, 它引起，就是吸烟会，会，吸烟会，会提升你的这个病的机率, and a change into [reading] 还有不见健康的饮食，你就应该去用一种健康的这种饮食才会降低你的 [indistinct]

Translation:

P032 Here, rapid *social development* is introducing, is introducing [indistinct while reading] *rapid social development is bringing lifestyle changes, because of this we often want*, it's ok, asso... [attempts 'associated'] *I can't think what the word I am looking at is, we also want, I can't think what this word is* [points to 'associated'] , *that is, this disease can also have that kind of* physical*, it's probably said like that, physical treatment, I don't know if it's said like this*, physical activity [indistinct while reading] increase, *it leads to, that is, smoking can, can, smoking can, can increase you rate of this disease*, and a change into [indistinct while reading] , *as well as an unhealthy diet, you, probably you can use a kind of healthy, this kind of diet can reduce your* [indistinct]]

Translated excerpt 16

Original:

P060: Unfortunately, these harmless behaviours are widely predicted to increase in the years to come, 这些，这些不好的行为，额，逐渐 predicted to increase，predicted to increase in the years to come, 就是这些不好的行为逐渐在增多，在最近几年里，在之后的几年, predict，I think that means the things will happen

Translation:

P060: Unfortunately, these harmless behaviours are widely predicted to increase in the years to come, *these, these bad behaviours, the amount, gradually* predicted to increase, predicted to increase in the years to come *that is, these bad behaviours are gradually increasing, in the past few years, in the years after this*, predict, I think that means the things will happen

Translated excerpt 17

Original:

P032: Unfortunately, 还有那种不良的生活习惯 [underlines harmless in sentence (ii)]，harmful, 有害的习惯, and [indistinct while reading] predicted, 猜测 to，这个词应该是猜测的意思把 [circles predicted in sentence (ii)]，我记得，是不是猜测的呀，predicted，猜测 [indistinct while reading] 就是这些，是这些方式 [gesticulates loosely to all earlier text] 有可能是导致很多人都得了心血管儿疾病

Translation:

P032: Unfortunately *there are still those kinds of bad habits* [underlines harmless in sentence (*ii*)], harmful, *harmful habits* and [indistinct while reading] predicted *guess* to *this word probably means guess* [circles predicted in sentence (*ii*)], *I remember, is it guess*, predicted *guess* [indistinct while reading], *that is those, its these habits* [gesticulates loosely to all earlier text] *can possibly cause many people to catch cardiovascular disease*.