Language Processing in Bilingual Children and Adults: Evidence from Filler-Gap Dependencies and Garden Path Sentences

PhD in Psycholinguistics

School of Psychology and Clinical Language Sciences

George Pontikas

August 2019

Abstract

The present thesis examines morphosyntactic processing in bilingual children and adults. It bridges gaps in the existing literature in three ways. Firstly, previous work on bilingual children has focused on inflectional morphology but has not examined the timecourse of processing in terms of misinterpretation and real time reanalysis or use of information of different sources to aid disambiguation. Second, it extends the use of the visual world eye-tracking paradigm to research in morphosyntactic processing in bilingual children. Third, for the adults, it compares early/native bilinguals to monolinguals and late bilinguals / L2 learners. Two linguistic phenomena were investigated with adults and children; which-questions and garden-path sentences in English. Overall, both bilingual children and adults showed qualitatively similar patterns of processing to their monolingual counterparts. All groups experienced greater difficulty with structures where there was ambiguity and a need for syntactic reanalysis, i.e. object which-questions and garden-path sentences, suggesting incremental processing. The main difference between monolinguals and bilinguals is that of speed; bilinguals appeared to process sentences slower than monolinguals even when their comprehension accuracy was equally as high. This difference was found for both children and adults, as evidenced by the reaction times or changes in the gaze data, and was generally not more pronounced in sentences where reanalysis is required. With regards to the bilingual adults, the early/native bilinguals clustered with the L2 learners in terms of processing speed but more so with the monolingual adults in terms of accuracy. The bilingual groups showed a reduced utilisation of information from various sources to aid processing. Bilingual adults and children made use of number mismatch between the two noun phrases in the study on which-questions to facilitate disambiguation; however, they showed this effect just for off-line comprehension accuracy and not for real time processing, i.e. in the gaze data. The bilinguals also did not show consistent use of referential context to disambiguate in the study on garden path sentences, although this was also the case for the monolingual adults and children. In sum, the results from the studies in this thesis suggest the both bilingual adults and children were equally as able as their monolingual counterparts at an end stage but differed to the monolinguals on more fine-grained measures of real time processing. These measures point to qualitatively similar but more protracted over time processing for bilinguals and with more limited use of facilitatory information to disambiguate.

Declaration of original authorship

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

George Pontikas

Acknowledgements

There are a number of people I would like to thank for their support. This was not just essential for me to complete my PhD but also made this long journey a lot more pleasurable even in the most frustrating of times. First and foremost, I would like to thank my three supervisors. My initial first supervisor, Professor Ianthi Maria Tsimpli, introduced me to Psycholinguistics in my 4th year of undergraduate study but also during my postgraduate studies at the Aristotle University of Thessaloniki in Greece. Professor Theodoros Marinis supported me throughout my PhD initially as my second supervisor and subsequently as my first supervisor. Dr Ian Cunnings originally supported me as a Postdoctoral Fellow at the University of Reading, subsequently became my second supervisor as a lecturer and in the last year my first supervisor. I thank all three for their inspiration, encouragement and input regarding theoretical, methodological and logistic issues and, of course, their endless patience to provide invaluably meticulous feedback on my thesis, conference abstracts and manuscript on more than one occasion.

I am also grateful to the children who participated in my research as well as their parents who agreed for them to take part. I would also like to thank the teachers and headteachers at the schools where I was able to recruit participants for championing my research and making data collection possible. I would also like to thank the University of Reading staff and their acquaintances whose children volunteered to take part in my research. Finally, I would like to extend my gratitude to the adult participants in my study.

I would also like to thank Dr Lotte Meteyard for her introducing me to mixed effects modelling and R as well as Dr Arpita Bose for enabling me to present my work at the Acquired Brain and Communication Disorders Lab and for her help on numerous occasions. I would also like to acknowledge numerous people with whom I either shared an office or have been in the same corridor over the years; Abhijeet Patra, Alessandra Valentini, Lewis Williams, Julia Hofweber, Daniel Fryer, Raluca Briazu, Adrian White, Lynn Bell, Georgina Dodd, Emma Pilkington, Jo Billington as well as my friend Dr Markos Kyritsis. From the more mundane tips on how to get things done on a day-to-day basis or the quick favours of reading through an abstract to the substantial conversations about how we can develop as researchers and progress professionally, having you around was a timely reminder that what I perceived to be as a personal failure was, in fact, just part of a normal learning curve called "life as a PhD student". Cheers!

I must also thank my parents for their support throughout my journey, not least their financial support in the first few months.

During my studies, I worked as a Teaching Fellow as a way to fund my PhD, initially in the Department of English Language and Applied Linguistics (DELAL) and subsequently in the School of Psychology and Clinical Language Sciences (SPLCS). As a self-funded student, this enabled me to undertake my PhD. Therefore, it is only fair that I thank my line managers, Professor Jane Setter in DELAL, and Professor Vesna Stojanovik and Professor Theodoros Marinis in SPCLS who employed me and guided me throughout these years as a very junior member of staff. The jobs were not merely a helping hand towards my studies, but also offered me with invaluable experience and opportunity well before my time. I am naturally also grateful to all members of staff at both DELAL and SPCLS who made me feel welcome and offered guidance, in particular my first mentor in DELAL, Professor Clare Furneaux.

Parts of this PhD have been presented or have been accepted to be presented at the following conferences:

Oral presentations:

Pontikas, G., Cunnings, I. & Marinis, T. (2019). Wh-question processing in bilingual children: Evidence from the visual-world paradigm. Talk given at 44th Annual Boston University Conference on Language Development (Boston, MA, 10th November 2019)

Pontikas, G., Cunnings, I. & Marinis, T. (2019). Bilingual children process wh-questions similarly to monolingual children: evidence from a visual-world paradigm study. Talk given at 2nd Conference on Multilingualism (Leiden, Netherlands, 2nd September 2019)

Pontikas, G., Cunnings, I. & Marinis, T. (2019). Incrementality of which-question processing in bilingual children. Talk given at Child Language Symposium 2019 (Sheffield, 11th July 2019)

Pontikas, G., Cunnings, I. & Marinis, T. (2017). Processing of which-questions in bilingual children: Looking while listening. Talk given at EuroSLA Doctoral student workshop (Reading, UK, 30th August – 2nd September 2017)

Pontikas, G., Cunnings, I. & Marinis, T. (2015). Bilingualism and SLI: evidence from eye-tracking. Talk given at Language Studies Working Conference, Department of English Language and Applied Linguistics, University of Reading (16th March 2015)

Pontikas, G. (2014). Bilingualism and SLI: evidence from language comprehension. Talk given at the Child Development Group, School of Psychology and Clinical Language Sciences, University of Reading (May 2014)

Pontikas, G., Marinis, T. & Tsimpli, I.M. (2014). Bilingualism and SLI: evidence from language comprehension. Talk given at the Language Studies Working Conference, Department of English Language and Applied Linguistics, University of Reading (14th March 2014

Posters:

Pontikas, G., Cunnings, I. & Marinis, T. (2018). On-line processing of which-questions by English-speaking monolingual and bilingual children: a visual word paradigm study. Poster presented at the Child Language Symposium, Reading, UK. (25th-26th June 2018)

Pontikas, G., Cunnings, I. & Marinis, T. (2018). Bilingual children process which-questions in the same way as monolingual children: a visual word study. Poster presented at the 2nd International Symposium on Bilingual and L2 Processing in Adults and Children (ISBPAC), University of Braunschweig, Germany. (24-25th May 2018)

Pontikas, G., Cunnings, I. & Marinis, T. (2018). Bilingual children process which-questions in the same way as monolingual children: a visual word study. Poster presented at Language Abilities in Children with Autism (LACA) Workshop. University of Reading, UK. (Reading, 23rd June 2018)

Contents

Table of figures

Chapter 1 Introduction

1.1 Introduction

The last decade has seen a significant increase in published works on morphosyntactic processing in bilingual children focusing on potential qualitative differences between monolingual and bilingual children (e.g. Blom & Vasić, 2011; Chondrogianni & Marinis, 2012; Chondrogianni, Vasić, Marinis & Blom, 2015; Lemmerth & Hopp, 2018; Marinis & Saddy, 2013; Lew-Williams, 2017; Vasić & Blom, 2011). The results have so far varied but a recurring finding is one of qualitatively similar patterns overall but with sometimes nuanced differences. Bilingual children tend to be slower than monolinguals even when they are equally accurate in a task (e.g. Chondrogianni et al. 2015; Vasić & Blom, 2011). Alternatively, effects observed in monolingual children may also be present in bilinguals but under more specific and restrictive conditions in terms of linguistic properties examined (e.g. Blom & Vasić, 2011; Lemmerth & Hopp, 2019) A further difference in the effects found in the bilinguals is that they may be contingent on the type of bilingualism based on the age of acquisition of the additional language (Lemmerth & Hopp, 2019; Roesch & Chondrogianni, 2016). The main distinction between types of bilingualism in children in the literature are simultaneous bilingualism (exposure to two languages from birth), early sequential (the additional language is acquired after the onset of the mother tongue, usually before the age of 3 or 4 years) and late sequential bilingualism (the additional language is acquired after the threshold for early sequential bilingualism; see also De Houwer, 2009; Genesee & Nicoladis, 2009).

There are at least two ways in which our understanding of processing of morphosyntax in bilingual children can be expanded as these areas remain largely unexplored; the primary one is in terms of the linguistic features examined and a second contribution is the methodology used. Most studies in bilingual morphosyntactic processing have focused on grammatical violations in the form of either errors of omission or errors in agreement of case and gender between determiners, adjectives and/or nouns (e.g. Chondrogianni & Marinis, 2012; Chondrogianni et al. 2015a, 2015b; Vasić & Blom, 2011; Vasić et al., 2012). There are still very few studies which examine either complex morphosyntax or key issues in sentence processing that have previously been examined in monolingual adults and children. This includes issues related to the timecourse of processing such as incrementality and revision (for the latter see, Lemmerth & Hopp, 2018; Lew-Williams, 2017). This thesis aims to bridge these gaps by examining how bilingual children process wh-questions and garden-path sentences, thus expanding the research evidence of how bilingual children process sentences in real-time by adopting visual world paradigms. These are two linguistic phenomena that involve a degree of temporary ambiguity. The parser is likely to initially misinterpret them and subsequently reanalyse them. In this sense, they are two suitable linguistic features to test incrementality in processing in bilingual children as well as their ability to revise erroneous interpretations. The manipulations in experimental design in this thesis also allow for us to address the last gap in the literature about how bilingual children can utilise information from various sources.

The second way research into child bilingual processing of morphosyntax can be enriched is related to research methodology. Most previous studies (e.g. Chondrogianni & Marinis, 2012; Chondrogianni et al. 2015a, 2015b; Vasić & Blom, 2011; Vasić et al., 2012) have used violation paradigms with a combination of off-line accuracy in the form of picture selection or comprehension question and experimental paradigms tapping into real-time processing using reaction times. The former captures the end result of comprehension but provides limited insights as to the nature of real-time processing on its own (Marinis, 2003; 2010, 2013, 2015, 2018, Marinis & Cunnings, 2018). Moreover, they measure participant behaviour in terms of response but do not capture the cognitive processes to reach this end state. As they only capture the end result and are not timed, they may also be the result of conscious reflection. Therefore, they are not the best insight into language processing (Marinis, 2003; 2010, 2013, 2015, 2018, Marinis & Cunnings, 2018).

Those studies which have reported reaction times used self-paced listening (e.g. Chondrogianni et al. 2015; Marinis & Saddy, 2013; Vasić & Blom, 2011) or word-monitoring tasks (Chondrogianni & Marinis, 2012). This measure is valuable in two ways; it can be used to infer sensitivity to morphosyntactic violations in a sentence and locates the segment in a sentence which is hard to process (Marinis, 2003; 2010, 2013, 2015, 2018, Marinis & Cunnings, 2018. The self-paced listening paradigm has also been used for garden-path sentences and sentences with temporary ambiguity where some degree of predictive processing would be needed in monolingual children (Felser, Marinis & Clahsen, 2003). An advantage of this type of study is that the information obtained about real-time processing can be informative about the structural representations or interpretations formed while hearing a sentence. One of the criticisms of self-paced listening and word-monitoring tasks is the potential lack of ecological validity. In self-paced listening tasks, the stimuli are presented word-by-word or phrase-by-phrase; this is unnatural as we don't typically listen to a sentence in this way but in a more continuous stream. Word-monitoring tasks, on the other hand, are also unnatural as participants do not need monitor an individual word and press a button during natural processing.

Therefore, a further contribution of this thesis is it that builds on the existing literature using a relatively novel experimental method in research with bilingual children. While the studies in this thesis report end-stage accuracy, they move away from the violation paradigm to examine sentence processing in bilingual children using an eye-tracking paradigm and, in particular, the visual-world paradigm. Eye-tracking with listening provides an improvement in comparison to previous methods used because sentences are presented in a natural way in the way we listen to sentences in the real world. This method has revealed differences in processing in children relative to adults which may not be captured in offline measures (e.g. Jackson, Felser & Clahsen, 2011). Given that looking while listening is a comparatively natural process and that eye-tracking provides a relatively high degree of temporal resolution, it is an ideal experimental method to investigate the timecourse of sentence processing in bilingual children and to address issues such as incrementality, revision or real time utilisation of information to aid processing. However, there are very few published works with bilingual children which use eye-tracking, and these do not test L2 English; both Lew-Williams (2017) and Lemmerth & Hopp (2018) tested gender cues on an adjective as a predictor for the upcoming noun in L2 Spanish and L2 German respectively. Therefore, a further contribution of this thesis is that it adopts an existing methodology to overcome limitations in the existing literature to provide a more fine-grained and naturalistic insight into sentence processing in bilingual children and is perhaps, the first of its kind in children's L2 English.

While research in morphosyntactic processing in bilingual children has been limited, this has not been the case for adults. In adults, debate has been strongly influenced by the Shallow Structure hypothesis (Clahsen & Felser, 2006a, 2006b), a claim that non-native speakers who acquired the second language late and with non-naturalistic exposure process language in a less purely grammatical way. Instead, as the syntactic representations built are assumed to not be fully specified under this account, their processing is argued to rely more on cues of a different nature, such as phonological, lexical or pragmatic cues. Numerous works have so far suggested that this is not the case (e.g. Hopp, 2015; for evidence from processing of syntactic ambiguity, also see Cunnings, 2017 for a more general review). Alternative accounts have proposed similarities in the nature of processing grammar in natives and non-native L2 learners but with more nuanced differences; it may be slower and less efficient, or it may be more resource intensive and thus, differences may emerge under more challenging conditions (Hopp, 2006; 2010; MacDonald, 2006, see also Cunnings, 2017 for a broader review on the issue).

Crucially, the debate in adult morphosyntactic processing has been about native and non-native speakers, i.e. L2 learners. However, it remains an open question as to what the nature of morphosyntactic processing is in individuals who have acquired two languages earlier in life and who, as adults, have become native speakers in both of them. A difference between late and early bilinguals is well established in the literature with the prevailing assumption being that the latter show nativelike performance and the former appearing deficient in this respect (Birdsong, 1999; Johnson & Newport, 1989). The reason for this has been argued to be the fact that, in the case of late learners, acquisition of the additional language has not taken place within a limited time window of biological changes, a critical period, after which acquisition is assumed to be still possible but will not be nativelike. Differences between early bilinguals with naturalistic exposure and monolinguals are also not predicted under accounts such as the Shallow Structure hypothesis (Clahsen & Felser, 2006a, 2006b, 2006c) as the quality, quantity and timing of the input both groups receive are similar. Accounts such as this which attribute these differences to the quality, quantity and timing of the input both groups receive being different are conceptually related to the above assumption. Evidence against this postulation comes from Hopp (2010) which supports the notion that late learners of a language can reach nativelike proficiency in terms of accuracy but that the L2 system is less efficient than in native speakers and more susceptible to L1 transfer.

Very limited work involves language processing in early bilingualism in adulthood, i.e. adults who were exposed to the additional language either at birth or at a young age (usually by about the age of 3 or 4 years) and who have are assumed to have attained nativelike competence relative either to monolinguals or late bilinguals / L2 learners. These adults are referred to as "(native) bilinguals" throughout this thesis in contrast to monolingual native speakers who are referred to as "monolinguals" and those who acquired an additional language at an older age mostly with non-naturalistic exposure who are referred to as "L2 learners". It is reasonable to expect differences between native and non-native speakers of any given language in terms of processing due to the different conditions under which they acquired the language; in fact, even the output may be different even between native speakers and highly proficient non-natives. However, it is less clear what will be the case with early bilinguals. Under the critical period hypothesis, early bilinguals will reach native proficiency and will be practically indistinguishable from monolinguals.

On the other hand, bilinguals will inevitably differ from monolinguals, regardless of their proficiency in the additional language, as they have two linguistic systems. Research from lexical processing suggests bilinguals may be slower than monolinguals (de Bruin, Della Sala & Bak, 2016) and this has been attributed to the possibility that both languages remain active during processing and may interfere when bilinguals or one of their languages (e.g. Blumenfeld & Marian, 2007; Green, 1986, 1998; Grosjean, 1997, 1998; Li, 1996; Marian & Spivey, 2003). This has been showed predominantly for lexical production (Costa, Caramazza & Sebastian-Galles, 2000; Colomé, 2001). It is conceivable that one language may influence the other during processing at the level of syntax as well. Albeit from offline measures, available evidence, mainly from syntactic priming, points in this direction (Dijkstra Grainger & van Heuven, 1999; Bernold, Hartsuiker & Pickering 2007; Desmet & Declerq, 2006; Loebell & Bock, 2003; Hartsuiker, Pickering & Veltkamp, 2004; Meijer & Fox Tree (2003) Schoonbaert, Hartsuiker &Pickering, 2007). In sum, contrary to what one might intuitively expect about early sequential or simultaneous bilinguals who reach nativelike proficiency in both languages, "[t]he bilingual is not two monolinguals in one person" (Grosjean, 1989: 203)[1]. The last contribution of this study is that it examines language processing in bilingual adults by investigating differences between early bilinguals with naturalistic exposure to both monolingual adults and bilinguals with late and predominantly classroom-based exposure to English. To this end, the same morphosyntactic phenomena are examined as with the children, wh-questions and garden-path sentences, again using the visual-world paradigm.

The remainder of this chapter presents the linguistic phenomena that are studied in this thesis in more detail and subsequently reviews the empirical evidence on bilingual adult and child language morphosyntactic processing, as well the most influential theoretical accounts, which have driven or are informed by the empirical research. The chapter concludes with a statement of the research questions and an overview of the structure of this thesis.

1.2. Linguistic phenomena examined in this Thesis

This thesis explores the processing of two linguistic phenomena which have been extensively studied in bilingual adults within the native vs. non-native debate but remains largely unexplored in bilingual children. The two phenomena are wh-questions and garden-path sentences. This section introduces these two phenomena from a linguistic perspective and outlines some theoretical assumptions about the reasons for which they are expected to be challenging to process. As this is relevant to both the children's and adults' studies, it is introduced at the onset of this thesis. Research more specific to either children or adults is included in the respective chapters.

---

[1] Further evidence of differences between monolinguals and early bilinguals has been found at the neural level for semantic processing (Proverbio, Čok & Zani, 2002; see also Werker & Byers-Heinlein (2008) for a review of differences in children)

1.2.1. Wh-questions

Wh-questions and, more generally, filler-gap dependences have been widely investigated in the literature on first language acquisition and first language processing in adults (e.g. Avrutin, 2000; Deevy & Leonard, 2004; Friedmann, Belletti & Rizzi, 2009; Goodluck, 2005; Jakubowicz & Gutierez, 2007; Kaan, Harris, Gibson & Holcomb, 2000; Stavrakaki, 2006). Long-distance dependences, of which wh-questions are a subtype, are syntactic constructions where a link is formed between overt and null elements in a sentence which can stretch over a number of words and syntactic constituents. The word order is not the canonical one with the overt element position at the front of the sentence with the default position empty. Subject wh-questions are constructions where the wh-phrase element in the non-canonical position is the subject of the verb – as such, the questions have the canonical word order of declarative sentences, as in (1a). Whereas object wh-questions are those with the verb's object as the fronted constituent, as in (2a), and this does not correspond to the canonical word order in a declarative.

(1a) Which donkey is carrying the zebra?

(2a) Which donkey is the zebra carrying?

These examples have been taken from English; it is not the case the all languages have interrogatives where question words obligatorily appear sentence-initially (Dryer, 2013a). According to the generative framework the fronted constituent's non-canonical position is due to syntactic movement. Within the generative framework, a dependency is formed between the moved element, known in the literature as the filler, and the original position, known as the gap which is now phonetically null (Chomsky, 1995). This is depicted in (2a) and (2b) for subject and object questions respectively:

(2a) Which donkey$_i$ e$_i$ is carrying the zebra?

(2b) Which donkey$_i$ is the zebra carrying e$_i$?

In generative syntax which assumes movement, this type of movement is known as A'-movement (Ouhalla, 1994; Radford, 2004) which is different to A-movement. The latter is assumed to be syntactic movement where a constituent changes syntactic function but maintains its argument structure. Examples of A-movement are passives and unaccusative verbs. On the other hand, A'-movement is assumed to be syntactic movement where the constituent is displaced but does not change syntactic function. In the case of wh-questions, the wh-phrase is fronted but remains the subject, object or adjunct to the verb. It should be noted that not all generative accounts, postulate movement for the subject wh-questions but do for object questions.

Other linguistic accounts, however, do not postulate an empty category per se but assume that the filler is connected to the lexical subcategoriser. In other words, the fillers are connected to the verb which requires a specific thematic position to be filled and this, in turn, will be syntactically projected as the verb's object (Pickering & Barry, 1991; Pollard & Sag, 1994; Sag & Fodor, 1995). These accounts typically also don't assume movement.

For English and many other languages, object-extracted wh-questions have been found to show greater difficulty than subject-extracted questions; this has been termed the "subject-object" asymmetry (Avrutin, 2000; Deevy & Leonard, 2004; De Vincenzi, Arduino, Ciccarelli & Job, 1999; Goodluck, 2005; Stavrakaki, 2006; Stromswold, 1995). Numerous theories have attempted to account for this discrepancy between the two structures either for wh-questions or for other long-distance dependencies in general. One strand of accounts is centred on working memory as a cornerstone of processing. A main tenet is the equation of increased difficulty with more working memory resources required to maintain a representation in working memory during processing (Just & Carpenter, 1992). The difficulty with object wh-questions has been argued to be due to the increased linear distance between the filler and the gap (Grodner & Gibson, 2005) or the structural distance between the filler and the gap relative to that in subject wh-questions or other long-distance dependences (Gibson, 1998).

Alternative probabilistic accounts don't attribute the increased difficulty to inherent properties of the structures themselves or to individuals' memory abilities but to the input speakers of a language have been exposed to. For example, some proposals argue that the relative infrequency of object which-questions relative to subject questions, leads to the formation of the expectation that an upcoming wh-question should be interpreted as a subject-question (Hale, 2001, Levy, 2008; Roland, Dick & Elman., 2007). This is not a problem if the question in the upcoming wh-question is indeed subject-extracted. But this violation of one's expectation is assumed to be the source of difficulty with object-extracted wh-elements. From a probabilistic perspective, speakers of English will adopt the heuristic to initially interpret the first NP as the subject of the wh-question due to their experience (e.g. Hale, 2001, Levy, 2008; Roland, Dick & Elman., 2007); in other words, they will assume that the wh-question is a subject question because they are more likely to encounter one.

The crucial implication of accounts which involve processing is that they rely on the temporary ambiguity of all which-questions and the incrementality of language processing. Speakers of a language, when listening, are not assumed to wait until the completion of a sentence before commencing processing. Rather, processing starts as early as possible and relies on the available content listeners have already heard (MacWhinney, 2001). This may lead to initially erroneous interpretations of sentences when there is temporary ambiguity. Sentences such as (2) are temporarily ambiguous because due to the lack of case morphology the first noun phrase may be either interpreted as the subject or the object of the verb. It is only after the second NP that the ambiguity is resolved. However, speakers of English do not wait until the second NP and will begin to construct a syntactic representation and an interpretation of the sentence after the first NP and are more likely to interpret it initially as a subject which-question (Bever, 1970; MacWhinney & Bates, 1982 for children; Contemori, Carlson & Marinis, 2018 for data from adults and children). For object-questions, this interpretation will need to be revised after the second NP.

From a processing perspective (e.g. Frazier 1987, Frazier & Clifton, 1989), the comprehension of filler-gap dependences involves identifying a dislocated element, i.e. a syntactic constituent whose canonical position would be one other than where it finds itself in the sentence. This is assumed to trigger the anticipation of a syntactic gap; i.e. the empty syntactic constituent later on in the sentence, which is expected to be filled. Erroneously filling the gap at the earliest point possible is assumed to be the cause of misinterpretation and need for reanalysis. This is evidenced by the filled-gap effect (Crain & Fodor, 1985; Stowe, 1986) in self-paced reading tasks; this is a slowdown when the participant encounters phonetically overt content at an assumed gap position relative to the equivalent position where no gap is expected. For example, in (3) a gap is expected after the verb *bring* but not for the respective verb in (4);

(3) My brother wanted to know who Ruth will bring us home to at Christmas.

(4) My brother wanted to know if Ruth will bring us home to Mom at Christmas.

In (3), who is initially interpreted to be the object of the verb bring; the slowdown at the pronoun us is due to the fact that, given this interpretation, the parser was not expecting the gap position to be filled. In (4), there is no such expectation and as such the parser does not experience this reanalysis-induced slowdown. There are, however, alternative accounts for the filled-gap effect. Under Head-Driven Phrase Structure Grammar (Pollard & Sag, 1994;

Sag & Fodor, 1994), filled-gap effects are explained on the assumption that the filler and the empty constituent (known as the gap) share features which are propagated from the filler to the foot one syntactic tree node at a time. Under direct association (Pickering & Barry, 1991), no gaps are assumed to exist and the filler is connected to the verb via its subcategorization frame (i.e. the information about its syntactic and argument structure), therefore, the effects are anticipated to occur at the verb position and not at the gap position. In each case the difficulty resides due to incrementality and the tendency of the parser to interpret the filler to be the direct object of the verb when it has an overt argument.

Regardless of theoretical account, the tendency towards a subject-bias is fairly well attested for filler-gap dependences in English even in typical adult L1 speakers using a variety of methodologies (Kaan, Harris, Gibson & Holcomb, 2000 for healthy adults and ERPs; Avrutin, 2000, Hickok & Avrutin 1996 for production and offline comprehension in young children and adults with aphasia). This can be evidenced from both behavioural measures such as slower reaction times for object questions (Stowe, 1989, who showed greater difficulty for wh-questions relative to declaratives with an object NP but not with a subject NP; see also, Grodner & Gibson, 2005 for relative clauses) but also electrophysiological measures such as sustained anterior negativity after the filler and late posterior negativity at the gap (Phillips, Kazanina & Abada, 2005). In another study which used eye-tracking, Staub found longer reading times for object relative clauses than for subject relatives and more regressive saccades for relative clauses than for subject clauses, indicative of greater processing difficulty and re-analysis of syntactic structure respectively (2010). Similar results have been obtained for other languages; for example, Schlesewsky, Fanselow, Kliegl & Krems (2000) have shown a strong preference for interpreting neuter singular case-syncretic wh-questions in German which are locally ambiguous as subject-questions using both on-line and off-line measures. Likewise, for Dutch, a strong subject advantage was observed for both wh-questions and declaratives in a grammaticality judgement task (Frazier & Flores d'Arcais, 1989).

The difficulty with object relative to subject gaps has been discussed so far in terms of incrementality. Object questions are, in sum, assumed to be harder because they are initially ambiguous and the parser tends to misinterpret them, thus they require reanalysis. A second parameter to the additional difficulty is related to retrieval difficulties due to the non-canonical order of the constituents; this entails the need for the preservation in working memory and their retrieval at the canonical position, as the distance between the two crosses other phrases which may have an interfering function.

Within the generative framework, difficulty in processing object extracted wh-questions could further be attributed to locality effects predicted under Relativized Minimality (Rizzi, 1990, 2004, for empirical evidence in children, see Friedmann et al. 2009, for an alternative view, see Goodluck, 2010). Under this account, the syntactic dependency established through A'- movement, in this case, between the fronted wh-phrase and its original position would be interrupted by the presence of an additional constituent, an intervener, in this case, the subject of the sentence. Crucially, under Relativized Minimality, the A' dependency is interrupted because the intervener and the initial constituent share morphosyntactic features (such as number) and, as a consequence, can act as a potential candidate as the filler in the syntactic relation. This entails a significant prediction which has been supported from previous studies (e.g. Contemori et al. 2018; Friedmann et al. 2009); that the intervening element will have this effect only in sentences where the subject and object share the same number as in (5a) and (5b) but not where they are different as in (5c) and (5d) where the mismatch in number will aid comprehension.

(5a) Which donkey is the zebra carrying?

(5b) Which donkeys are the zebras carrying?

(5c) Which donkey are the zebras carrying?

(5d) Which donkeys is the zebra carrying?

Other memory-based accounts have focused not solely on maintenance of content but on retrieval and reactivation of constituents with the increased difficulty of object-extracted long-distance dependences being due to cues having reduced diagnostic value. One such alternative processing account which could account for this asymmetry is similarity-based interference accounts (Gordon et al. 2001, 2002, 2004). According to this account, the challenge in sentence processing lies in susceptibility of memory representations to items which have similar features as the information to be encoded and remain active during sentence processing. The parser will attempt to retrieve the relevant constituent stored in working memory but others will also remain active during processing and will impede the selection of the correct one. This is particular the case when the target to be retrieved shares similar features with its competitors. Such features could be grammatical gender or number or other semantic features such as animacy. Crucially, when some of these features are different between the target element (e.g. NP at the verb's object position) to be retrieved and its competitors, retrieval and processing is facilitated. For example it is predicted that (5a) will be more difficult to process than (5c); the reason for this is that as the match in number of the two NPs means that neither can be excluded as the subject of the sentence and hence will remain available at the point of the verb's object. This means, they are unable at the retrieval point to uniquely select the filler amongst potential distractors as they share similar features such as number and/or gender and as such more than one could be a potential match (Gordon et al., 2001, 2004).

Relativized Minimality is similar to similarity based-interference proposed by Gordon et al. (2001, 2002, 2004), as the source of the difficulty is predicted to be the similarity in features. However, it is an account which originates in generative grammar as opposed to similarity-based interference accounts which are framed within cognitive science.

1.2.2. Garden-path sentences

Building a syntactic structure for incoming material and a representation of a sentence is assumed to happen instantaneously. With unambiguous sentences, this is comparatively straightforward, with ambiguous sentences less so. Ambiguity means that there are two options available for the parser to build so the parser will only select one of the two. There are two types of structural ambiguity; global and local. Globally ambiguous sentences are not resolved by the end of the sentence as in (6);

(6) Visiting relatives can be a nuisance

It is unclear in (6) whether the noun "relatives" is the object of the verb visiting or the subject of the verb *be*. The sentence does not provide any clues as to which reading is the correct one.

In sentences like (6), the interpretation the parser selects is one of the two interpretations. As the sentence is not disambiguated, the interpretation of the sentence is unlikely to be revised. But not all ambiguous sentences are like (6). Locally structurally ambiguous sentences are ambiguous only up until a point at which the content disambiguates between the two possible interpretations. In this case, only one interpretation is ultimately viable.

Garden-path sentences are a particular type of sentence with local structural ambiguity where the parser is misguided towards an erroneous initial interpretation of the sentence. An example of a garden-path sentence is (7) or (8)

(7) He put the frog on the napkin on the table.

(8) While Mary dressed the baby slept.

Until a point (7) and (8) have two potential interpretations yet after the disambiguating material one of the two becomes unavailable.

The difficulty with (7) is that the parser will build a syntactic structure for the sentence but at the point of the second PP, the sentence becomes unambiguous and one of the two interpretations becomes unavailable. In the case of "*He put the frog on the napkin on the table*", speakers may consider the first PP to be a modifier of the verb and interpret it as the destination (e.g. Trueswell et al. 1999). However, after the second PP, this becomes unviable and the first PP can only be interpreted as a modifier of the noun frog[2]. If the parser has already built a syntactic representation inconsistent with the ultimately correct reading of the sentence, this needs to be revised and a new one built after the parser has become aware of the need to do so. Similarly in (8), the parser is guided towards interpreting the NP "the baby" as the object of the verb "dressed"; it is only after the second verb "slept" that one realises that the NP has been misinterpreted and that it is the subject of the second verb (e.g. Trueswell et al. 1999). There is considerable research into whether children can do this in the same way as adults (see Chapter 4 for a more detailed review) but whether or not bilingual children can do so remains an empirical question.

Garden-path sentences have traditionally been used to test models of language processing. Structure-based accounts (Frazier, 1987a) posit that initial decisions regarding syntactic

[2] Technically, in the aforementioned example, the first PP can still be interpreted as the destination with the second PP being interpreted as a modifier of the first PP (i.e. Put the frog on the napkin that is on the table). This possibility is not discussed in the original studies on garden-path effects (e.g. Trueswell et al., 1999) potentially because the visual stimuli would not support such an interpretation, as they wouldn't in the studies on garden-path sentences in this thesis.

structure formation are exclusively driven by the reader’s knowledge of the syntactic structures and possible constituents. The semantics of the verb or the distributional properties are not initially relevant in this type of account. Two parsing strategies have been postulated to guide the parser; minimal attachment and late closure. Minimal attachment (Frazier, 1987a, 1987b; Frazier & Rayner, 1982) is a tendency for the parser to attach incoming material parsimoniously in order to create the simplest syntactic structure possible. Therefore, in a sentence such as (9)

(9) The judge believed the defendant …. and threw out the charges / was lying”

(example from Osterhout, Hollcomb & Swinney, 1994:786), although both the option of a direct object to the verb and an embedded clause are possible, the human parser will most likely adopt the first interpretation because a sentence with an embedded clause would be structurally more complex and therefore, avoided until it was clear this was the appropriate syntactic representation of the sentence. It is under minimal attachment that the PP “*on the table*” in (7) will be interpreted as the adjunct to the verb *put* rather than a complement to the noun *frog*. For (8), the parser will attach the NP “the baby” as the object of “dressed” in order not to create an embedded clause and therefore, keep the sentence as simple as possible. The second parsing strategy is late closure (Frazier & Fodor, 1978). Under late closure, incoming material is assumed to be attached to the local clause that is processed and not to previous clauses. This entails that within embedded clauses new phrases are attached to the embedded clause and not the main clause.

Alternatively, lexically-driven accounts posit that syntactic interpretation is built on the semantic knowledge of the complement taking verbs (McDonald, Pearlmutter & Seidenberg, 1994; Lapata, Keller & Schulte im Walde, 2001). Lexically-driven accounts of processing are strongly related to expectation-based probabilistic accounts (Levy, 2008; Roland, Dick &

Elman, 2007) under which the probabilistic knowledge of the distributional properties of different linguistic features impacts their processing. Therefore, in (8), the second NP will be interpreted as the object of the verb because "*dress*" is more frequent as a transitive rather than an intransitive verb. As such, the parser is lead to expect "dress" to be a transitive rather than transitive verb. For (9), it is the semantics of the verb "*believe*" and the nouns "*judge*" and "*defendant*" that make a transitive interpretation more likely, in other words, the meaning of the words makes this reading more plausible. In the aforementioned examples, different accounts don't make conflicting predictions. This is not always the case.

The garden-path model is an account within a broader theoretical framework where processing is rigidly serial or stepwise. In this type of account, other factors such as word meaning or pragmatics are not assumed to play a role in initial parsing decisions (Ferreira, 2003; Ferreira & Henderson, 1991; Fodor, Bever & Garrett, 1974; Fodor & Ferreira, 1998, Fodor & Inoue, 1994; Frazier, 1987; Frazier & Rayner, 1990). Processing is assumed to be initially purely structural and independent of other sources of information. It is only at the later stages where non-syntactic cues becoming relevant. In this sense, processing under the garden-path model is assumed to be highly modular. Lexically-driven and probabilistic models belong to the opposite strand of theoretical accounts known as constraint-based models (MacDonald, Pearlmutter & Seidenberg, 1994; Seidenberg & McClelland, 1989). In these models multiple sources of information such as semantics or pragmatics are active from the onset of processing and assist in determining the interpretation. Processing is not assumed to be serial but also cues are utilised in parallel. As such processing is not posited as something modular but interactive. Evidence so far suggests that the parser in adults makes use of non-syntactic information to facilitate processing; such information may be related to lexical properties of the words in the sentences (e.g. Ferreira & Clifton, 1986; Trueswell, Tanenhaus & Kello, 1993 for the effects of animacy), prosody (Nakamura, Arai & Mazuka, 2012; Papangeli & Marinis, 2010), discourse (Ferreira & Clifton, 1986). Another source of information which may be utilised is non-linguistic; the presence of two referents in a visual scene has been shown to facilitate processing by reducing or helping to avoid effects of misinterpretation (Tanenhaus, Spivey-Knowlton, Eberhard & Sedivy, 1995; Trueswell, Hill, Logrip & Sekerina, 1999; Spivey, Tanenhaus, Eberhard & Sedivy, 2002). The motivation for this is that the number of referents has a pragmatic function as the presence of the second referent for sentences like (7) triggers the need to provide a motivation for a modifier to a phrase in order to disambiguate, thus avoiding garden-path effects.

An alternative account which does not fall within either of these two strands is the theory of good enough processing (e.g. Ferreira, Bailey & Ferraro, 2002; Ferreira, Engelhard & Jones, 2009, Karimi & Ferreira, 2016). Under this account, processing is claimed to not always be structural but more heuristic in sentences which are difficult. In these cases, multiple different types of cue are assumed to be used as in constraint-based models. Language processing is posited as potentially only partial and the arising representations incomplete (Ferreira et al., 2009) or lacking in detail and that the parser may remain with the erroneous interpretation. Evidence for this can be drawn from the fact that people may respond positively to the question "Did Mary dress the baby?" when hearing a sentences as in (8), "While Mary dressed the baby slept" (Christianson et al. 2001). This is the wrong interpretation as the verb "dress" in this case is intransitive rather than transitive; the second NP in the sentences is, thus, in the absence of overt morphological case, interpreted as the object of the verb. The initial interpretation needs to be abandoned at the second verb as the ambiguous NP will be interpreted as a subject. The fact that participants respond to the comprehension question in a way that is consistent with a transitive reading of the verb "dress", entails that the original misinterpretation persists. Under the good enough account, the representation built is posited to also be informed by schematic knowledge the speaker of a language may have.

It should be noted that more recent accounts of good-enough processing do not ignore syntactic structure; reanalysis is assumed to take place. In fact, in more recent accounts, it is considered largely complete; the main tenet of good-enough processing accounts regarding the global interpretation of ambiguous sentences is that the initially assigned interpretation is not fully erased but that both interpretations seem to partially co-exist (for evidence for this see Slattery, Sturt, Christianson, Yoshida & Ferreira, 2013).

In sum, these two linguistic features tested are suitable to examine real time processing of sentences in bilingual children and adults, and to test key features observed in monolingual children but also adults such as incrementality, revision and reanalysis, and use of information from other sources in real time. The remainder of the chapter provides an overview of the main theoretical issues in sentence processing in bilingual children and bilingual adults as well as empirical evidence. It concludes with an outlining of the central research questions and an overview of the structure of this thesis.

1.3. Language processing in bilingual children

1.1.1. Theoretical background – What has motivated studies so far?

Research into language processing in bilingual children has largely been motivated by questions regarding the nature of their syntactic representations, the rationale being that different processing patterns would reflect potentially different representations (e.g. Chondrogianni & Marinis, 2012). A number of theoretical accounts originate from adult L2 acquisition but have been expanded to children. One of the longest standing theoretical questions in studies on L2 acquisition in children concerns the nature of syntactic representations; given the fact that production has been shown to not always be targetlike (e.g. see Chondrogianni & Marinis, 2012 for a review for on tense morphology in L2 children), it is plausible that the representations of functional categories are also deficient (e.g. Haznedar, 2001; Minimal Trees Hypothesis, Vainikka & Young-Scholten, 1993, 1996a, 1996b, 2010). Accounts which assume some degree of deficit normally posit this for linguistic features not present in the L1 and hypothesize the need for these to be acquired within an early time window (e.g. Failed Functional Features Hypothesis, Hawkins & Chan, 1997, Hawkins & Hattori, 2006; Interpretability Hypothesis, Tsimpli, 2003; Tsimpli & Dimitrakopoulou, 2007; Tsimpli & Mastropavlou, 2008). Alternative accounts have assumed no deficit in the representation of linguistic categories as such (Epstein, Flynn & Mortohardjono, 1996) but a difficulty in their lexical realisation in production (e.g. Feature Reassembly Hypothesis, Lardiere, 2009; Missing Surface Inflection Hypothesis, Prévost & White, 2000). A further issue which is relevant to this thesis is whether or not bilingual children utilise the same sources of information and the same processing strategies for the L2 or additional language as monolingual children do for their L1[3]. The studies in this thesis do not directly test these theoretical accounts about the nature of representation or access to morphosyntactic features although the results are discussed in terms of their consistency with the predictions made by these. Nonetheless, this thesis examines processing strategies in monolingual and bilingual children and adults and also how they integrate sources of information.

1.1.2. Previous works on ungrammaticality

The earlier works on morphosyntactic processing in bilinguals have focused on inflectional morphology or agreement at the phrase level and have used a violation paradigm through self-paced listening or word monitoring and/or sentence-to-picture matching (e.g.

[3] There are at least two more questions which have attracted substantial research, and which could impact on language processing in bilingual children; whether or not there is transfer of representation – and by extension, processing strategies, from the L1 and whether the developmental trajectories differ between L1 and L2 acquisition. This thesis does not test them explicitly and therefore, they are not included in the literature review.

Chondrogianni & Marinis, 2012; Marinis & Saddy, 2013; Vasić, Chondrogianni, Marinis & Blom, 2012). This type of study was motivated by the debate about syntactic representations in bilingual children. If they lacked a fully-fledged representation of inflectional features, it was assumed that the bilingual children would not be equally able as the monolingual children to detect ungrammaticality. Sensitivity to ungrammaticality is assumed to be reflected in an increase in reaction times at the segment which contains the grammatical error. While the linguistic features in this thesis are not the same as in these studies, they do examine real time processing – similarly to the studies in this thesis - in addition to end stage proficiency as their experimental method (self-paced listening), this allows one to pinpoint the segment in the sentence which causes difficulty during processing. The studies in this thesis extend the findings from these studies by examining two so far unexplored linguistic phenomena in bilingual children in terms of real time processing and provide an insight into the interpretations formed as they hear them.

Numerous studies have used this experimental paradigm to test both verbal (Chondrogianni & Marinis, 2012 for tense morphology in English) and nominal inflectional morphology. Studies in verb morphology have included tense and non-tense morphology in English (Chondrogianni & Marinis, 2012), subject-verb agreement mismatch in Greek[4] (Kaltsa, Tsimpli, Marinis & Stavrou, 2016), passives (Marinis, 2007; Marinis & Saddy, 2013). Studies on nominal inflectional morphology have examined gender agreement between determiner, adjective and noun in Dutch (Vasić & Blom, 2011; however, this study does not provide a group comparison), gender agreement in Dutch and Greek in separate studies (Vasić, Chondrogianni, Marinis & Blom, 2012), subject or object definite articles and object

[4] Greek permits subject-verb agreement mismatch in number if the subject is post-verbal but not pre-verbal.

clitics[5] in Greek (Chondrogianni, Marinis, Edwards & Blom, 2015a), indefinite articles in English (Chondrogianni & Marinis, 2016), determiners in English (Chondrogianni et al., 2015b), pronouns and reflexives in English (Marinis, 2008).

The majority of studies have tested Turkish-English or Turkish-Dutch bilingual children in the age range of 6-9 years, although some studies tested older children (e.g. Kaltsa et al. tested Albanian-Greek children aged 10-13 years) while Chondrogianni & Marinis (2016) tested younger children (5-8 years). Bilingual children had slower reaction times than the monolingual children overall across most studies and both groups were slower for ungrammatical segments relative to grammatical ones. However, the absence of an interaction between group and grammaticality suggests that the bilingual children slowed down to the same extent as the monolingual children in the ungrammatical sentences. This has been taken to reflect equal sensitivity to ungrammaticality and similar processing mechanisms in both monolingual and bilingual children. As such, the findings have been interpreted as being in support of accounts such as the Missing Features Inflection or the Feature Reassembly Hypothesis[6], which posit a non-deficient syntactic representation as the difficulties that had been previously observed in production were not found for comprehension. The results from comprehension accuracy or accuracy of grammaticality judgement in the aforementioned studies are not consistent with accounts which assume an absence of syntactic representation as, there were few differences between monolinguals and bilinguals (with the exception of the studies on the passive, Marinis, 2007; Marinis & Saddy, 2013). In these studies, the bilinguals had lower accuracy than the monolinguals and this was found disproportionately for the passive sentences relative to the active ones. However, even

[5] There are no subject clitics in Greek

[6] However, studies on nominal morphology are less supportive of the Missing Features Reassembly Hypothesis as errors of omission found in the elicited production task, which also included, are not predicted.

though these studies showed greater difficulty for the bilinguals with passives, they did not indicate qualitative differences in processing mechanisms as similar effects, in the same direction, were found at the same segments for both groups. Moreover, bilinguals' ability to utilise cues in real time appeared to be contingent on grammatical transparency, i.e. for linguistic features which are rule-based, and was not found when the cues were lexically idiosyncratic, i.e. applicable to specific lexical entries but not generalizable to others (Vasić et al., 2012).

1.1.3. Complex morphosyntax

This study builds on existing work on sentence processing in bilinguals. Existing work provides insights into real time processing for morphosyntactic phenomena at the phrase level but work on complex morphosyntax is still comparatively limited.

Long-distance syntactic relationships and, in particular, filler-gap dependencies, which wh-questions are a subtype of (see section 1.2.1 of this chapter) are one area of morphosyntax which remains underexplored in bilingual children in terms of processing. The first study to investigate comprehension of filler-gap dependencies is a study by Roesch & Chondrogianni (2016). This study tested comprehension of German subject and object which-questions as in (12a) and (12b) respectively in L1 French L2 German bilingual children aged 4-6 years.

(12a) subject question

Welche Maus malt den Frosch an ?

WhichFEM-Ø mouseMASC- Ø paints theNOM-ACC frogNOM-ACC?

'Which mouse is painting the frog?'

(12b) object question

Welche Maus malt der Frosch an ?

WhichFEM- Ø mouseFEM- Ø paints theMASC-NOM frogMASC-NOM?

'Which mouse is the frog painting?'

The study was based on the assumption that the subject-object asymmetry observed for English is also attested in German and on the observation that object which-questions are locally ambiguous in German as they are in English. If sentences are parsed incrementally, it is likely that object questions in German will initially be misinterpreted as subject questions as is the case in English. This study used an off-line picture selection task where participants heard sentences, in this case, which-questions, and needed to select a picture. The study compared simultaneous and early sequential bilinguals – with a limited length of exposure to German of about 1-3 years – to L1 German age-matched controls. The study in question investigated how monolingual and bilingual children could use disambiguating cues to help processing of which-questions in German. In particular, Roesch & Chondrogianni tested how the children used nominal case to disambiguate and manipulated the position of the cue. The latter was either sentence-initial, i.e. on the wh-word (the noun was masculine so that nominative case could be distinguished from the accusative) or late, i.e. it appeared only on the second NP. The third condition was the double cue where there was a disambiguating cue both on the wh-word and on the second NP

The manipulation was based on the gender of the nouns. Masculine nouns in German carry overt case marking contrast between nominative and accusative for the determiner and the adjective (and potentially the noun) which is not the case for feminine and neuter.

Roesch & Chondrogianni found lower accuracy for object which-questions relative to subject questions and an advantage that was more pronounced in the double cue condition. Overall the monolinguals had higher accuracy scores than both groups of bilinguals. The most beneficial cue for disambiguation was the double cue followed by the cue at the start of the sentence. The three groups benefited from the disambiguating cue differently. The monolinguals utilised the cue in all three conditions but the simultaneous bilinguals benefited only from the presence of a disambiguating cue at the start of the sentence. This benefit was absent from the sequential bilinguals. These results are consistent with Meisel (2009) who argued that early sequential bilinguals differ to simultaneous bilinguals and monolingual children. It is also one of the studies which contradicts earlier research previously reviewed as qualitative differences between bilingual and monolingual children are indeed observed. However the fact that the bilingual children made both errors relating to the reversal of thematic roles but also selected a distractor image suggests that their parsing strategies were neither random nor purely reliant on linear word order. Therefore, the difficulty the sequential bilinguals have is arguably not one of representational deficit but rather of using the appropriate cues in real time when these are not adequately prominent.

Garden-path sentences are another type of syntactically complex sentence. Papangeli & Marinis (2010) examined how L1 Russian L2 Greek bilingual children aged 9-12 years process sentences with temporary ambiguity in Greek in comparison to monolingual controls using a self-paced listening task. Papangeli & Marinis manipulated a verb that was either optionally transitive or always intransitive and the reading of the NP was either the subject of the verb or the object (based on agreement between the ambiguous NP and the verb following it). In the event of the optionally transitive verb, the subsequent NP is ambiguous; it may the subject of a forthcoming verb but it is most likely going to be interpreted as the object of the verb the parser has encountered. Examples of the experimental items are depicted in (13a-d);

(13a) Optional transitive verb; transitive interpretation of the verb

Eno maγireve ta psaria kaike sto furno

While cooking-PST-3SG the-NEU.PL.Ø fish-NEU.PL.Ø, burned-PST.3SG in.the-MASC.ACC oven-MASC.ACC

‘While he/she was cooking the fish, she/he was burned in/by the oven”

(13b) Optional infinitive – intransitive interpretation (garden-path)

Eno maγireve ta psaria kaikan sto furno

While cooking-PST-3SG the-NEU.PL.Ø, burned-PST.3PL in.the-MASC.ACC oven-MASC.ACC

‘While he/she was cooking, the fish were burned in/by the oven’

(13c) Intransitive verb – transitive interpretation (ungrammatical)

* Eno etrehe ta psarya kaike sto furno

While running-PST-3SG the-NEU.PL.Ø, burned-PST.3SG in.the-MASC.ACC oven-MASC.ACC

‘While he/she was running the fish.ACC, he/she was burned in/by the oven’

(13d) Intransitive verb - intransitive interpretation

Eno etrehe ta psarya kaikan sto furno

While running-PST-3SG the-NEU.PL.Ø, burned-PST.3PL in.the-MASC.ACC oven-MASC.ACC

‘While he/she was running, the fish.NOM were burned in/by the oven’

The ambiguous NP contained a neuter noun as these are ambiguous for nominative or accusative case in Greek. This way, it was unclear what the syntactic function of the NP was until the verb. In the case of the intransitive verbs, the object interpretation of the NP was ungrammatical but in the sentences with optionally transitive verbs both readings are grammatical. Participants heard the sentences in segments and then had to judge the sentences' grammaticality. Papangeli & Marinis conducted two experiments; one where the intonation of the sentence was flat providing no cues towards one of the two interpretations and another where the sentences were read with natural intonation where cues guide towards one reading over the other and garden-path effects are thus not expected. In the experimental without prosodic cues, the monolingual children showed a slowdown at the post-critical segment for the sentences with an optionally transitive verb and a subject reading of the NP relative to the other three conditions. This reflects garden-pathing; the L1 Greek children were garden pathed and interpreted the ambiguous NP as the object of the verb where this was permitted. The slowdown at the disambiguating segment suggests that they are sensitive to the agreement mismatch between the second NP and the second verb and that this signals that the original interpretation is wrong. This was not found for the bilingual children. On the other hand, the monolingual children did not show this slowdown in the experiment with the presence of prosodic cues suggesting that using available cues they were not garden-pathed. Instead, they slowed down generally for the ungrammatical sentences relative to the grammatical ones at the post-critical segment. Between-experiment comparisons revealed an interaction of group by prosody further supporting the notion that only the monolingual children utilised prosodic cues to avoid initial misinterpretation of locally ambiguous sentences. Conversely, there was no difference between the two experiments for the bilinguals; they were not garden-pathed in the experiment with the intonational cues, as would maybe be expected. Crucially, they did not show effects of garden-pathing when intonational cues to guide them towards a correct interpretation were absent and, thus it would be expected that incremental processing would lead to misinterpretation. This is a qualitative difference between the two groups and, in this sense, not in line with previous studies which showed no qualitative differences between monolingual and bilingual children in processing. This would indicate that the bilinguals were not able to utilise the subcategorization information of the verb to build an expectation of an upcoming object.

Differences between bilingual and monolingual children have also been observed with regards to how they use available information to predict the upcoming content of the sentence that they are listening to. It should be noted that filler-gap dependencies and garden-path sentences also entail the need for prediction as the parser commits to a specific interpretation which subsequently needs to be revised. The reason for this is that the parser does not wait until disambiguation when it encounters ambiguity. Instead, it utilises available knowledge about the relevant syntactic structures from previous exposure to build a syntactic structure and interpretation even though this interpretation is only a possibility prior to disambiguation. The need to revise comes from the failure of the disambiguating segment to confirm the parser's expectations. For example, in wh-questions, the parser anticipates an expected gap on the basis of the filler[7]. The parser will subsequently predict that the first plausible gap for the filler is the real one. For garden-path sentences, the parser utilises the subcategorization information of the verb, i.e. that it is potentially transitive, and therefore, expects an object. Consequently, it will interpret the upcoming NP as the verb's object even though the sentence so far does not indicate this definitively.

Lemmerth & Hopp (2019) investigated how bilingual children use gender on the determiner and/or adjective to predict upcoming nouns in a visual-world paradigm study. Lemmerth &

[7] This is framed in terms of a generativist syntax approach; this would not be inconsistent with alternative accounts but would need to be worded differently.

Hopp tested simultaneous and early sequential (mean age of onset 2 years) L1 Russian L2 German bilingual children aged 6-9 years and compared them to age-matched monolingual German speakers. Participants heard questions in the form “Where is *DETERMINER ADJECTIVE NOUN*?”, as demonstrated in (14a-c) while looking at four pictures.

(14a) masculine gender nouns

a. der kleine Käse

theMASC smallAMBIGUOUS cheeseMASC

‘the small cheese’

(14b) feminine gender nouns

die kleine Karte

theFEM smallAMBIGUOUS cardFEM

‘the small card’

(14c) neuter gender nouns

das kleine Auge

theNEUT smallAMBIGUOUS eyeNEUT

‘the small eye

Gaze data were collected while the participants heard the question and until they responded. Lemmerth & Hopp used nouns in German that either had the same or different grammatical gender as in Russian. They also manipulated the gender cue to be either on the determiner alone (in this case, a definite article) or on the determiner and the adjective preceding the noun. Lemmerth & Hopp found that the simultaneous bilinguals behaved similarly to the monolinguals and used gender to predict upcoming nouns. This was irrespective of whether the noun had the same or different grammatical gender in Russian to German. The sequential bilinguals, however, use the cue on the adjective or the determiner only when the noun in German had the same grammatical gender as it did in Russian. Lemmerth & Hopp argued on the basis of their findings that grammatical gender in the sequential bilinguals was mediated by the lexicon in the first language although this was not the case for the simultaneous bilinguals.

In another study using the visual world paradigm, a study by Lew-Williams (2017) examined whether bilinguals make use of grammatical gender cues as they appear on Spanish definite articles to predict upcoming material, i.e. whether looks to a picture of objects which matched those features increased after hearing the definite article. Participants looked at a set of two pictures and heard either a question in the form of "Where is the…?" or an instruction in the form of "Find the …?". It should be noted that the determiner in Spanish is marked for gender and number and can, thus be used to predict the upcoming noun. That is if one of the two objects matches the grammatical gender of the determiner, the parser will anticipate this noun to be the upcoming noun. If both objects have the same gender, which matches that of the determiner, this information cannot be used. In a subsequent experiment with the same design, Lew-Williams (2017) manipulated the biological gender of the nouns. In a third experiment, Lew-Williams manipulated the number of the objects in the visual stimuli (singular or plural) following the same experimental design. Lew-Williams measured the time needed for participants to reorient their looks towards the correct picture, i.e. the one in the instructions or question as a dependant variable, a type of reaction time. Results showed the school aged bilingual children were able to orient their gaze to the correct picture earlier when there was a disambiguating cue similarly to the monolingual children for number and biological gender but not for grammatical gender. This study, however, tested children in an immersion context where the L2 was not the majority language and compared them to simultaneous Spanish-English bilinguals.

In sum, the results have so far shown some consistencies but at the same time are varied. An emerging consensus is that bilingual children have knowledge of syntax and morphology. This is evidenced by the fact that the bilingual children show sensitivity to ungrammaticalities as reaction times at the point where these occur, or immediately thereafter, are longer than in the respective sentences where there are no ungrammaticalities (e.g. Blom & Vasić, 2011; Chondrogianni & Marinis 2012, 2016; Chondrogianni et al. 2015a, 2015b). This effect is also observed in monolinguals. This entails that any differences observed between monolinguals and bilinguals in production (e.g. Chondrogianni et al. 2015a) are not due to lack of knowledge per se but to other factors which may burden language production. Bilingual children also show similar processing mechanisms; the absence of an interaction between group and grammaticality suggests that both monolingual and bilingual children slowed down equally for sentences with ungrammaticality. The main difference between bilingual and monolingual children in the studies reviewed is that the bilinguals are overall slower in terms of reaction times, suggesting less efficient processing. A further difference is that bilinguals may have greater difficulty with particular conditions which are expected to be harder across groups even if the effects are observed in the monolinguals as well. In terms of accuracy, the majority of studies find lower accuracy for the bilingual group relative to the monolinguals, although both groups still have high above-chance performance. A second difference found between bilingual and monolingual children in terms of morphosyntactic processing is in their reduced ability to utilise available information for purposes of predictive processing or facilitating disambiguation. This ability may be absent completely for some aspects of morphosyntactic processing (Lew-Williams, 2017 for use of grammatical gender on the determiner to predict upcoming nouns, Papangeli

& Marinis 2010 for absence of garden-path effects in bilingual children). Alternatively, it may be more restricted in terms of context (use of gender on the determiner to predict upcoming noun only when the noun has the same gender in both languages; Lemmerth & Hopp, 2018) or type of bilingual (effects found for simultaneous but not early sequential bilingual; Roesch & Chondrogianni, 2016). Available studies have tested late sequential bilinguals with an L2 other than English (Lew-Williams, 2017 for Spanish; Lemmerth & Hopp, 2018; Roesch & Chondrogianni, 2016 for German, Papangeli & Marinis, 2010). There are currently no studies on morphosyntactic processing where there is a need for prediction or disambiguation to be made during processing in bilingual children for English.

1.4. Morphosyntactic processing in bilingual adults

Research in language processing in adult bilinguals has been shaped by the question as to whether sentence processing in the L1 and L2 is qualitatively different or not. The distinction has been framed between monolingual speakers of a language and late sequential bilinguals, often with predominantly classroom exposure, in other words non-native speakers, or L2 learners. One of the most influential theoretical proposals is the Shallow Structure Hypothesis under which L2 learners differ from natives in that their processing is less syntactic in nature as the grammatical constraints are less robust than in monolingual natives (Clahsen & Felser 2006a; 2006b; 2006c). Although L2 learners' syntactic representations were assumed to be shallower and less detailed than those of native speakers (Clahsen & Felser 2006a: 3) and that this entailed fundamental and qualitative differences between the two groups (Clahsen & Felser 2006b: 118), this does not entail deficit. Clahsen & Felser claimed than L2 learners underuse syntactic information (2006a: 21) and are thus, more reliant on information from other sources (e.g. lexicon, pragmatics, intonation) which is assumed to be inadequate compensation. This account originated from the long-standing observation that L2 learners have difficulty with grammar and is grounded in research into real time processing in L2 learners. The latter has showed nativelike attainment for some linguistic features, including some aspects of morphosyntax but also persistent difficulties with other aspects of morphosyntax or complex structures even in learners with a high degree of proficiency (see Clahsen & Felser, 2006a for a review). The Shallow Structure Hypothesis was an attempt to explain differences observed in L2 processing which could not be accounted for under the assumption of a global deficit in L2 learners' capacities but also to integrate L2 processing within existing psycholinguistic approaches to acquisition and processing (Clahsen & Felser, 2018).

One study which lends support to the Shallow Structure Hypothesis is by Marinis, Roberts, Felser & Clahsen (2005) which investigated processing of wh-questions with intermediate gaps as in (15a) in comparison to sentences without the intermediate gap as in (15b) in English L2 learners.

(15a) The nurse who$_i$ the doctor argued e_i ' that the rude patient had angered e_i is refusing to work late.

(15b) The nurse who$_i$ the doctor's argument about the rude patient had angered e_i is refusing to work late.

The difference between the two sentences is that, in (15a), there is a possible filler site for who after the verb "*argued*"; it is likely that the parser will utilise subcategorization information about the verb "argued" and interpret the filler as the subject to an upcoming verb, but would then need to revise the interpretation and insert the filler in the object position of the verb "angered". This earlier occurring possibility is not possible in (15b). Marinis et al used a self-paced reading task followed by a comprehension question similar to Gibson & Warren (2004) but with L2 learners whose native language was Chinese, Japanese, German or Greek with at least an upper intermediate level of proficiency. Native English speakers were also tested as controls. At the segment of the intermediate gap, Marinis et al (2005) found a slowdown in reaction times for the native speakers for sentences with an intermediate gap in comparison to the respective segment in the control sentences. Conversely, they found a facilitation effect, i.e. faster reaction times, two segments thereafter, at the subcategorising verb "angered" for sentences with an intermediate gap compared to sentences without one. This was not found for any of the groups of L2 learners. This was interpreted as evidence that the L2 learners, irrespective of the L1, did not posit any intermediate gaps during processing as expected by filler-driven structural accounts suggesting they "underuse syntactic information" (2005:70). However, Dedykspotter, Schwartz & Sprouse (2006) reanalysed some of the data from this study and found the same effect for the L2 learners but at the subsequent segment indicating the effects were delayed but present. This, however, may reflect the storage of the filler in working memory. A limitation of the Marinis et al. (2005) study is that two types of sentences differed in the word length preceding the critical segment due to differences in structural complexity.

Differences between L2 learners with naturalistic relative to those with classroom exposure were found in another study by Pliatsikas & Marinis (2013a). In this follow up study to Marinis et al. (2005), Pliatsikas and Marinis compared L2 learners and tested wh-questions with intermediate gaps as in "The politician who the journalist predicted that the government report would bother is calling a press conference" using a self-paced reading task and the same material as in Marinis et al. (2005). At the intermediate gap segment, only the native speakers showed a slowdown for sentences where there was an extraction suggesting the L2 learners did not process the intermediate gap. All three groups showed a slowdown for the sentences with an extraction relative to those without one at the following segment. At the segment of the subcategorising verb, the native speakers and the L2 learners who had naturalistic exposure to English showed facilitation, i.e. faster reading times, for the sentences with a VP extraction relative to those with an NP extraction. This was absent for the L2 learners with only classroom exposure and is suggestive of a more lexically-driven processing. The facilitating effect of the intermediate gap was interpreted as evidence for the use of syntactic cues. The results from Pliatsikas & Marinis indicate that the nature of processing in L2 learners changes with exposure; learners who become bilingual with a more naturalistic exposure become more nativelike in how they process sentences even for structures such as intermediate gaps which are infrequent in spoken English. The results challenge the Shallow Structure Hypothesis as the predictions it makes are empirically confirmed only for a subset of L2 learners relative to its predictions; only the advanced L2 with classroom exposure showed different and less structural patterns of processing to the natives[8].

In another study, Felser, Roberts, Marinis & Gross (2003) investigated how L2 learners with German or Greek as their native language process sentences in English with ambiguous NP attachment such as "*The dean liked the secretary of the professor who was reading a letter*". The ambiguity in this sentence is who was reading the letter, the secretary or the professor and speakers of different L1s tend to have a preference for either one or the other depending on the language. Both groups undertook an offline task where they heard sentences as in the example and had to answer comprehension questions relating to the attachment and then did a self-paced reading task with the same type of sentences. For both tasks, Felser et al. manipulated the preposition to be either "*of*" or "*with*". Moreover, in the self-paced reading task, the number of the NPs was manipulated so that only one type of attachment was possible and this became apparent at the auxiliary verb in the embedded clause. In other

[8]For different results on processing of inflected verb forms where both L2 learners with naturalistic exposure and ones with classroom exposure perform in a nativelike manner, see Pliatsikas & Marinis (2013c)

words, if the participants' preferred attachment was inconsistent with the one available at the point of the auxiliary, due to subject verb agreement, a slowdown is anticipated. The L2 learners whose native language was German showed a preference to attach the relative clause to the second NP with the preposition "with". This was evidenced by decreased reaction times at the critical segment when subject-verb agreement forced NP2 attachment. This was also found for the Greek speakers. Native speakers also showed this preference for sentences but with both the preposition "with" and "of". The L2 learners, however, showed no differences between conditions for "of" suggesting there was no particular attachment preference for sentences with the genitive. Similar results were found for the off-line comprehension questions. In sum, Felser et al. argued that the participants could use lexical information from the preposition "with" but this was not possible with "of". Felser et al. interpreted the results to show that L2 learners could use lexical or semantic information during processing. The difference between the two prepositions is assumed to be that "with" provides semantic information as it can assign a thematic role whereas "of" does not. Furthermore, they argue that there was little evidence for an influence from the L1 as German and Greek speakers displayed the same processing patterns even though the two languages have contrasting attachment preferences. The native NP2 attachment preference in the "of" conditions was interpreted as the result of attaching the NP to the structurally closest constituent to the predicate phrase (here, the NP2 interpretation). In this sense, the natives followed a syntactically-driven strategy of interpreting ambiguity but the L2 learners did not. Therefore, Felser et al (2003) interpreted the findings as in moderately support in shallow structural processing in L2 learners[9]. Papadopoulou and Clahsen (2003) conducted a similar

[9] It should be noted that not all studies which examine processing of relative clauses support this; for example, Hopp (2014) shows that relative clause attachment is dependent on working memory capacities and lexical processing with L2 learners with higher performance on both measures, on a par with native speakers, showing nativelike patterns of processing.

study to Marinis et al. but in Greek testing L1 and L2 Greek (L1 Spanish, German or Russian) and found similar results.

An alternative theoretical account which has been applied to processing to account for differences between adult natives and L2 learners is the Interface Hypothesis[10] (Sorace, 2011). Under the latter, L2 learners may not reach nativelike attainment for linguistic phenomena which involve a combination of syntactic knowledge and other cognitive abilities. As applied to processing, this entails an inability for the L2 learners to integrate syntactic information with other information in real time. Several studies have been argued to support the Interface Hypothesis. For example, Belletti, Bennati & Sorace (2007) studied the production and interpretation of overt, null and post-verbal subjects in Italian in L1 English L2 Italian adults. All 3 are grammatical in Italian but the use of each one is argued to be determined by pragmatic, in particular discourse factors. In a series of elicited production tasks, Belletti et al. demonstrated that adults with near native L2 Italian used overt pronominal subjects more frequently than L1 native Italians and were more likely to interpret overt subject pronouns in an embedded clause as coreferential with the subject of the matrix clause relative to the control group. On the other hand the L2 learners produced more post-verbal subjects than the native controls but just as many null subjects, even though the latter aren't available in English. In sum, Belletti et al. (2007) show that even at high levels of L2 proficiency, some difficulties persist for features of morphosyntax the use of which interacts with discourse. In other words, Belletti et al. argue that the morphosyntactic features have

[10] Both the Shallow Structure Hypothesis and the Interface hypothesis account for differences between native speakers and L2 learners of a language in terms of the information utilised during processing but do not assume differences in representation. Other models do posit such differences but are not discussed further as they are less relevant to this thesis - for example, the declarative/procedural model distinction (Ullman 2001, 2005, 2015). However, support for this type of model is derived from studies on morphological processing rather than morphosyntactic processing (e.g. Pliatsikas & Marinis, 2013a, 2013c, possibly 2013b as well, albeit less directly)

been acquired for L2 Italian; within a principles and parameters approach, they argue that the null-subject parameter has been set for Italian in the L2 learners. The source of the difficulty is assumed to be the relevant settings of the L1 which are assumed to be accessed in real time processing of the L2.

In another study supporting the Interface Hypothesis in adult L2 learners, Sorace & Filiaci (2006) examined the interpretation of anaphora within bi-clausal sentences in a similar population of L2 Italian learners. Sorace & Filiaci manipulated the subject pronoun to be either null or overt and which clause was the first in the sentence (i.e. the main clause preceded the subordinate clause and there was forward anaphora or the subordinate clause preceded the main clause and there was backward anaphora). The study used a picture verification task as in Belletti et al. where participants needed to select a picture to match the sentence they heard. Sorace & Filiaci found no differences between native speakers of Italian and respective L2 learners in terms of null subject pronoun interpretation; for forward anaphora both groups interpreted the null subject in the embedded clause equally as coreferential to the subject or complement of the main clause but mainly as coreferential to the subject of the main clause in the backwards anaphora condition. Differences between the two groups emerged in the interpretation of the overt subject pronouns. In the forward anaphora condition, the native speakers were more likely to interpret the overt pronoun as coreferential with the complement of the main clause whereas the L2 learners preferred to interpret it as coreferential with the subject of the main clause. In the backwards anaphora condition, the L2 learners interpreted the pronoun as the subject of the main clause, something the natives rarely did, but the natives interpreted it as referring to an extrasentential antecedent, which was also uncommon for the L2 learners. Sorace & Filiaci argue that advanced L2 learners had a representation of the null and overt subjects in Italian but had difficulties using the pragmatic constraints that differentiate between the two in real time.

Other accounts advocate no qualitative differences between processing in the L1 and the L2 but posit that L2 processing is cognitively more demanding (e.g. McDonald, 2006; Hopp 2006, 2010). Such accounts consequently predict that differences between L1 and L2 processing can be predicted by individual differences, i.e. they will be larger in L2 learners with more limited memory capacity, or may become more evident in tasks depending on the cognitive demands involved. In a study involving a range of tasks, McDonald (2006) compared L2 learners of various linguistic backgrounds with native speakers of English. The tasks involved a working memory span, a gated word recognition task (in which the segments of the word were progressively presented), a word detection task to measure speed of processing and a grammaticality judgement task, all of which were administered in English. With the exception of response speed on grammaticality judgement, the L2 learners underperformed on all measures relative to the native speakers. McDonald carried out a second experiment with native speakers. In this experiment, the participants were placed under stress conditions by manipulating either the working memory load (high vs. low; participants needed to remember four or seven digits) or speed (normal speed vs. compressed speech) or making decoding harder (by listening through white noise) with each participant being assigned to one condition. Participants undertook the same tasks as in the first experiment. In all measures except the low memory span condition, native speakers' performance dropped and for the noise condition performance on the grammaticality judgement task was not significantly better than that of the L2 learners from the first experiment that were under no condition of "stress". This indicates that L2 processing is more resource intensive that native language processing.

More indirect evidence in support of this notion comes from studies testing L2 learners which show more nativelike patterns of processing with near nativelike proficiency. One study in this direction comes from Hopp (2006); in this study, advanced L1 English and Dutch speakers of German were compared to native speakers using a self-paced reading task to test subject-object ambiguities. In the self-paced reading task, the ambiguous sentences where those which were not disambiguated by case but by subject-verb agreement which was clause final as the clauses were embedded. All groups showed lower accuracy for those sentences which were ambiguous but differed in the reading times; only L2 learners with a near native proficiency score in German as established by a close test behaved like the natives and showed a slowdown at the point of disambiguation.

In another study, Hopp (2010) compared English-German, Spanish-German and Russian-German adult speakers to native German speakers using a range of tasks on subject-verb agreement, grammatical gender agreement and case marking. In a grammaticality judgement task, near-native L2 learners differed from advanced L2 learners and clustered with native speakers in that they did not accept sentences with an ungrammatical word order. Furthermore, in a self-paced reading task similar to the one in Hopp (2006), near-native L2 learners behaved similarly to native speakers in that they were sensitive to both case marking and subject-verb agreement (similarly to Hopp, 2006), as they showed a slowdown at critical segments were there was disambiguation. This was not found to be the case for the advanced L2 learners of any of the language combinations[11].

The studies reviewed so far have treated the differences in L2 proficiency as categorical. More recent studies have examined proficiency and various measures of L2 competence and cognition as a continuum in order to predict individual differences. One such study is Hopp (2014) who examined relative clause attachment in German-English speakers compared to

[11] Hopp (2010) found similar results for two speeded grammaticality judgements as well.

English native speakers on an eye-tracking reading study of temporarily ambiguous sentences. Hopp showed that attachment preference was modulated by reading span and word level processing (as evidenced by a lexical decision task). Non-natives matched with natives in capacity for working memory and lexical processing demonstrated nativelike parsing. Reading times were significantly higher at the disambiguating region and overall for sentences where the ultimate interpretation favours high attachment than low attachment. As lexical recognition automaticity increases, preferences become more nativelike, i.e. favouring low attachment. These findings are directly contradictory to the Shallow Structure hypothesis as near-native L2 learners show parsing that is similar to that of natives. Moreover, the impact of reading span and lexical decision speed suggests that processing is modulated by factors of language ability. This is again not consistent with the Shallow Structure Hypothesis as the differences between natives and non-natives are shown to be potentially overcome even in cases of late or non-naturalistic exposure. Similar results have been found using an eye-tracking reading study with the same population of L2 learners which tested the processing of subject-object garden path sentences in English (Hopp, 2015).

1.5. This thesis

1.5.1. Aims, research questions and predictions

In sum, there are two fundamental gaps in the literature. 1) Research in bilingual children has focused predominantly on language production. Research on language processing in bilingual children has examined processing of ungrammaticalities in both nominal and verbal morphosyntax. However, studies so far have relied on reaction times and have not examined eye-movements which provide more fine-grained information about the timecourse of processing. Therefore, they have not addressed the issue of incrementality of processing in bilingual children, their ability to revise misinterpretation or the extent to which they utilise information from various sources to facilitate processing. 2) The second gap is that, while there is ample research on language processing in bilingual adults, research has focused on the differences in language processing between native speakers and non-native L2 learners / late bilingual speakers of the language. There is limited work on language processing in bilingual adults who are native speakers of the additional language as well as of their native language (i.e. early bilinguals with naturalistic exposure).

This thesis aims to bridge the aforementioned gaps in the literature on language processing in bilinguals in two ways. Firstly, it includes two eye-tracking studies on morphosyntactic processing in bilingual children; these studies each address a linguistic phenomenon which has so far been explored only in monolingual children, namely long-distance dependences in the form of wh-questions and garden-path sentences. Second, it replicates the same two eye-tracking studies with bilingual adults who have acquired English at an early age with naturalistic exposure, similarly to the bilingual children and compares them to both monolingual adults and L2 learners of English.

The main research questions are:

1) Do bilingual children differ in morphosyntactic processing to monolinguals in terms of overall accuracy and efficiency?

2) Do bilingual children successfully build syntactic representations in real time as monolingual children do? Do they revise them, if needed, as monolingual children?

3) Do bilingual children make use of various sources of additional information (such as cues from feature (dis)similarity or referential context) during language processing to facilitate processing in the same way as monolingual children?

4) Do bilingual adults process morphosyntax similarly to monolinguals or L2 learners in terms of overall accuracy and efficiency?

5) Do bilingual adults successfully build syntactic representations in real time as monolingual adults? Do they revise them, if needed, similarly to monolinguals or have difficulty similarly to adult L2 learners?

6) Do bilingual adults make use of various sources of additional information (such as cues from feature (dis)similarity or referential context) during language processing to facilitate processing in the same way as monolingual adults? Do they differ in this respect to L2 learners?

1.5.2. Overview of chapters

The rest of this PhD thesis is organised as follows. Chapters 2 and 3 present the results from two eye-tracking – visual-world paradigm – studies with bilingual children and adults respectively (Studies 1A and 1B) examining wh- questions, while chapters 4 and 5 present similar eye-tracking studies with garden-path sentences for children and adults respectively (Studies 2A and 2B). Research questions 1-3 are addressed in Chapters 2 and 4. Research questions 4-6 are addressed in Chapters 3 and 5. Each experimental study has specific research questions which fit within the broader questions outlined in section 1.5.1. To address the research questions, three metrics are used; accuracy, reaction times and gaze data, with the latter used to examine processing over time. More detail on the conceptual motivation, the design and the results and their interpretation are included in each chapter individually.

Chapter 6 summarises the findings from all studies and presents a broader picture of morphosyntactic processing in bilinguals given the evidence and in light of current theories of language processing. It aims to connect the findings to the broader research questions of this thesis.

Chapter 2 Wh-question processing in bilingual children

2.1. Introduction

There is ample research on the acquisition of long-distance dependencies, including wh-questions in children but the majority has focused on production or offline end stage comprehension tasks (e.g. Deevy & Leonard, 2004; Goodluck, 2005; Omaki, Davidson-White, Goro, Lidz & Phillips, 2014; Stromswold, 1995). There is a substantially smaller body of evidence for online language processing in children, in particular using eye-tracking; this has either focused on local ambiguity as a window into online processing or has looked at other types of long-distance dependencies (e.g. Love, 2007; Roberts, Marinis, Felser & Clahsen, 2007). There is even less available work on real-time language processing in bilingual children. The latter has focused on morphosyntactic features, in particular, sensitivity to ungrammaticalities in case and gender (for a more detailed review, see Chapter 1). How bilingual children process which-questions in real time remains largely underexplored. This study aims to bridge this gap in the literature.

This chapter begins by summarising the main theoretical views on filler-gap dependencies and how these conceptualise their processing. It then reviews studies which examine how children process wh-questions and subsequently, studies demonstrating how monolingual children use cues to aid processing of filler-gap dependencies. It then outlines how the current study fits into the existing body of research. The experimental design is then presented, followed by the results and discussion.

Wh-questions

Wh-questions have been shown to be a challenging linguistic structure to be acquired in children and there is evidence that object questions create a processing difficulty even in adults. Whereas object wh-questions appear early on in child language and around the same time as subject wh-questions in production (Stromswold, 1995), difficulties in comprehension of object which-questions have been shown to persist until early school years (Goodluck, 2005; Deevy & Leonard, 2004) but not in other types of wh-question (Avrutin, 2000). Similar difficulties with object wh-questions have also been observed across numerous typologically varying languages, such as French (Jakubowicz & Gutierez, 2007), Italian (De Vincenzi, Arduino, Ciccarelli & Job, 1999), Greek (Stavrakaki, 2006) but also Hebrew (Friedmann, Belletti & Rizzi, 2009).

The asymmetric difficulty in processing is predicted by several language processing theories (for more detail see Chapter 1). Under structure-based accounts such as the Active-Filler hypothesis, speakers will recognise that the wh-question is a candidate filler and will insert it at the earliest possible gap, i.e. the position of subject of the verb (Frazier 1987, Frazier & Clifton, 1989). The source of the difficulty for one type of question relative to the other lies in the fact that this strategy does not work for object questions as it leads to misinterpretation. On the other hand, under expectation-based probabilistic models, object questions are harder as they are less frequent in the input and, as such, speakers of English have been tuned to expect a subject question when hearing the which-phrase (Hale, 2001, Levy, 2008; Roland, Dick & Elman., 2007).

Beyond the subject-object asymmetry found across wh-questions, which-questions are assumed to be of increased difficulty relative to other types of wh-questions (e.g. who or what). One reason for this is that which-questions are d(iscourse)-linked and, thus, require a link to preceding discourse. In other words, one is asking about a particular entity out of several which was previously mentioned. In contrast, this is not the case for questions with who or what. This has been assumed to result in an additional burden on cognitive resources during processing leading to greater difficulty processing which-questions. This has been shown in a number of studies with several populations and experimental methodologies (Kaan, Harris, Gibson & Holcomb, 2000 for healthy adults and ERPs; Avrutin, 2000, Hickok & Avrutin 1996 for production and offline comprehension in young children and adults with aphasia).

Moreover, some which-questions have a more restricted subset than other which-questions from which one needs to select a potential referent. For example, sentence (16) has a more specific set of referents than example (17) as the semantic properties of the candidate entities in (16) are more than those in (17);

(16) Which donkey is carrying the zebra?

(17) Which person is carrying the zebra?

In (17) one only needs to select an animate entity which is human. In contrast, (16) requires the selection of an animate entity which happens to be an animal and in particular, a certain type of animal. Using a self-paced listening task with truth-value judgement, Donkers, Hoeks & Stowe (170163), compared participants' processing of wh-questions with who, what and which-person vs. which-N - as in (17) vs (16) – in adult L16 Dutch speakers. They were able to show lower judgement accuracy and longer listening times for which-questions relative to the other two types and greater difficulty for which-questions with a more restrictive subset of candidates (i.e. Which-N) than for ones with more generic subsets (which person). This study focuses on which-questions because they appear to be a particularly challenging phenomenon.

Incrementality of wh-question processing in children

One study which examines real time processing in children at the age of the participants in this study is one by Marinis & van der Lely (2007). Marinis & van der Lely used a cross-modal priming paradigm where children, aged around nine years, heard sentences and responded to a visual stimulus which they needed to judge in terms of animacy. The experimental sentences were indirect object wh-questions which followed a short one-sentence story while the visual stimuli were presented at three positions; onset of verb, the trace and a control position between the two. The visual stimuli were either the same as the antecedent the participants heard in the short story or unrelated. It is expected reaction times will be faster for the related word at the trace position relative to the unrelated word or the other positions. This facilitation was indeed found for the typically developing children in this study and was interpreted as the result of antecedent priming at the gap position. This entails a structural dependency between the filler and the gap.

One study which tests incremental processing in child language using wh-questions is the study by Omaki, Davidson-White, Goro, Lidz & Phillips (2014). They examined comprehension of embedded wh-questions in L1 English and L1 Japanese 4-and-a-half to 6-and-a-half-year-old children as well as adult controls. In both languages, these are ambiguous as it is not clear which gap position the fronted wh-phrase fills. Using an off-line comprehension task, participants were given stories and were asked a comprehension question, such as (18).

(18a) "Where did Lizzie say that she was gonna catch butterflies?".

(18b) "Where did Lizzie tell someone that she was gonna catch butterflies?".

(18a) "Where did Lizzie say to someone that she was gonna catch butterflies?".

Responses were coded as reflecting a main clause interpretation where the wh-word was interpreted as an adjunct to the main verb or an embedded clause interpretation where it was associated with the embedded verb. Both adult controls and children provided fewer responses consistent with a main verb interpretation for sentences as in (18a) than in (18b) and (18c). For Japanese sentences with local syntactic ambiguity, there was a preference for an embedded clause interpretation of sentences such as (3)

(19a) Doko-de Yukiko-chan-wa [kouen-de choucho-o tsukameru to] itteta-no?

where-at Yukiko-DIM-TOP pro park-at butterfly-ACC catch COMP was telling-Q

"Where was Yukiko telling someone that she would catch a butterfly at the park?"

However, due to differences in word order between Japanese and English (Japanese is a head final language whereas English is head initial), this mirror image reflects the same mechanism; the embedded clause in Japanese is centre-embedded and is the first clause of the two to have a position available for a gap for a gap. Therefore, all groups preferred to insert the filler in the earlier gap position available for globally ambiguous questions. When an additional prepositional phrase was added to specify the location of the embedded clause event, hence making the question locally ambiguous (English example: "Where was Yukiko telling someone that she would catch a butterfly at the park?"), the adults gave a main clause interpretation more often whereas the children did not. This suggests that the children had difficulty re-analysing the sentence.

The Omaki et al. study shows incremental processing and difficulties recovering from garden-path effects, but these conclusions are extrapolated based on off-line findings. Similar evidence for incremental processing has been found in picture selection tasks for relative clauses in French with young children (Bentea & Durrleman, 2018). Other studies using reaction times, a more fine-grained off-line method, also suggest similar processing mechanisms for both adult and children for filler-gap dependences (e.g. Love, 2007; Marinis & van der Lely, 2007; Roberts, Marinis, Felser & Clahsen, 2007).

Eye-tracking studies with filler-gap dependences are only now beginning to emerge. Atkinson, Wagers, Lidz, Phillips & Omaki (2018) investigated locally ambiguous filler-gap dependences such as “Can you tell me what Emily was eating the cake with?” with children aged 5 and 6 years. The temporary ambiguity in this example is the syntactic function of the word “what”; until the VP “was eating”, it is assumed that it is the object of the verb. However, in the subsequent segments, this reading becomes unsustainable and the listener reinterprets the wh-word as the instrument of the verb. In this study, Atkinson et al. find a bias towards filling the gap as early as possible for adults and older children (although the children showed this in the spill over NP rather than the VP region as the adults did). The results are consistent with the existing literature which suggests that children process language incrementally. The younger children, however, did not show this bias.

Cues in incremental wh-question processing

To our knowledge, the first study to investigate processing of which-questions in children using online measures is a recent study by Contemori, Carlson & Marinis (2018). The study in question investigated this particular type of wh-question as these have been shown to be particularly difficult for children and where the subject-object asymmetry is most pronounced (Avrutin, 2000). In an eye-tracking task, monolingual English-speaking children aged 5-7 years as well as monolingual adult controls, heard subject- and object-extracted which-questions such as (20a) and (20b) respectively whilst looking at two pictures with exactly the same content (agent, action, theme, all involving animals).

(20a) Which donkey is carrying the zebra?

(20b) Which donkey is the zebra carrying?

The difference between the two pictures was that the argument structure depicted was reversed so that one picture corresponded to (20a) while the other did so to (20b); in other words, one picture depicted a donkey carrying a zebra; the other one showed a donkey being carried by a zebra. The corresponding picture (i.e. the target) for the subject questions was the incorrect picture (i.e. competitor) for the equivalent object question and vice versa. Participants then needed to select the picture which matched the question heard by clicking on it. Analysis of accuracy and gaze data showed a persistent disadvantage for the object questions for the children. The gaze data showed similar processing mechanisms for adults and children through similar patterns for each of the two question types. For subject which-questions looks to matching picture the target increased over time but for object which-questions they initially decreased and then increased. This change indicates that object-questions such as “Which cow is the goat pushing?” were initially interpreted as a subject question and only after the disambiguating second NP did the hearer reanalyse the question and build a different syntactic representation.

Contemori et al. tested Relativized Minimality in which-question processing in children and therefore, manipulated the number of the two noun phrases in the which-questions so that it would either be the same (match) as in (21a) and (21b) or different (mismatch) as in (21c) - (21d).

(21a) Subject – match: Which cow is pushing the goat?

(21b) Object – match: Which cow is the goat pushing?

(21c) Subject – mismatch: Which cow is pushing the goats?

(21a) Object-mismatch: Which cow are the goats pushing?

Examples (21b) and (21d) are both object questions but only the latter has two NPs with mismatching features. Under Relativized Minimality, this mismatch in features will facilitate the parser to process (21d) relative to (21c). Under Relativized Minimality, it is the mismatch of features per se that has the facilitatory effect as the second NP is interpreted as the only grammatically possible subject in order to avoid a subject-verb agreement error and is, thus, blocked from functioning as a competitor to fill the object position. A second reason that (21d) might be easier to process than (21b) is that disambiguation is possible earlier on; in (21d), it is possible with the onset of the auxiliary verb but with (21c), this is only possible after the second NP. This may allow the parser more time to form a novel, correct interpretation of the sentence but is not related to Relativized Minimality.

There is no motivation for number mismatch to benefit processing of subject-questions, i.e. (21c) relative to (21a). It was expected that (21a) and (21c) would not be particularly problematic for processing and that therefore, the number mismatch would not provide any particular observable aid as there is no need to reanalysis because the subject bias would guide speakers to a correct interpretation. On the other hand, (21b) and (21d) were predicted to be harder as speakers would be prone to misinterpret them as subject extractions, and in this case the number mismatch in (21d) would have facilitated processing relative to (21b). This was indeed the case in Contemori et al. Looks to target over time were faster for object which-questions when the number of the two NPs was different than when it matched.

The study by Contemori et al. faces two limitations; it does not include a full paradigm; singular first NPs were always used, so the match and mismatch conditions are singular-singular and singular-plural respectively. The plural number however has been described as the marked number option relative to the singular in linguistics. Marked features have been associated with additional complexity or difficulty (for a review on markedness in morphology see, Haspelmath, 20021). Both morphological and structural properties of a language typically labelled as marked are usually acquired later than unmarked features (e.g. Harley & Ritter, 2002). It is worth noting that this effect of markedness has also been attested in processing even in adult L1 speakers of English. For example, Wagers, Lau & Phillips (2009) have shown that marked features cause more disruption to computing subject-verb agreement, another type of syntactic dependency, than unmarked features. More specifically, plural nouns caused more attraction errors than singular nouns. The second limitation is related to the age of the participants. It is expected that number mismatch will have a facilitative role in developing linguistic systems. It may still have an impact in fully developed adults' systems, but the effect is likely to be smaller as the benefit will be more redundant. Confirming this, in the Contemori et al. study there was an effect of number match on accuracy for the 21 7–year-old children while the adult data was not analysed as it reached ceiling. The impact of number mismatch as a facilitatory feature is unclear in older school age children whose processing may be more akin to that of adults.

The studies reviewed so far examine wh-question processing in monolingual children. Yet, few studies have examined wh-question in bilingual children. As discussed in Chapter 1, one study to do so is Roesch & Chondrogianni (2016). The study tests comprehension of German subject and object which-questions in bilingual children. To this aim, a picture selection task was used with monolingual children and two types of bilingual children, simultaneous and early sequential bilinguals. Similarly to Contemori et al (2018) but using off-line measures only, this study examined how bilingual and monolingual children are facilitated by the presence of disambiguating cues for processing which-questions in German. More specifically, Roesch & Chondrogianni tested how nominal case in German helps children disambiguate which-questions and controlled for the position and number of disambiguating cues; the latter were either on the wh-word, the second NP only – in both cases there was a single cue or a disambiguating cue on both phrases (double cue). Accuracy was higher for subject questions relative to object questions and when there were two disambiguating cues relative to when there was only one. Furthermore, the presence of the single disambiguating on the wh-word was more beneficial than the single cue at the second NP. The monolingual children outperformed both groups of bilinguals. The impact of case differed across groups; the monolinguals made consistent use of the cue, accuracy was higher when case disambiguated whether an NP was a subject or object, while the simultaneous bilinguals did so only in the initial position while the sequential bilinguals did not do so at all. Age of onset but not cumulative length of exposure was a predictor of accuracy for the object questions.

In sum, so far research suggests that even young monolingual children process sentences incrementally and can use number mismatch between two noun phrases where there is ambiguity to facilitate processing. The timecourse of which-question processing in bilingual children remains an empirical question. Available evidence suggests their ability to utilise cues in real time to aid processing is more nuanced than in monolingual children. This observation holds for both use of number mismatch in which-question processing but also for other types of grammatical information (see Chapter 1).

The current study

The present study examines how bilingual children process which-questions and how they may differ from monolinguals. It is also conceptually similar to Roesch & Chondrogianni in that it investigates how bilingual children utilise disambiguating features when processing which-questions in English; however, it differs in several ways; First, the most important difference is that this study uses an online task in addition to offline measures. Second, the disambiguating feature is number, not case. The reason for this is that English does not have overt nominal morphology for nouns other than number. Third, we do not control for the position or number of disambiguating cues (the match or mismatch always becomes apparent after hearing the first NP and the subsequent auxiliary). A further important distinction is that this study tests older children with a longer length of exposure but does not differentiate between early sequential and simultaneous bilinguals.

Building on previous work, we examined the processing of wh- questions in bilingual and monolingual children using the same experimental methodology as Contemori et al (2018). To our knowledge, it is the first eye-tracking study on wh-questions in bilingual children. The aims are to examine which-question processing in bilingual children relative to monolingual children and whether this differs between subject- and object-questions. We further explore the timecourse of processing of the two structures across the two groups. In other words, it investigates whether there is incrementality in syntactic processing in bilingual children as has been established for monolinguals. Moreover, it examines the impact of number mismatch of two NPs as a facilitatory cue in line with predictions made under Relativized Minimality for bilingual children and for older monolingual English-speaking children. Lastly, this study addresses the limitation of the Contemori et al. study of using only singular first NPs and investigates the impact of the number of the first NP on which-question processing.

Our research questions are,

1. Do bilingual children differ in their performance to monolinguals in processing which-questions in English on a range of measures reflecting the end result?
2. Does the timecourse of processing differ between bilingual and monolingual children? Is there evidence of incrementality in processing in bilingual children?
3. Do bilingual children make use of morphological cues to facilitate processing of which-questions in the same way monolingual children do?

4. Does the number of the first noun phrase have an impact on the processing of which-questions?

The current literature provides little insight into how the bilingual children may perform specifically for the gaze data although it is likely there will be few differences on accuracy overall, but reaction times may be slower for the bilinguals (Chondrogianni & Marinis, 2012; Chondrogianni et al., 2015). It is unclear what will be the case with the gaze data due to the dearth of relevant studies; yet some hypotheses can still be made. If bilinguals process language incrementally, it is expected that there will be an initial misinterpretation of object questions as subject ones consistent with both the Active Filler Hypothesis and expectation-based models; looks to target will drop below chance and increase thereafter. This is not expected to be the case for subject questions where looks will increase from the beginning and will plateau earlier. This would be consistent with what was found for both adults and monolingual children in Contemori et al. (2018). A difference between the two groups in the way they process each question type would be manifest by an interaction of group with structure. One possible difference is that the timecourse will be different. Contemori and colleagues found a slower increase in looks towards target for the children relative to the adults. This can be interpreted as slower but qualitatively similar processing. Bilingual children may also show a slower increase in looks towards the correct picture relative to the monolingual children mirroring the slower reaction times found for self-paced listening studies (Chondrogianni & Marinis, 2012; Chondrogianni et al., 2015).

If bilingual children do make use of the number mismatch in accordance to Relativized Minimality, there will be an interaction with structure and number match; in other words, the effect of number (mis)match should be present only in the object questions. In Contemori et al. (2018), this was found for accuracy although reaction times were not analysed. It is, therefore, expected to be found in the present study for accuracy and it is conceivable that reaction time will also show a similar effect. Garden-path or expectation-based models of language processing do not make explicit predictions about the effect of number markedness. However, the available literature from adult processing (Wagers et al., 2009) would predict that the unmarked forms will be easier to process for both groups consistent with the literature on their acquisition and use as well. Attraction errors, whereby subject-verb agreement is violated by the verb agreeing with the linearly closest NP, are more common with plural NPs (Bock, 1995). Attraction errors are known to have been found for questions with wh-movement where the verb agrees with the object instead of the subject (Dillon, Staub, Levy, Clifton, 2017).

2.2. Method

2.2.1 Participants

A total of 68 children from Grades 3-6 participated in this study. The sample included 37 monolinguals aged 7;10-11;6 (M=9;7, SD=1;1, 16 girls and 21 boys) and 31 bilingual children 7;4-11;5 (M= 9;6, SD=1;2, 17 girls and 14 boys). There was no difference in age between the two groups $F (1, 65) = .265$, $p=.608$, $\eta_p=.004$). None of the children had a history of language impairment or learning difficulty based on a parental questionnaire and teachers' reports[12].

The bilingual children came from a variety of linguistic backgrounds; 6 children had French as their L1, 3 had Greek, 2 had Chinese, Serbian and Spanish and two were bilingual Polish-Serbian. One child spoke Arabic, Flemish, Italian, Latvian, Lithuanian, Macedonian, Slovak, Surashtra, Tibetan and Urdu as a first language and two children were bilingual from home (Russian & Georgian and Turkish & Urdu). All monolingual participants were born in the UK except two who were born in Australia. About half the bilingual children were born in

[12] For an overview of the children's scores in formal assessments see Appendix X

English majority countries (all in the UK except one who was born in Canada). Four children were born in France and one was born in China (AR Tibet), Georgia, Italy, Libya, Lithuania, Russia and Spain[13]. The bilingual children born in the UK or Canada were reportedly exposed to English from birth mainly or early on through day care. For children born in non-English speaking countries, the picture varied in terms of age of onset and length of exposure (see Table 2 for details); one child was born in France and moved to the UK aged 9 but was exposed to English from birth, as s/he was growing up in a one-parent-one-language setting. Four children were exposed to English after their L1 as early sequential bilinguals, between the age of 2 and 3 years, while children were exposed to English as late sequential bilinguals between the ages of 5 and 8 years. These children were born in non-English speaking countries and both parents spoke a language other than English at home. All children in this study had a minimum exposure to English of 2 years (for details, see Table 2); for the children born in the UK and Canada, this was equated to their chronological age (average 9; 3, SD= 13.85 months, range 6;2-10;3); for the children who were reportedly exposed to English after birth, average length of exposure was under 5 years (M=4;10, SD= 25 months) and ranged between 2;5 and 7;4. The children who were first exposed to English before 4 years of age had a mean exposure nearly 7 years (M=6;8, SD= 5.74 months, range 6;2-7;4). The children exposed to English at a later age had a substantially shorter length of exposure (M=3;4, SD= 17.42 months, range 2;2-5;6).

It should be noted that recent research has tended to examine bilingualism as a spectrum of linguistic abilities and cognitive states (Serratrice, 2018; Luk & Bialystok, 2013) depending on either proficiency in one or both languages or circumstances under which the additional language was acquired. For the purposes of this thesis, the bilingual children were included in a single group and compared with the monolingual children. This is acknowledged as a

[13] The country of birth was not provided for one child

limitation in Chapter 6 of this thesis. Tables 1 and 2 summarise key demographic data about the two groups of children.

Ethical approval for this study was granted from the School of Psychology and Clinical Language Sciences Research Ethics Committee. Children were recruited either through mainstream schools in the area of Reading and Southampton (UK) or privately through a University of Reading webmail for employees or word of mouth. Separate information sheets and consent forms were completed by children and parents.

Table 1 Age of participants

	Monolinguals	Bilinguals
N	37	31
Range (in years; months)	7;10-11;6	7;4-11;5
Mean (in years; months)	9;7	9;6
SD (in months)	12.50	13.63

Table 2 Age of Onset (AoO) and Length of Exposure (LoE) to English for bilingual children

Exposure to English	N		Range (in years; months)	Mean (in years; months)
Around birth	20	AoO	0[14]	0[1]
		LoE	6;2-10;3	9;3
Early (<4 years)	4	AoO	2;0-3;0	2;6
		LoE	6:2-7;4	6:8
Later (>5 years)	5	AoO	5;6-8;0	6;8
		LoE	2;4-5;6	3;4

[14] By definition that would be 0 with no variation

2.2.2 Design

The study is a visual world eye-tracking task. In this task, participants, heard a wh-question and looked at two pictures. Both pictures contained two animate entities (animals) with one doing something to the other (e.g. carrying). The two pictures differed in that the thematic structure had been reversed so the agent in the one picture was the theme in the other and vice versa. In each trial, one picture had depicted the event with the argument structure corresponding to the one in the question the participants heard (henceforth “target”); the other depicted the reverse argument structure (“competitor”). After hearing the verbal stimulus and looking at the pictures, participants clicked on the picture that answered the question. Between-subjects variable was language group, i.e. monolinguals vs. bilinguals. This task examined subject and object wh-questions with the question word “which” as wh-questions headed by “which” have been found to be comparatively more difficult relative to questions headed by “who” or “what” (e.g. Avrutin, 2000). Moreover, the number of the two noun phrases was manipulated so that the two could be either the same or different (i.e. match vs. mismatch) following Contemori et al. (2018). The third within-subjects variable was the number of the first noun phrase (singular first vs plural first) which was not included in Contemori et al. This gave rise to 8 (2x2x2) conditions. For each condition, there were ten trials with the same words across conditions as can be seen in Table 3.

Table 3 Sample of experimental items per condition

Subject Question Singular-Singular	**Object Question Singular-Singular**
Which bear is chasing the camel?	Which bear is the camel chasing?

Subject Question Singular-Plural	Object Question Singular-Plural
Which bear is chasing the camels?	Which bear are the camels chasing?
Subject Question Plural-Singular	**Object Question Plural-Singular**
Which bears are chasing the camel?	Which bears is the camel chasing?
Subject Question Plural-Plural	**Object Question Plural-Plural**
Which bears are chasing the camels?	Which bears are the camels chasing?

This study did not use fillers. There are two reasons for this. First, due to the complexity of the research design, a relatively large number of trials was needed. To include the full experimental design in addition to filler trials would have made the task excessively long, tiresome for the children participants and would have impacted the size of the battery of background assessments of other experimental tasks administered. Reducing the number of conditions would not have been conducive to addressing open questions from the Contemori et al. (2018) study – in particular, the effect of the number of the first noun phrase. To reduce the number of trials in order to allow for fillers would have reduced statistical power. The latter is particularly problematic for studies using physiological measures such as gaze data where participants may vary significantly by trials (e.g. in this case, lack of concentration, particular interest in one specific stimulus etc). For the same reason multiple variants of the task in the form of a Latin square were not created. The second reason is that fillers are perhaps redundant in this task. Fillers are habitually used in order to prevent practise effects or effects of the participant building expectations about upcoming trials. This is unlikely to be the case in this study as knowledge about previous trials is not informative about upcoming ones for two reasons. First, all conditions cannot be distinguished without the participant hearing and processing the entire sentence. Second, the participants hear all eight conditions interchangeably, therefore any biases in expectation are cancelled out. To ensure that individual trials did not impact the results, analyses included random intercepts for trial. These were usually among the first to be removed when the model failed to converge suggesting that the participants tended to behave similarly across trials.

2.2.3 Materials

For each trial, one which-question and two pictures were used. 80 which-questions were created by forming which-questions with all ten lexical sets across all conditions. Each set of lexical items involved a transitive verb and two animate entities, more specifically animals. This way all sentences in the experimental trials were semantically reversible. The verbs used were the same as in Contemori et al. The verbal stimuli were digitally recorded by a male L1 speaker of British English in a sound booth. There were no fillers in this task due to time constraints.

The visual stimuli have been derived from Contemori et al. (2018) but additional pictures for the novel conditions (PL-SG, PL-PL for both subject and object questions) in this study were created by copying to ensure maximal visual similarity. The visual stimuli were two pictures both depicting the action in the question and involving the same animate entities. One picture corresponded to the answer of the question heard (target) whereas the other one did not (competitor). The competitor showed the same action and entities involved as the target but with the thematic roles reversed. The size and visual features of target and competitor were similar to the greatest degree possible.

An example of the visual stimuli where both NPs match in number (Singular) is shown in Figure 1. The picture on the right is the target for Subject SG-SG and competitor for Object SG-SG; the reverse is true for the picture on the right. For the trials with the mismatch condition, the singular and plural entity was the same in both pictures. Figure 2 shows the visual stimuli for the same set of item in the conditions where NPs differed in number;

condition Subject SG-PL and Object SG-PL. Again the target for the subject condition is the competitor for the object condition and vice versa. The position of target and competitor was counterbalanced across conditions. As a result, except for the structure of the sentence heard, there were no cues to adjudicate between target and competitor.

Figure 1 Sample visual stimuli for the example listed in the materials for conditions where both NPs are singular. The same pictures were used for both subject ("Which bear is chasing the camel?", target on the left side; competitor on the right) and object ques

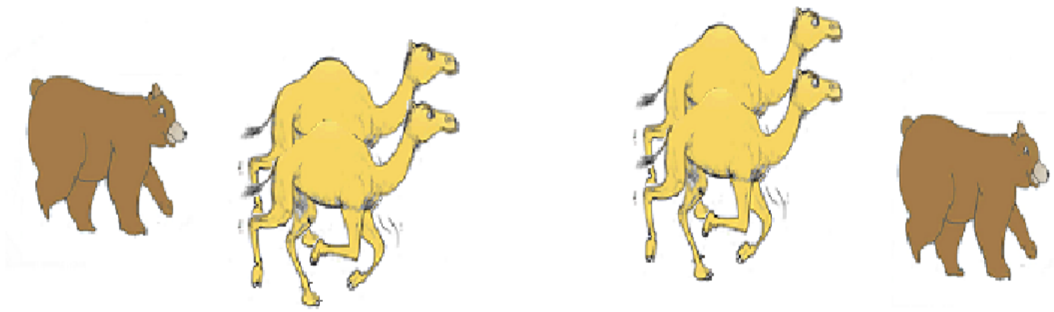

Figure 2 Sample visual stimuli for the example listed in the materials for conditions where the NPs differ in number. The same pictures were used for both subject (“Which bear is chasing the camels?”, target on the left side; competitor on the right) and object questions

2.2.4 Procedure

The study took part in a quiet room with the participants wearing headphones. A Tobii X120 (Tobii Technology AB, Sweden) eye-tracker was used to measure participants’ eye-gaze with a resolution of 120Hz. The remote Tobii eye-tracker is placed in front of a laptop, which is raised to a height of 30cm as it is placed on the eye-tracker case, and allows a freedom of head movement of 44 cm x 22 cm x 30cm. The eye-movement data reported are an average of both eyes. Stimulus presentation and eye-gaze data collection was conducted using E-prime (Schneider, Eschman & Zuccolotto, 2002). The testing started with a 5-point calibration procedure using Tobii Studio. The experimenter (first author) judged the quality of the calibration by examining the calibration plot for the five points. Participants sat on a chair at about 60 cm from the screen, although this was adjusted somewhat to facilitate calibration.

During the task, a fixation cross in the centre of the screen appeared before the onset of each trial which participants needed to fixate upon for 1000ms for the trial to begin. Participants heard a question over a set of headphones and saw two pictures on each side of the screen. Following the question, a question mark appeared, and a cursor appeared on screen. The two pictures were kept constant and the participants needed to click on a picture on the screen to select the target while looking at the pictures. There was no time limit for participants to select a picture. The order of the stimuli was pseudorandomised to avoid the same condition in adjacent trials which were split into two blocks of 40 trials. Total testing time for the children was about 10-15 minutes per block excluding the time needed to calibrate before starting.

2.2.5 Analyses

Accuracy, reaction time and gaze data were collected and analysed. Two areas of interest (AOI) were defined a priori in E-prime capturing the left and right half of the screen, corresponding to each picture presented in each trial. Eye-movement data were time locked to the onset of the auxiliary verb. A window of 200ms was allowed for the time it takes to program a saccadic eye-movement and as such any movement prior to this would not reflect linguistic processing related to the stimulus (Matin, Bao & Boff, 1993). Eye-movements were analysed for a period of 2 seconds (200-2200ms post auxiliary). Incorrect trials were removed from the analyses[15]. The time period examined was divided into ten equal bins of 200ms. For each bin the proportion of looks to target relative to target and competitor was calculated.

[15] Trials were originally analysed as a whole and then separately for correct and incorrect trials; as the patterns observed in the incorrect trials were inconsistent, these were assumed to be noise and as such excluded from further analyses.

These proportions of looks were quasi-logit transformed to compute the empirical logit which better handles cases where the proportion is extremely high or low (likelihood near 0 or 1) and where the observations are not purely independent from one another, as is the case with eye-movements (Barr, 2008). The decision to segment the observations into 200ms windows is due to the sampling frequency of the Tobii eye-tracker (120Hz, 1 measurement per 8.3 milliseconds). A 200ms time window meant the proportion was calculated as a ratio out of 25. This allows for numeric values with a distribution that does not approximate the binomial one while still allowing for relatively fine-grained analysis in terms of timecourse.

The analysis was conducted using logistic mixed effects for accuracy and linear mixed effects for reaction time and gaze data with crossed random effects for subjects and lexical items (Baayen, Davidson & Bates, 2008) implemented in the lme package in R (Bates, Maechler, Bolker & Walker, 2015, version 3.5-0). For the gaze data, a growth curve model was fitted to access changes over time (Mirman, Dixon & Magnuson, 2008; Mirman, 2014). The latter is a particular type of hierarchical mixed effects model where a polynomial is created for time and is used as a predictor variable to model time course of change in addition to other variables which may have an overall effect or an effect on the change over time. These polynomials are orthogonalised; they have been transformed so that they are independent from one another in order to avoid collinearity and as such their individual contribution to the change over time can be assessed. Using power polynomials has several advantages. First, it allows us to represent curvilinear change over time which is not exclusively linear. Second it avoids the need to segment the data into windows for separate analyses with reduced statistical power whilst reducing the risk of false positives due to multiple comparisons[16]. Third, allows for an interaction of time with the variables assumed to impact gaze data. Fourth the bias from the arbitrariness of establishing time windows is reduced (for a

[16] Although this may not be an issue if the effect is reliable across the majority of time bins (e.g. see Ito, Corley, & Pickering 2018 for a study with individual analyses for each 50ms time window).

discussion of this, see Mirman et al 2008). This type of analysis examines changes for each group and within subject variable by fitting multiple curves to the data, each corresponding to one of the orders of the polynomial terms, and accessing them as predictors of the data. It does not focus on the effect of time as a multilevel factor which would subsequently require pairwise comparisons but examines the trajectory of change over time. The intercept term reflects the average height of the curve, i.e. the height when all variables are zero, to phrase it differently the constant difference in height of the curve by condition. The linear order polynomial captures the general degree of change over time (i.e. slope) while the higher order polynomials capture the non-linear change (i.e. sensitivity to change at particular time points or revision which would be expected if the participants process incrementally and need to reanalyse). For this study, a fourth[17] order polynomial was fitted to the data as visual inspection of the data indicated a not purely linear change over time and which appeared to differ for the two types of syntactic structure. One should note, this is not the only way of analysing gaze data with or without including time as a variable (e.g. Ito et al., 2018). The motivation for choosing this type of analysis is that it follows the analysis in Contemori et al. (2018) and can also be applied to Studies 2A and 2B.

Sum coding (-1, 1) was used for between subject variables (monolingual vs. bilingual) and fixed main effects of 'structure' (subject vs. object which-question), 'number matching' (match between the two NPs vs. mismatch) as well as 'first NP' (singular vs. plural) for all three metrics. Time was scaled, and a fourth order orthogonal polynomial[18] was fitted in the range of bins in order to conduct the growth curve analysis. The dependent variable for the eye-tracking was the empirical logit transformation of the aforementioned ratio in order to reduce skew in the data (Barr, 2008). Trials where there were no looks to either target or object were not included, as the computed empirical logit value would be infinity (0 divided

[17] It was a better fit than models with fewer terms

[18] Contemori et al used a restricted cubic spline with four knots; this is a conceptually similar analysis with a different type of transformation of the data; it is not expected that this would impact the results

by zero) and were thus treated as missing data. Weights were added to each observation based on the reciprocal of the variance (i.e. 1/weights).

For all three previously listed metrics, the maximal model permitted by design that converged was used with correlation parameters removed (Barr, Levy, Scheepers & Tily, 2013). This included all dependent variables and by-subject and by-item random intercepts and slopes for all fixed effects. For the eye-tracking data, a single model was fitted for all data instead of multiple models, with bin (i.e. time) as an additional fixed effect. This resulted in each trial having multiple interdependent data points per trial. Therefore, a third random intercept was allowed, that of trial ID (the unique pairing of subject number and lexical item which defined a trial). When a model failed to converge, the random effects that accounted for the least variance were iteratively removed until the model converged.

2.3. Results

2.3.1. Accuracy

The mean accuracy by group for each condition is shown in Figure 3 and Table 4.

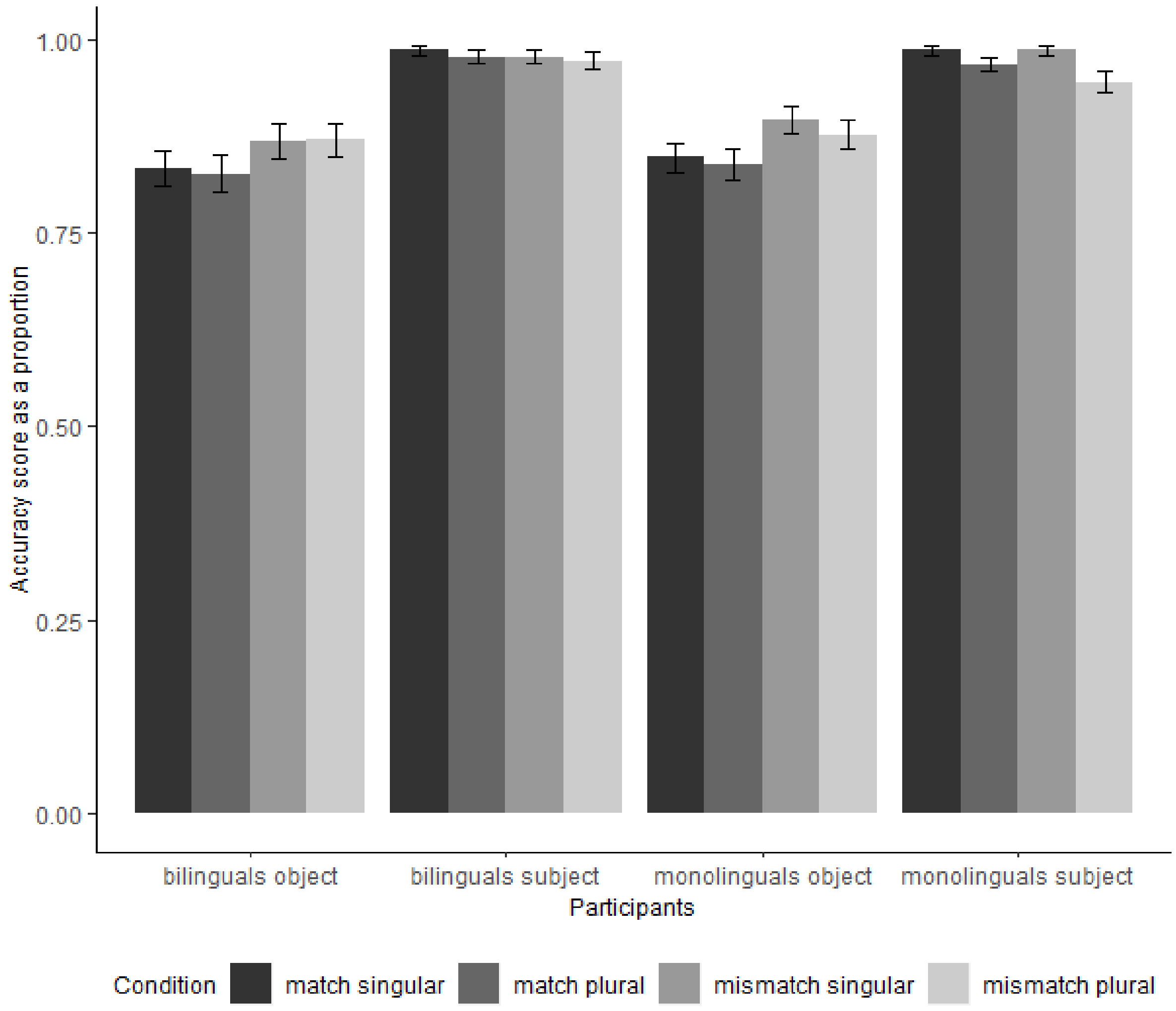

Figure 3 Mean accuracy as percentage (SE) by group and condition

Table 4 Overall accuracy by condition as a percentage (SD) for each group

	Monolinguals	Bilinguals
Subject questions – Number match – First NP Singular	97.9% (14.5)	98.5% (12.2)
Subject questions – Number match – First NP Plural	95.5% (20.8)	97.7% (14.9)
Subject questions – Number match	96.7% (18.0)	98.1% (13.6)
Subject questions – Number mismatch – First NP Singular	98.4% (12.6)	97.7% (14.9)
Subject questions – Number mismatch – First NP Plural	94.3% (23.2)	97.2% (16.6)
Subject questions – Number mismatch	96.6% (18.2)	97.5% (15.7)
All subject questions	96.6% (18.0)	97.8% (14.6)
Object questions – Number match – First NP Singular	82.3% (38.2)	83.3% (37.4)
Object questions – Number match – First NP Plural	81.9% (38.6)	82.6% (38.0)
Object questions – Number match	82.1% (38.4)	83.0% (37.6)
Object questions – Number mismatch – First NP Singular	87.7% (32.9)	86.8% (33.9)
Object questions – Number mismatch – First NP Plural	86.0% (34.7)	87.0% (33.7)
Object questions – Number mismatch	86.8% (33.9)	86.9% (33.8)
All object questions	84.3% (36.4)	84.8% (36.0)
All questions	90.5% (29.3)	91.4% (28.0)

There was a significant main effect of syntactic structure (β = 1.07, SE = 0.13, z = -7.93, p <.001) but no significant effects of group (β = -0.01, SE = 0.16, z = -0.45, p = 0.972) or number (β = 0.05, SE = 0.10, z = 0.46, p = 0.644). Accuracy was substantially lower for object which questions but reached near ceiling for subjects. Differences between the two groups were miniscule and not statistically significant. There was also a trend for an effect of First NP with higher accuracy for trials where the first NP was singular (β = -0.18, SE = 0.10, z = -1.76, p = 0.079) and a trend for an interaction of structure by number of first NP (β = 0.16, SE = 0.08, z = 1.89, p = 0.058).

Given that subject and object wh-questions have shown differential effects in previous studies (e.g. Contemori et al. 2018) and the trend for a structure by first NP number interaction, separate analyses where conducted for subject and object which-questions. For subject which questions there was a significant main effect of first NP number (β = -0.30, SE = 0.13, z = -2.263, p = 0.0234) which was absent in the object which questions (β = -0.03, SE = 0.09, z = -0.36, p = 0.717) driven predominantly by the monolinguals. Accuracy was lower for the subject which-questions with a plural first NP. On the other hand, there was an effect of number match in the object which questions (β = 0.21, SE = 0.10, z = 2.13, p = 0.033) which was not found in the subject which questions (β = -0.07, SE = 0.17, z = -0.387, p = 0.702). Accuracy was higher in the object which-questions when there was a number mismatch between two NPs but there was no difference whether the first NP was singular or plural.

2.3.2 Reaction times

Figure 4 and Table 5 provide a visual overview of the average reaction times by condition for each group.

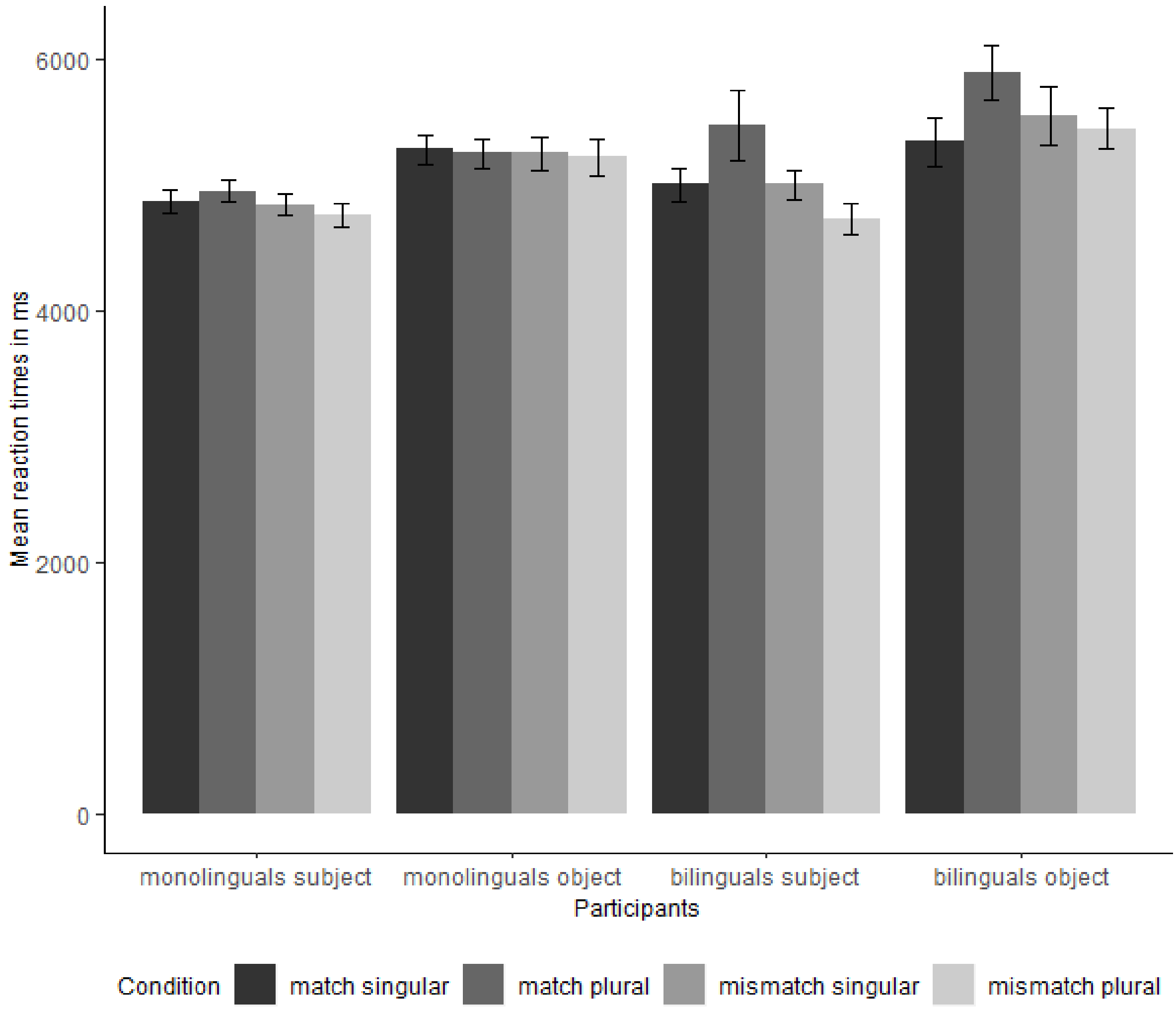

Figure 4 Average reaction times in ms (se) by group for each condition

Table 5 Mean reaction times in ms (SD) by condition for each group

	Monolinguals	Bilinguals
Subject questions – Number match – First NP Singular	4933 (1805)	4996 (2106)
Subject questions – Number match – First NP Plural	4958 (1601)	5465 (4507)
Subject questions – Number match	4945 (1707)	5229 (3515)
Subject questions – Number mismatch – First NP Singular	4906 (1524)	4996 (1943)
Subject questions – Number mismatch – First NP Plural	4787 (1460)	4728 (1833)
Subject questions – Number mismatch	4854 (1497)	4877 (1898)
All subject questions	4902 (1611)	5063 (2873)
Object questions – Number match – First NP Singular	5226 (1910)	5343 (2863)
Object questions – Number match – First NP Plural	5248 (1907)	5887 (3301)
Object questions – Number match	5237 (1907)	5615 (3098)
Object questions – Number mismatch – First NP Singular	5279 (2004)	5546 (3165)
Object questions – Number mismatch – First NP Plural	5253 (2278)	5440 (2365)
Object questions – Number mismatch	5266 (2150)	5490 (2768)
All object questions	5250 (2025)	5556 (2946)
All questions	5062 (1820)	5288 (2916)

There was a significant main effect of structure (β = 245.50, SE = 33.05, t = 7.43, $p < .001$) and a significant main effect of number match (β = -96.39, SE = 33.10, t = -2.91, p = 0.004) but no significant main effect of group (β = 182.71, SE = 152.32, z = 1.20, p = 0.235). Reaction times were shorter for subject than for object questions and shorter for trials where the number of the two noun phrases did not match. The statistically significant interactions were number match by number of first NP (β = -87.22, SE = 32.90, t = -2.65, p = 0.008) and the three-way interaction of group by number match by first NP number (β = -72.46, SE = 32.79, t = -2.21, p = 0.027).

The reaction time data were subsequently analysed by group individually to explore the differential impact of number match and first NP number for each group. In the monolingual children, there was only a main effect of structure (β = 189.42, SE = 61.42, z = 3.08, p = 0.011) as reaction times were on average longer for the object than for the subject questions. This difference was consistent regardless for the number match between the two NPs and/or the number of the first NP (see Figure 3). For the bilingual children, there was both a main effect of structure (β = 326.26, SE = 104.56, z = 3.12, p = 0.008) but also an interaction of number match by number of first NP (β = -168.41, SE = 75.73, z = -2.22, p = 0.037). As with the monolinguals, object which-questions had longer reaction times than subject questions. A visual inspection of the data suggests that a slowdown particularly with the plural number of the first NP when there is a number match between the two NPs and not when there is a mismatch. Conversely, the mismatch benefit is only present in questions where the first NP is in the singular. This, however, was entirely absent in the monolinguals (see Figure 3).

The bilingual data were further compared pairwise based on number of first NP. For the trials where the first noun phrase was in the singular, there was a main effect of structure only (β = 327.08, SE = 148.17, t = 2.21, p < .039). In the trials where the first NP was plural, there was both a main effect of structure (β = 289.56, SE = 113.16, t = 2.560, p = 0.030) but also a trend for number match (β = -275.39, SE = 124.79, t = -2.21, p = .053) with reaction times longer when the number of the two NPs matched. There was no significant structure by number match interaction in either subset (β = 14.35, SE = 74.44, z = 0.19, p = 0.849 for the first NP in the singular condition; β = 82.03, SE = 117.87, z = 0.70, p = 0.503 for the first NP in the plural condition).

2.3.3 Gaze data

We report main effects and interaction on the intercept, the linear and the quadratic term. The former reflects overall effects of any given condition irrespective of timecourse. The latter two reflect the impact of any condition on the change over time (i.e. is there a difference in the rate of the increase or in the nature of change over time for one level relative to the other). The main effects of the polynomial terms are also included as they describe the type of change over time in the data. The full overview of main effects and interactions on all polynomial terms is provided in the Appendix. Figures 5 and 6 provide an overview of the data by group and condition.

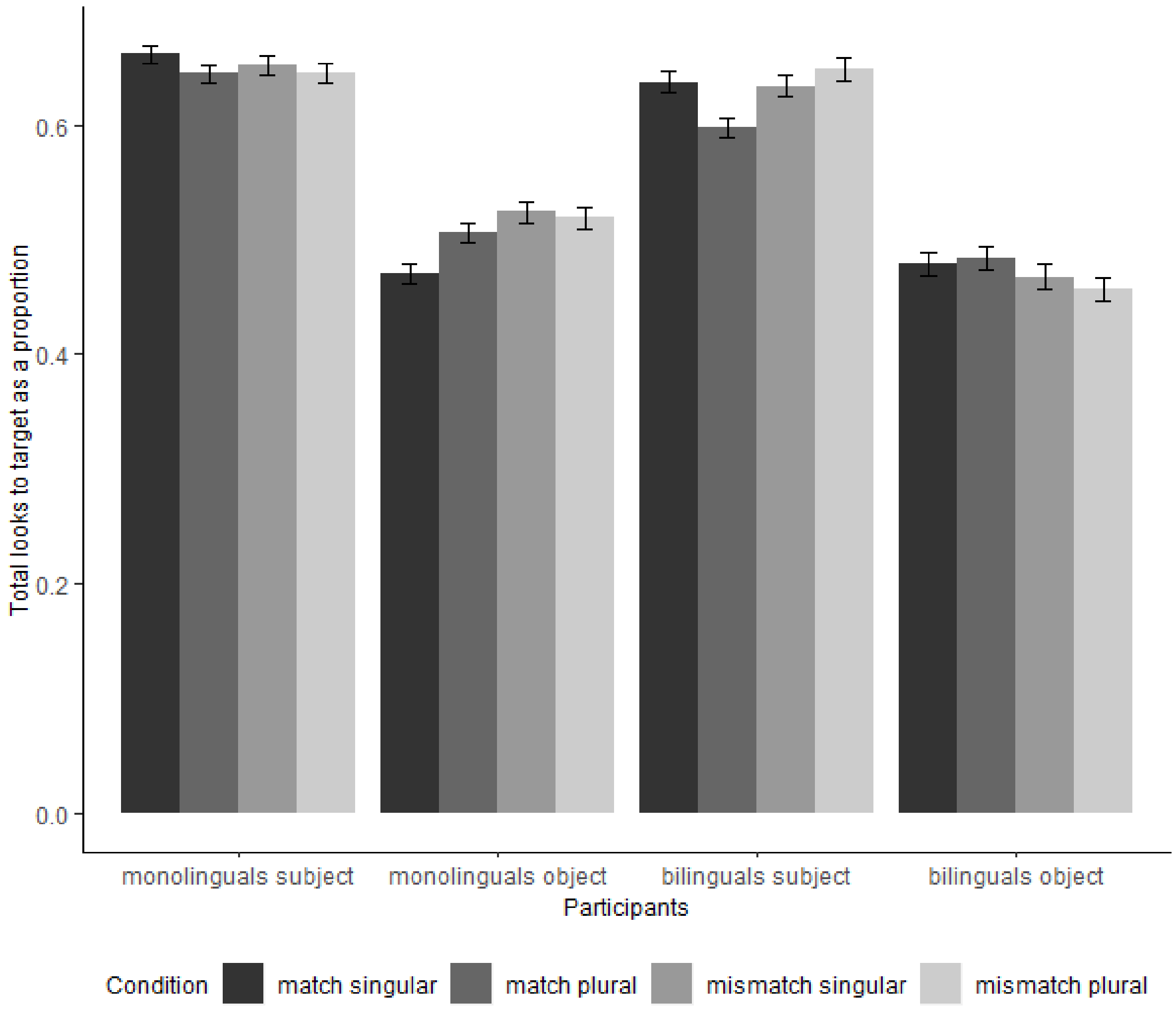

Figure 5 Mean looks to target picture overall as a proportion (SE) by group and condition for the time window examined (200-2200ms after onset of auxiliary)

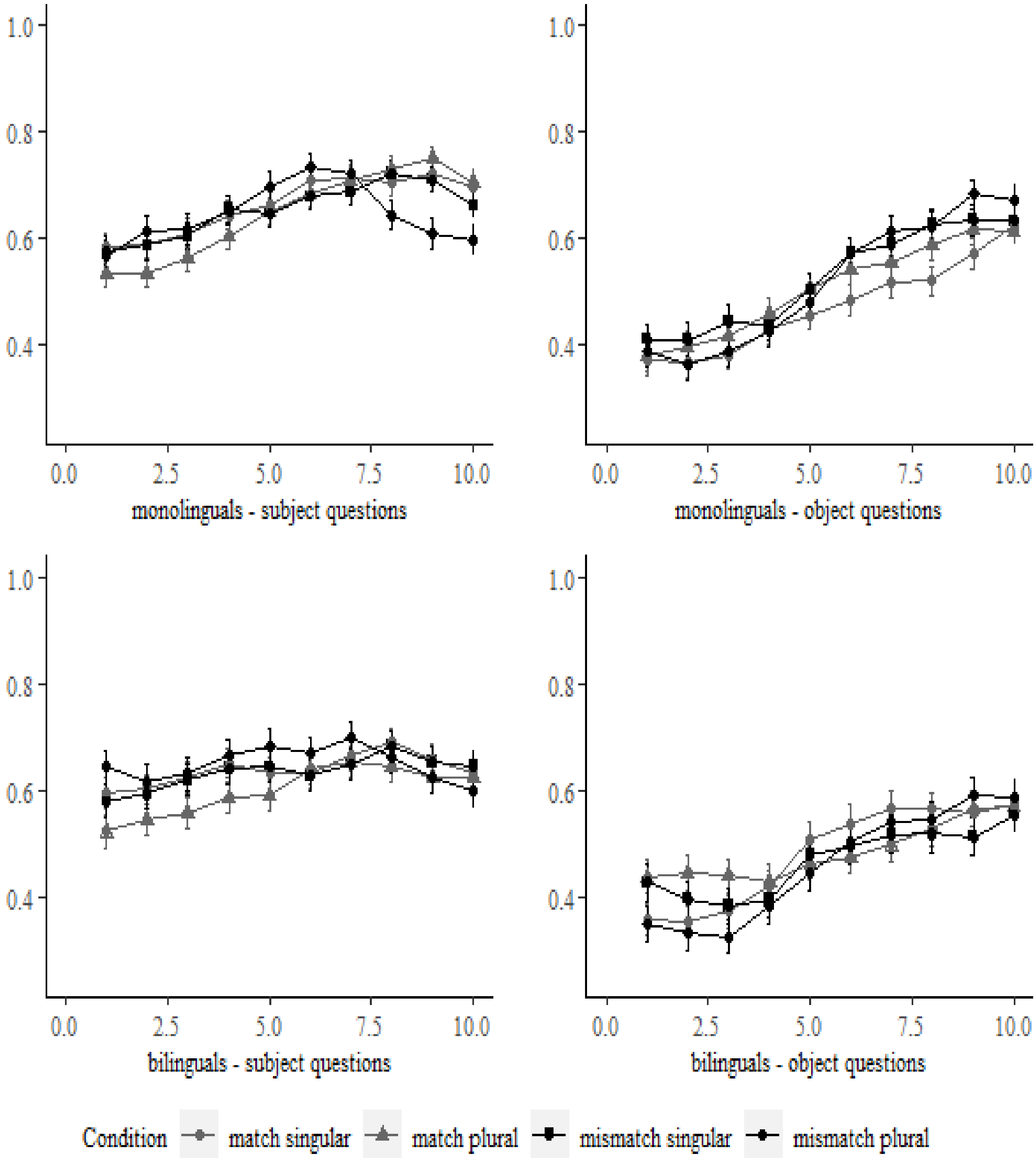

Figure 6 Looks to target picture over time as a proportion (SE) by group and condition for the time window examined (200-2200ms after onset of auxiliary)

The only significant main effect on looks on the intercept term was syntactic structure (β = -0.16, SE = 0.02, t = -8.92, p <0.001) but there was no overall main effect of group (β = -0.03, SE = 0.03, t = -0.95, p = 0.348). There were more looks towards the target relative to the competitor for the subject than for the object questions. The bilingual children looked at the correct picture over time as much as the monolingual children did. There was no significant effect of number match (β = 0.00, SE = 0.02 t = 0.03, p = 0.979) or First NP number. The main effects on the intercept term are shown in Figure 5.

The linear and cubic polynomial terms were statistically significant (β = 26.38, SE = 0.99, t = 26.63, p < 0.001 for the linear term; β = -3.29, SE = 0.99, t = -3.33, p = 0.001 for the cubic term) while there was a trend for the quartic term as well (β = 1.89, SE = 0.97, t = 1.92, p = 0.055). The quadratic term did not reach statistical significance (β = 0.15, SE = 0.98, t = 0.156, p = 0.876). The significance of the linear term as a predictor of participants' looks to target suggests that looks towards the target relative to the competitor increase over time. The significance of the cubic and trend for the quartic term indicate this increase does not happen in a purely linear manner. Looks to the target picture by group and condition over time are shown in subsequent graphs when reporting specific effects.

There was a significant main effect of group (β = -7.10, SE = 0. 99, t = -7.17, p <. 0.001) and structure on the linear term (β = 7.33, SE = 0.99, t = 7.40, p < 0.001). The effect of group on the particular term indicates that the increase in looks towards the correct picture was different across the two groups with a sharper increase in the monolinguals and a more protracted one in the bilinguals although both groups shows an initial decrease in looks in the object conditions reflecting misinterpretation of the wh-phrase as the subject (see Figure 6).

This can be interpreted as slower processing for the bilinguals who require more time to orient their gaze towards the correct picture. The effect of structure was significant on all four polynomial terms (β = 7.33, SE = 0.99, t = 7.40, p 0<.001 for linear term; β = 2.34, SE = 0.98, t = 2.38, p = 0.018 for quadratic term; β = -2.53, SE = 0.99, t = -2.57, p = 0.010 for cubic term and β = 3.399, SE = 0.97, t = 3.45, p = 0.001 for quartic term). This reflects a qualitatively different pattern of change for the two types of question; the crucial difference is that looks to the correct picture actually decline for object questions (reflecting the original subject bias-driven misanalysis) and then increase more slowly (see Figure 7).

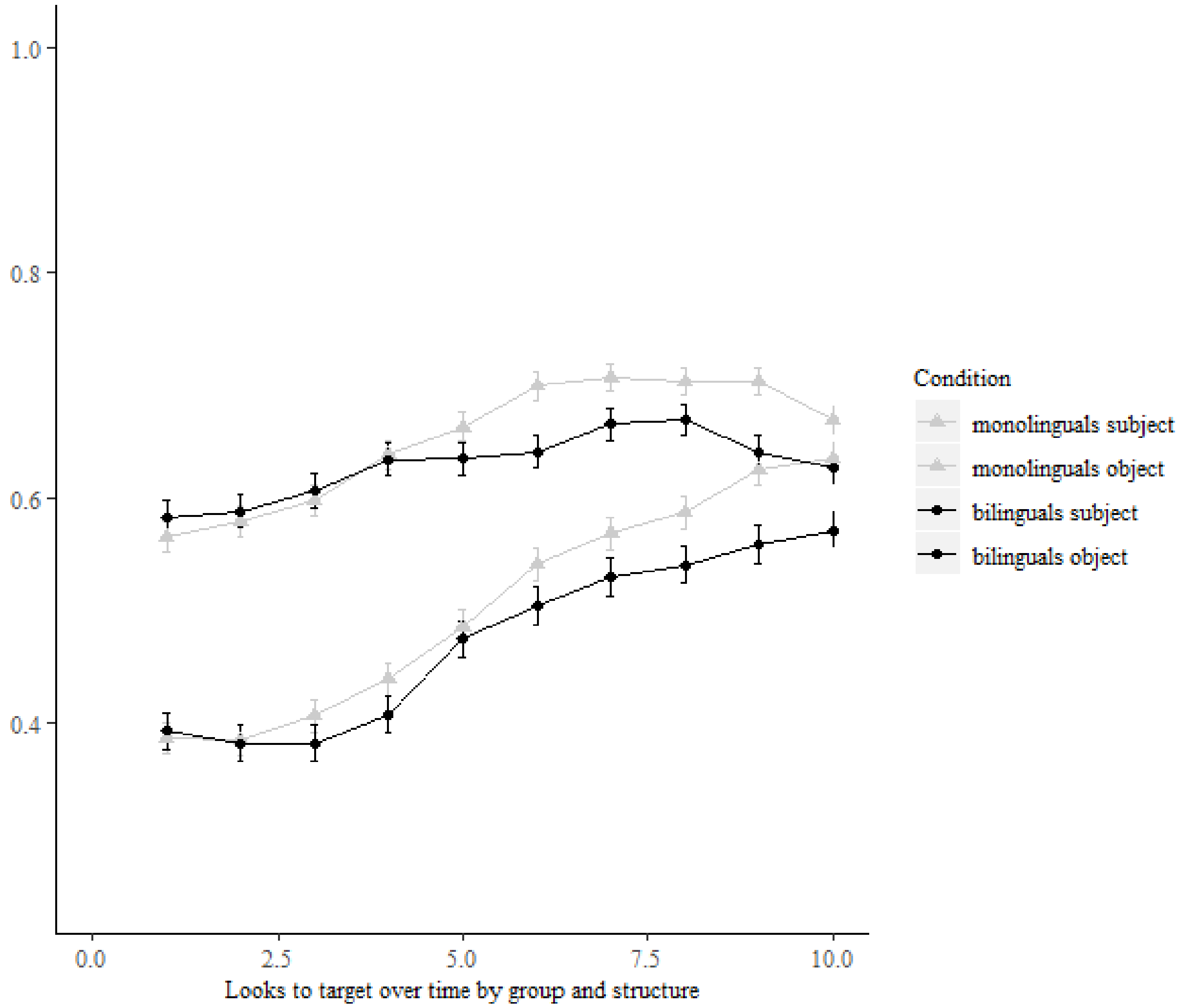

Figure 7 Looks to target picture over time as ratio (SE) by group and structure

On the linear term, which reflects the changes over time, there was a group by number match interaction (β = 2.20, SE = 0.98, t = 2.24, p = 0.025) as well as a three-way interaction of structure, number match and first NP number. This suggests that number matching impacts the increase in the looks towards the target differently for the two groups. Figure 8 shows that a mismatch in number leads to a faster increase in looks to target for the monolinguals but not for the bilinguals and shows a more rapid increase in looks to the correct picture for the former relative to the latter.

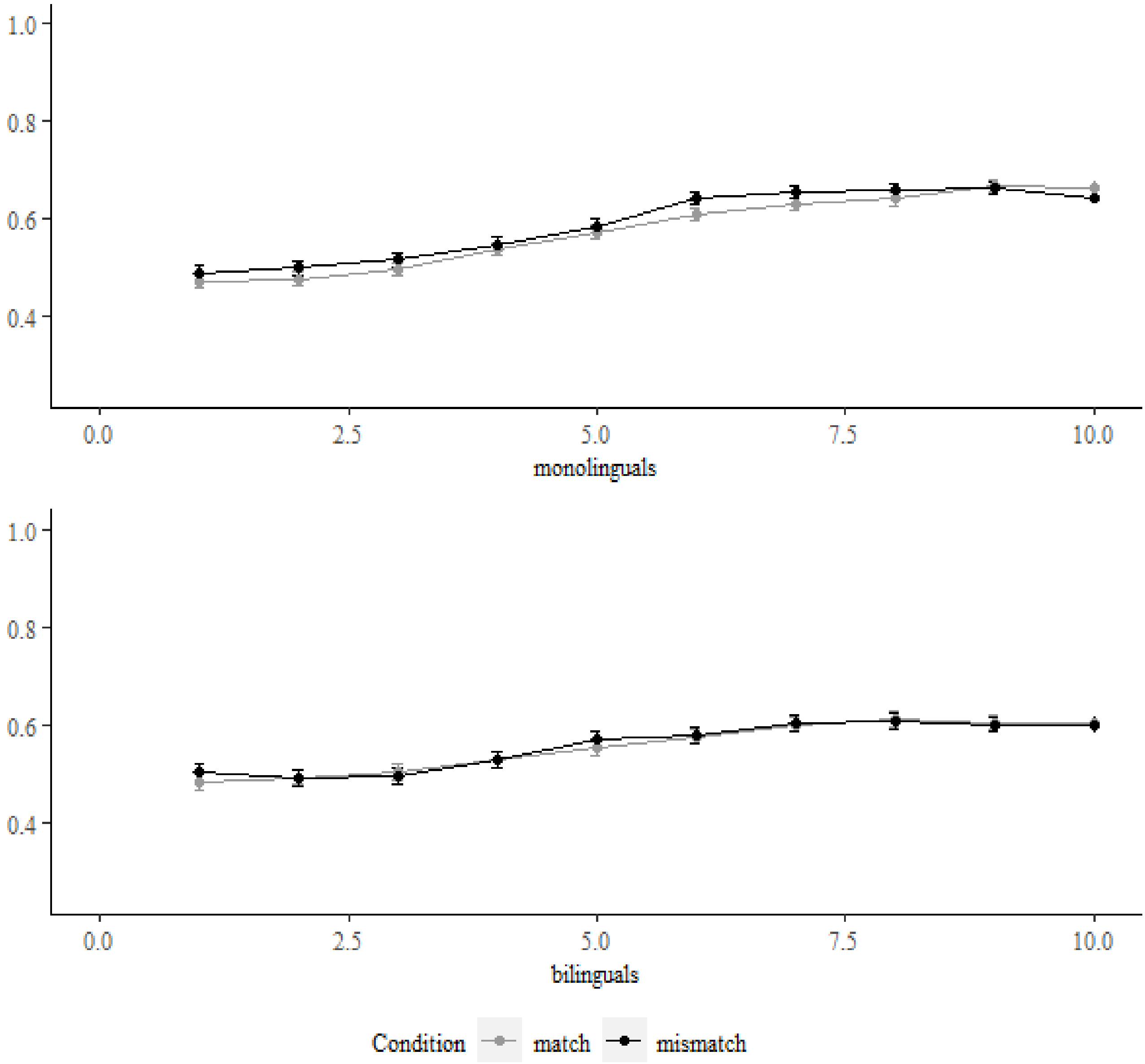

Figure 8 Looks to target over time by group and number (mis)match separately

Given the aforementioned interaction and the differential patterns between subject and object wh-questions, including for gaze data (Contemori et al. 2018), the gaze data for subject and object questions were analysed separately. The model for the subject only questions showed a main effect of the linear and quadratic term ($\beta = 15.04$, $SE = 0.98$, $t = 15.30$, $p < 0.001$ and $\beta = -2.04$, $SE = 0.98$, $t = -2.08$, $p = 0.038$ respectively) suggesting an increase in looks to target over time. There was no main effect of group on the intercept term, but the effect was significant on the linear term ($\beta = -0.04$, $SE = 0.035$, $t = -1.185$, $p = 0.24$ and $\beta = -5.78$, $SE = 0.99$, $t = -5.88$, $p < 0.001$ respectively). This suggests both groups looked to the target equally but that the increase was faster for the monolinguals (also, see Figure 7). There were no other main effects or interactions which were significant on the intercept or linear term.

The object questions model showed somewhat different results. There was an interaction of group by number on the intercept term ($\beta = -.034$, $SE = 0.01$, $t = -2.32$, $p = 0.025$). Figure 8 illustrates this interaction.

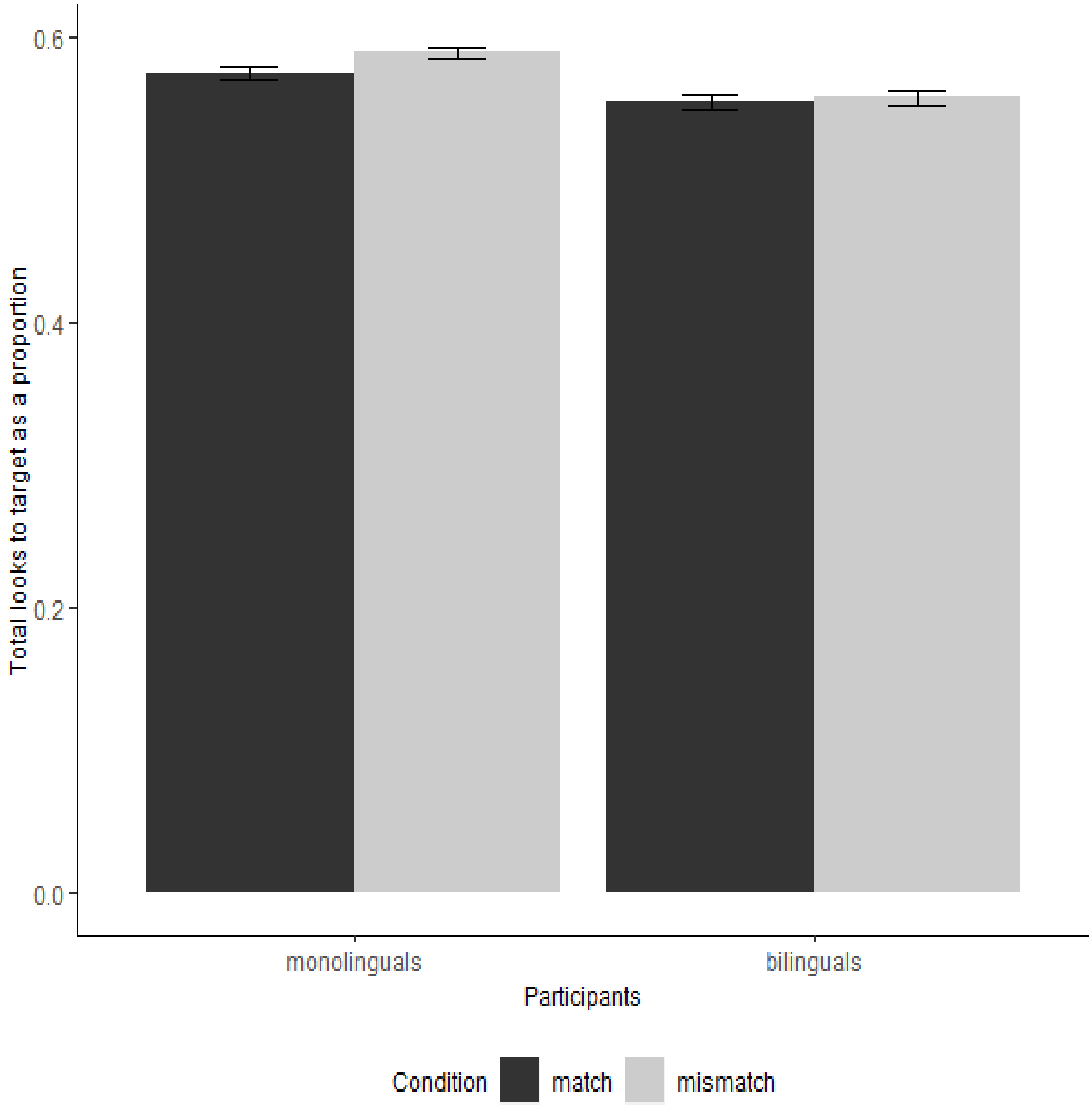

Figure 9 Looks to target picture overall as ratio (SE) by group and number match across the time window examined

This suggests that only the monolinguals looked significantly more to the target when the number of the two NPs was different. Similarly to subject questions, the first, second and fourth order polynomials were significant predictors of participants' looks reflecting the fact the looks increase over time but not purely linearly (linear term: β = 24.96, SE = 1.11, t = 22.54, $p < .0001$, quadratic term: β = -4.49, SE = 1.10, t = -4.08, $p < 0.001$ and quartic term:

β = 3.56, SE = 1.10, t = 3.23, p = 0.002). On the linear term, there was a significant effect of group, a number match by first NP interaction (β = 3.18, SE = 1.11, t = 2.86, p = 0.004) and a trend for a three-way interaction of group by number match by first NP number interaction (β = 1.89, SE = 1.11, t = 1.70, p = 0.089).

As the effect of number match was different for the two groups, the object question data from the two groups were analysed separately. For the monolingual participants, the linear, cubic and quartic terms were significant as looks to target increased over time (β = 22.21, SE = 1.07, t = 20.70, p < 0.001, β = -3.36, SE = 1.07, t = -3.15, p = 0.002 and β = 3.16, SE = 1.07, t = 2.97, p = 0.003). There was a main effect of number match on the intercept term as the monolingual children looked more towards the target picture in the number mismatch condition (β = 0.05, SE = 0.02, t = 2.80, p = 0.009, consistent with Figure 7). However, this effect was not significant on the linear term suggesting the number mismatch did not significantly impact the increase over time for the monolinguals only (β = -0.08, SE = 1.07, t = -0.08, p = 0.938).

For the respective bilingual dataset, the results were different. The two polynomial terms were significant as with the monolinguals (β = 13.40, SE = 1.12, t = 11.97, p < 0.001, β = -3.04, SE = 1.12, t = -2.72, p = 0.007, β = 1.91, SE = 1.16, t = 1.72, p = 0.087 for the linear, cubic and quadratic term respectively for which there was a trend) suggesting a similar pattern of increase in looks over time. Number match trended towards significance on the first order polynomial but unlike for the monolingual children it was not significant on the intercept term (β = -2.13, SE = 1.12, t = -1.89, p = 0.058 and β = -1.58, SE = 0.03, t = -0.62, p = 0.541). The aforementioned trend indicates a potential yet much weaker effect in the bilinguals than in the monolinguals.

To sum up the results, monolingual and bilingual children looked at the correct picture equally as much. For both groups, looks to target increased over time but more strongly so for the monolinguals. There were significant differences between subject and object questions irrespective of participant group with looks to target being significantly fewer for object questions than for subjects. The effect of structure was present for all time terms. Effects of first NP number were not found on the intercept or linear terms in the gaze data while the effects of number match were also inconsistent and limited to the monolingual group for gaze data and the bilingual group for the reaction times.

A significant overall interaction of structure, number match between NPs and first NP number prompted further analysis by splitting the dataset by structure. Analyses of the subject only questions showed no overall main effects but effects on particular time terms. No overall main effects or interactions were found. In contrast, there was an overall interaction of group by number for object questions with a similar effect of polynomial time terms as with subject questions. In subsequent pairwise analyses by group, there was an overall effect of number match for the monolinguals with looks to target being higher when there was a mismatch in number of the two NPs which was absent in the bilinguals.

2.4. Discussion

This study investigated the online processing of which-questions in bilingual children relative to monolingual children's using a visual-world paradigm task. Subject which-questions were compared to object which-questions. The number of the two NPs was manipulated to be either a match or a mismatch as with Contemori et al. 2018 although in addition to the paper in question the number of first NP was also manipulated. To this end, three metrics were used as dependent variables; the overall accuracy (offline), reaction times and gaze date during listening (both offline). The results differ across metrics, but certain common themes emerged.

RQ1: Do bilingual children differ in their performance to monolinguals in processing which-questions in English on a range of measures reflecting the end result?

There were no main effects of group on any of the measures adopted and none of the metrics suggested the bilingual children underperformed. Bilinguals did not differ significantly from the monolinguals in terms of accuracy. In fact, they were practically indistinguishable to monolinguals in terms of accuracy. Overall accuracy scores were at 90.5% and 91.4% for monolinguals and bilinguals respectively while for accuracy for object questions was at 86.8% and 86.9% respectively. There were also no group differences between monolinguals and bilinguals in terms of reaction times. Bilinguals did not underperform monolinguals in the gaze data either as looks to the correct picture relative to the competitor were as many as for the monolinguals. The only effect on accuracy in this study was that of structure with accuracy being lower for the object questions than for the subject ones. Effects on accuracy are usually interpreted as a sign that wrong interpretations may still linger and have been found even for adults (Slattery et al., 2013).

The bilingual children who were included in this study had been exposed to English for a minimum of two years and attended mainstream school in the UK without any particular difficulties reported either by parents in the questionnaires provided or by educators in recruitments. The participants are, therefore, highly proficient in English and that these results reported in this study may not be replicable with bilingual children with more limited proficiency in English. This is indeed true but in that case, any differences in processing would be confounded with differences in proficiency. This would be a potential further step in research; however, it would require obtaining data from the more proficient bilingual children initially as a benchmark. Moreover, to test children in the age range of 8-11 years with more limited proficiency in English would entail testing essentially L2 learners of the language which could yield effects of exposure to when comparing them to monolinguals.

RQ2: Does the timecourse of processing differ between bilingual and monolingual children? Is there evidence of incrementality in processing in bilingual children?

The gaze data provides an insight into the timecourse of which-question processing in bilingual and monolingual children. There was no overall effect of group on the intercept term as the two groups looked at the target equally. Yet there was a significant effect of group on the first order polynomial suggesting a steeper increase for the monolinguals. We focus on the first order term of the polynomial as it is the clearest in its significance and the most meaningful to interpret as it is a measure of linear change over time and hypothesize the effect of group found reflects faster processing in the monolingual children. The linear order polynomial was overall significant suggesting an increase in the looks of both groups towards the target picture over time. This can indeed be confirmed by visual inspection of the data and is to be expected. The main effect of group on the linear polynomial term indicates that the differences in curvature between the two groups were significant; the increase in looks to target was namely less steep in the bilinguals. This in turn can be interpreted as slower but not less accurate processing.

However, the third order of the polynomial term was also found to be significant with a trend for the fourth as well suggesting that while looks to target increase overall over time, linearity alone does not capture the full trajectory of the data. The different timecourse of looks to target for the object-questions may account for this. Whereas for subject-questions there is a generally linear increase in looks, in objects there is an initial decrease and then a gradual increase. There are more looks to the competitor as the ratio is below 0.5 and this begins to briefly decline. Looks to target begin to increase after about three 200ms time bins and then follow a sharper increase over time than the subject questions. This increase continues even at the last time point albeit less steep suggesting that participants continue to try to process the questions. This was also found in Contemori et al. (2018, see p. 430, Fig.2) and reflects an initial garden-pathing of the children who interpret the first NP as the subject of the question instead of the object and are then forced to reinterpret the question after the auxiliary and second NP. This substantially different timecourse is reflected in the fact that the effect of structure is significant on all polynomial terms. Despite a sharp increase in looks to target following disambiguation, looks to target for object-questions remain below those for subjects. This mirrors the general increased difficulty posed by object-extracted filler-gap dependencies observed in the literature for numerous measures.

This pattern was found for both monolinguals and bilinguals. The absence of an interaction of group by structure either on the intercept or on any of the polynomial terms suggests that even though the bilinguals may be slower at processing the questions, the process of reanalysis for object questions follows a similar timecourse for the bilingual and monolingual children and that the additional burden of the more difficult type of which-question does not translate into a further slowdown of processing for the bilinguals. Language processing in bilingual children is rapid and incremental as it is in monolinguals and in adults. They begin to process language with the onset of a sentence and form a syntactic representation of what they hear with the material available. Consequently there is the possibility of a wrong interpretation needing revision upon encountering information which does not support the initial structural representation.

Probabilistic accounts (e.g. Levy, 2008) argue that the difficulty in processing object wh-questions lies in the symmetry of frequency of their occurrence relative to subject wh-questions. That is, speakers of English have come to associate the first NP with the thematic role of agent[19] and with the syntactic position of subject of verb and to expect all which-questions to be subject questions. Processing cost is incurred when these expectations are not met. The results from this study support this view as the gaze data shows an initial misinterpretation of object questions as subject ones but with the caveat that subject wh-questions are indeed more frequent than object wh-questions in the bilinguals' exposure to English and this is an important prediction for probabilistic accounts. To address this uncertainty, appropriate corpora need to be built. What is plausible in the bilingual children is that any expectations formed for English are done so on the basis of quantitatively little input which is adequate for them to acquire which-questions.

The results of this study are also consistent with the Active Filler Hypothesis; as speakers of a language will attempt to place the filler in the earliest gap available (Frazier & Clifton, 1989), the dislocated[20] wh-phrase will be put in a potential gap before the verb and hence the sentence will be interpreted as a subject- rather than an object question.

RQ3: Do bilingual children make use of morphological cues to facilitate processing of which-questions in the same way as monolingual children do?

Contemori et al. tested whether the mismatch in number between the two NPs would aid disambiguation of the ambiguous object which-question. An overall effect of number match was shown with the two NPs having a different number being beneficial and specific to object questions in that looks to the correct picture increased faster at the point of

[19] However, the bias to interpret the first NP in a sentence needed be the result of frequency; Gertner, Fisher and Eisengar (2006) have shown a bias in the same direction even for novel verbs.
[20] Assuming movement from Spec-TP to Spec-CP

reinterpretation of the which-question. The results from this study are broadly in line with those of Contemori et al. but differ on specific metrics.

Unlike Contemori et al., there was no main effect of number match on accuracy[21], nor was there an interaction with any other independent variable. This may be due to the fact that the children included in this study were older than in Contemori et al. (i.e. 8-11 years as opposed to 5-8). As such their overall accuracy is expected to be higher[22] and it is harder to tease apart within subject differences due to limited variance but also due to the fact the older children may have been able to overcome their original misinterpretation by the time they selected the correct picture. In this study, there was no overall main effect of number mismatch in terms of gaze data but a more nuanced picture that was specific to the monolinguals. The effect of number (mis)match was not as large in this study as in Contemori et al.; consistent with the aforementioned study, it was absent in the subject questions and found exclusively in the object questions where looks to the correct picture was higher when the two NPs had a different number. This benefit however did not extent to the bilingual children where there was little difference between the two conditions.

The effect of number of number (mis)match between the two NPs is further supported by the reaction time data in a similar pattern is observed for the bilinguals as is with the gaze data for the monolinguals. Overall reaction times were faster when there was a number mismatch between the two NPs, i.e. between the first NP and the auxiliary verb. This effect differed across group depending on the number of the first NP. For the monolingual children, there was no effect of number mismatch regardless of the number of first NP and/or type of wh-question. For the bilinguals, number mismatch interacted with number of the first NP. That

[21] Although this was driven almost exclusively from the effect of number mismatch in the object questions as evidenced by the structure by number match interaction.
[22] This is indeed the case in particular for the object number mismatch condition; there were near ceiling effects for the monolingual children in this study and the younger children in Contemori et al.

means the benefit of number mismatch were particular to trials where the first NP was plural but absent if it was singular. Reaction times were also shorter for object which-questions numerically in the mismatch condition for the bilingual-only data, but that interaction did not reach significance.

In this sense, the data from this study lend partial support to Relativized Minimality, mainly for the monolingual children although only modestly as the effects are overall weaker in comparison to Contemori et al. A plausible explanation is the fact that there is little need for it. As the children are older, their language processing capacities are enhanced and more like those of adults relative to the younger children in Contemori et al. This being the case, they will have less difficulty processing which-questions and the benefit of number mismatch between the two NPs confers limited benefit. This is supported by the fact that the bilingual children benefit from the number mismatch in the reaction time data for the sentences with a first NP in the plural condition which is the marked option. However, eliminating one NP as a potential filler due to morphosyntactic cues under Relativized Minimality per se may not be the only plausible explanation for the effect of number mismatch on processing. For the sentences where the number of the two NPs is different, disambiguation is feasible earlier than for the sentences where the number of the two NPs is different, i.e. at the auxiliary verb vs. the second NP. It is not feasible to disentangle the impact of these two factors in the current experimental design.

The absence of number match effects in the gaze data for the bilingual children and the absence of the same effects in the reaction time for the monolingual children may be two sides of the same coin. These effects may be absent in the gaze data for the bilingual children as their processing of wh-questions may be overall slower and as such may cancel out any within-subject effects. Unsure about the correct syntactic representation, the bilingual children may have looked at both pictures more equally for a longer period of time until they are more certain. Alternatively, the bilingual children may have simply been more conservative and adopted a more exploratory[23] strategy of looking at the pictures more equally than the monolingual children. This may have attenuated or delayed some within-subject differences. The latter may in turn, not have yielded group differences on the intercept term, i.e. overall, but would lead to delayed effects or effects that are stronger in the later bins which would result in a different course of change over time; i.e. effects on the polynomial terms. However, it is reasonable to expect that this would spill over into the reaction times as the bilingual children may take some time after hearing the entire question before answering. Although the effect appears in different metrics for each group, the direction of the effect is the same. The data hence suggest that both monolingual and bilingual children make use of disambiguating cues in a similar manner and that the mismatch in number agreement between first NP and auxiliary can be used to recover from the initially wrong syntactic representation of object which-question, i.e. their misinterpretation as subject questions. The challenge for Relativized Minimality for the bilingual children is not in the different metrics where the effect of number mismatch is found but that the effect is, in fact, found across both question types. Relativized Minimality would predict an interaction with structure and number match as the effect would only be present in object questions where there is an intervener that would be act as a potential filler.

The results are also somewhat different those of Roesch & Chondrogianni (2016) in terms of the effect of number (mis)match for bilingual children. In Roesch & Chondrogianni, the simultaneous bilingual children showed reduced use of disambiguating cues relative to the monolinguals while the early sequential ones did not show any. In this study effects were found for the monolinguals in the gaze data and the bilinguals in the reaction times. In this

[23] This may explain the numerically lower ratio of looks towards target for the subject questions in the bilingual even though it is above chance performance and the higher ratio for the object questions in the first bins (less steep garden-pathing) for the object questions.

sense, it is not straightforward to characterise the effect in the bilinguals as "reduced" or "enhanced" relative to the monolinguals. There are, however, two differences between this study and Roesch & Chondrogianni; the nature of the disambiguating cue tested (nominal case in German vs. number in English) and the nature of the participants (in the present study they were older and had been exposed to the additional language for longer, i.e. 7-8 years on average, two thirds of them reportedly at an earlier age than those in Roesch & Chondrogianni). A second point of divergence of the results from this study with those from Roesch & Chondrogianni is that the differences in the latter between the two groups of bilinguals (simultaneous and early sequential) extend to overall accuracy. The sample size of this study makes splitting this group into subgroups of bilinguals challenging for purposes of by-group statistical comparison. To address this limitation, a correlation analysis was run for participant's overall accuracy, the measure in this study most comparable to Roesch & Chondrogianni, and their age of onset and length of exposure. Age of onset was not significantly correlated with accuracy ($\rho = 0.007$, $p = 0.923$); the same held true for length of exposure ($\rho = 0.072$, $p = 0.308$). One way to reconcile the differences between the two studies is the quantity of input the children have received. Given the demographics, the participants in the present study have received more input than those in Roesch & Chondrogianni. It is plausible that a certain quantity of input functions as a threshold beyond which additional input does not have an impact. This seems to be the case for accuracy in the present study, as evidenced by the essentially non-existent correlations outlined earlier and may be extendable to how participants utilise disambiguating cues.

The data are also possibly consistent with cue-based retrieval models of language processing (e.g. Gordon et al., 2004, Van Dyke & McElree, 2006). These predict that dissimilarity in the number of NPs will facilitate comprehension, or, alternatively processing is hindered when the features of disambiguating cues become associated with multiple items in memory with a matching feature (e.g. plural number), which can no longer act as facilitatory cues. More specifically, however, they predict greater difficulty in processing at the filler position. This has been supported by research which has used reading times as evidence in self-paced reading tasks (e.g. Van Dyke & McElree, 2006). The measures used in this study cannot test this; however, as the difference in looks to target between match and mismatch condition becomes more prominent towards the end of the time frame analysed indicating that the benefits of number mismatch could be observed around the gap position. A further complication is that the verb which would trigger retrieval in this study is not marked for number. What is marked for number is the auxiliary verb. It is not clear how, or if, the auxiliary verb would facilitate retrieval triggered at the main verb.

RQ4: Does the number of the first noun phrase have an impact on the processing of which-questions?

One of the limitations of the Contemori et al. study was that the mismatch condition always involved a first NP that was singular and a plural second NP. In the current study, the number of the first NP was counterbalanced allowing for a plural first NP and a singular second NP. For accuracy, there was a trend for the structure by first NP number interaction indicating the number of the first NP has a differential impact depending on the structure; the number of first NP was significant in the subject questions and accuracy was higher when the first NP was in the singular than when it was in the plural, but there was no effect in the object questions with the main effect for all data trending towards significance. This finding can be explained on the basis of linguistic markedness as the plural is the marked form (Haspelmath, 2005) and is expected to be more difficult (Harley & Ritter, 2002). The fact it is absent in the object questions may be related to another form of markedness; namely, that the object questions are the more marked grammatically and are, hence, a more difficult feature relative to subject questions, something which is found consistently in the literature for language processing in English. This would lead to overall lower accuracy which would mask the effect of first NP number, arguable a feature of less difficulty relative to object questions.

There is little to no evidence that number of the first NP had an impact on the other two measures. For reaction times, the effect was only present in the bilinguals. More specifically, the presence of a plural first NP lead to longer reaction times in the number match condition relative to when the first NP was in the singular in the same condition and the number mismatch condition regardless of first NP number. This suggests the effect of markedness was present only in the condition which was more difficult (number match). Once there were other cues to aid disambiguation, its role was redundant. There was no main effect of first NP number or interactions with group, structure or number match in the gaze data for either the intercept or any polynomial term.

2.5. Conclusion

This aim of this study was to examine the online processing of which-questions in bilingual children and compare them to that of monolingual children. It builds on a recent study on monolingual children (Contemori et al. 2018) but also expands the experimental paradigm to control for the number of first NP whilst also expanding recent findings to older monolingual children who were used as the age-matched control for the bilinguals. A visual world paradigm with picture selection was used where type of which question (subject vs. object), number of the two NPs (match vs mismatch) and number of the first NP (singular vs plural) was manipulated.

The results show that bilingual children did not underperform monolingual children in terms of overall accuracy; overall reaction times did not differ between the groups and bilingual children looked to the correct picture about as much as the monolingual children over a 2 second period after hearing the auxiliary verb. The difference between the two groups was observed in the timecourse as evidenced by the effects of group on the linear terms in the gaze data, i.e. shallower curvature in the trajectory of the looks over time for the bilinguals which would signal slower processing. However, these differences are found on a fine-grained timescale with the end result not being different to monolingual children. Qualitatively, the bilingual children do not differ from the monolinguals. They have greater difficulty with object relative to subject questions in the same way as monolingual children, as evidenced by the absence of interactions between group and structure. Moreover, the same factors or facilitative features (i.e. number matching between NPs) which have been shown in previous studies to impact language processing in monolingual children had a reduced and timewise delayed impact in bilingual children suggesting slower but similar processing mechanisms.

Chapter 3 Wh-questions in bilingual adults

3.1. Introduction

This chapter presents Study 1B. It is the same experimental task as in Study 1A but with monolingual and bilingual adults. This chapter begins by outlining recent research in filler-gap dependency processing as well as theoretical accounts which have motivated them or are informed by them. Due to the larger body of evidence available for bilingual adults relative to children, work on monolingual adults is generally not included with exception to monolingual adult controls in studies on bilinguals.

Filler-gap processing in bilingual adults / L2 Learners.

There is a growing body of works on bilingual language processing in bilingual adults. The latter has focused on differences in processing between native speakers and non-native speakers. As discussed in Chapter 1, one of the most influential theories of bilingual language processing posits that non-native speakers are less sensitive to grammatical information than native speakers and more reliant on lexical or pragmatic cues in processing (Shallow Structure hypothesis, Clahsen & Felser, 2006a; 2006b). Another account which predicts qualitative differences in processing is the Interface hypothesis (Sorace, 2011) according to which the difficulty for non-native speakers relative to natives is at the interface of syntax and other cognitive domains; it is predicted that the challenging linguistic features for non-natives will be those where information beyond, and in addition to, morphosyntax need to be integrated in real time. Alternative accounts assume that native and non-native processing are qualitatively similar in that both groups can utilise the same type of information during processing but less efficient in that it is more burdensome on the non-native speakers' cognitive resources (MacDonald, 2006; Hopp, 2006; 2010).

Filler-gap dependencies have been widely used in studies on bilingual processing (e.g. Frenck-Mestre & Pynte, 1997; Juffs & Harrington, 1995; Pliatsikas & Marinis, 2013c; Williams, Möbius & Kim, 2001). However, the bilinguals in these studies have typically been late sequential bilinguals often with non-naturalistic exposure to language. Little is currently known about how early sequential or simultaneous bilinguals with naturalistic exposure to English process filler-gap dependencies or, more generally, morphosyntax. Moreover, it is unclear how bilinguals will utilise morphosyntactic cues in real time to facilitate processing, e.g. by aiding disambiguation. Additionally, there is limited work on early bilinguals relative to L2 learners using visual-world eye-tracking studies

Previous studies have suggested that late bilinguals with naturalistic exposure and those with classroom exposure may differ qualitatively from one another in terms of establishing syntactic dependencies. Bilinguals with naturalistic exposure showed processing of sentences with filler-gap dependencies and intermediate gaps more akin to monolinguals than the ones with classroom exposure in a self-paced listening study (Pliatsikas & Marinis, 2013c). In other words, as discussed in Chapter 1, there was a speeding up of reaction times for sentences with movement when there was an intermediate gap than when there was not for the monolinguals and the bilinguals with naturalistic exposure but this was not found for the bilinguals with classroom exposure. However, in this study, even the bilinguals with naturalistic exposure showed differences in processing relative to the monolinguals at the initial critical segment of the sentence heard which is associated with the processing of the intermediate gap. Only the monolinguals showed a slowdown at this point for sentences where there was movement relative to ones without it. Even though the late bilinguals with naturalistic exposure differed from the monolinguals, it is an open question as to how early bilinguals will use morphosyntactic features in comparison to monolinguals.

There is a long-established body of evidence that supports the claim that, similarly to children, monolingual adults have greater difficulty with object NPs in wh-questions than either subject NPs in wh-questions or object NPs in declarative sentences (Crain & Fodor, 1085; Stowe, 1986). Whilst not exclusive to wh-questions, research has long suggested that – similarly to children - processing of filler-gap dependencies in monolingual adults is incremental involving the building of syntactic representations and semantic interrelations in real time. The filler will hence be inserted at the first plausible position for a gap even in the absence of disambiguating information beyond that (Stowe, 1986 for self-paced reading; Traxler & Pickering, 1996 for reading study with eye-tracking). Similar findings have been reported for L2 learners. In one of the earliest studies, in processing wh-dependencies in bilinguals, Juffs & Harrington (1995) tested highly proficient speakers of English who were native speakers of Chinese using a combination of a self-paced reading and grammaticality judgement task. Accuracy was higher and reading times after the embedded verb "think" were faster for subject-extracted sentences such as (22a) in relation to object-extracted sentences such as (22b);

(22a) What does the man think crashed into the car?

(22a) What does the man think the car crashed into?

One limitation of this early study was that it compared segments which contained different words in terms of syntactic category (determiner vs. verb) and with a different semantic content. It also only included only Chinese native speakers who may have difficulty with object-extractions as wh-movement is absent in Chinese, although similar findings have been obtained for English-French learners (Frenck-Mestre & Pynte, 1997) and for English-Spanish learners (Jegerski, 2012). In a follow up study, Juffs (2005) addressed these issues by modifying the experimental sentences to include some additional words as in (23a) and (23b) but also varied the L1 of the L2 learners.

(23a) Who does the nurse know saw the patient at the hospital?

(23b) Who does the nurse know the doctor saw in his office?

Juffs replicated the original findings as all groups of L2 learners had greater difficulty with object wh-questions (lower accuracy scores and slower reading times at embedded verb). In a study on wh-questions by Williams, Möbius & Kim showed (2001), monolinguals and L2 learners completed two self-paced reading tasks with sentences both containing two potential gaps but with one where both potential gaps are plausible or where only one would be plausible as (24a) and (24b) respectively demonstrate;

(24a) Which girl did the man push the bike into late last night?

(24b) Which river did the man push the bike into late last night?

In both (24a), the phrase "Which girl" can be inserted into two positions, either after push as an object to the verb *push* or after the preposition into (plausible condition). In other words, it is initially interpreted as the direct object of "push". However, in (24b), this interpretation is quickly rejected as being implausible. It's not as quickly rejected in (24a) as the initial gap is plausible. All groups showed an increase in reading times at the verb "push" and for the segments after the verb "push" in the implausible relative to the plausible condition suggesting an initial commitment to filling the gap at the earliest possible opportunity. At the phrase "the bike", however, this effect is reversed. Across groups there was a preference to interpret the phrase "Which girl" as a direct object to the verb and the slowdown at the phrase "the bike" reflects the garden-pathing (i.e. a filled-gap effect; Stowe, 1985) in sentences such as (24a) where this misinterpretation is possible. However, the monolinguals showed effects of plausibility at the determiner whereas the non-natives did so at the subsequent noun indicating that the monolinguals made use of the disambiguating information earlier on than the L2 learners. This suggests that the L2 learners were slower than the monolinguals to form a filler-gap dependency. In a follow up task in the same study, another group of monolinguals and L2 learners were asked to indicate which words from the trials from the first experiment did not make sense. There was no effect of plausibility for the monolinguals but the L2 learners judged sentences to not make sense after significantly more segments, i.e. much later on, in the plausible condition than in the implausible condition. This suggests that the L2 learners had difficulty recovering from the erroneous initial interpretation and continued to be more likely to find the plausible sentences unacceptable whereas that this was not the case with the monolinguals.

Early gap-filling was replicated in a follow up study with adjunct extracted wh-questions (Williams, 2006) as in (7) and (8);

(25a) Which machine did the mechanic fix the very noisy motorbike with two weeks ago?

(25b) Which friend did the mechanic fix the very noisy motorbike for two weeks ago?

Reaction times were again slower at the critical region, after the verb “fix” for sentences where the critical region was a plausible gap site as in (25a) than for those where it was implausible as in (25b).

Effects of plausibility were also found for monolinguals and L2 learners in a study by Felser, Cunnings, Batterham and Clahsen (2012) in a reading study with eye-tracking. Both monolingual and L2 learner participants spent more time reading the segment “read” in sentences as in (26a) where the antecedent (in this case “the shampoo”) NP is semantically an implausible object to the verb “read” than in sentences as in (26a) where it is plausible (in this case the magazine);

(26a) Everyone liked the magazine that the hairdresser read extensively and with such enormous enthusiasm about before going to the salon.

(26a) Everyone liked the shampoo that the hairdresser read extensively and with such enormous enthusiasm about before going to the salon

This entails that the parser has treated the NP as a filler and has tried to insert it into the earliest position available. Such plausibility effects in L2 learners signalling early gap-filing have also been found in ERP studies not at the group level but as a function of L2 proficiency with the more proficient L2 learners showing the same effects as monolinguals (Dalas, DeDe & Nicol, 2013)

A further finding is that this strategy to insert fillers into the earliest possible gap site is modulated by the syntactic construction. In the previous examples, the filler-gap dependency is formed within the same clause and as such there are no syntactic constraints on the formation of filler-gap dependences. This will not however be the case in sentences with an embedded relative clause where forming a filler-gap dependency between filler in the main clause and a gap site in the embedded clause is not permitted. Traxler & Pickering have shown plausibility effects in monolinguals when there are no syntactic constraints as in (26a) and (26b) but this effect is absent in sentences with relative clause islands such as (27a) and (27b).

(27a) Everyone liked the magazine that the hairdresser who read extensively and with such enormous enthusiasm bought before going to the salon.

(27b) Everyone liked the shampoo that the hairdresser who read extensively and with such enormous enthusiasm bought before going to the salon.

Felser et al. replicated these findings for L2 learners. Plausibility effects in reading times for both the critical region and the spill-over regions immediately afterwards were found for the monolinguals and the L2 learners in the sentences where there was no constraint (i.e. slower reading for the implausible condition (26b) than plausible (26a)) but not for the constraint conditions (27a) and (27b), suggesting both groups did not construct filler-gap dependencies in the island clause. In contrast to Williams et al. who reported delayed effects for the L2 learners, Felser et al. show effects of plausibility for first-pass reading times in the L2 learners; these were absent in the monolinguals and as such do not suggest slower processing. This effect appears to be present in bilinguals irrespective of the typology of the L1; Cunnings, Batterham, Felser & Clahsen show the same effects of structural complexity for L2 learners with German and Chinese as their L1 (2010) which were larger in the non-natives than the natives.

Other studies suggest larger variability by L1 background (Kim, Baek & Tremblay, 2015). The study tested L2 learners of English with Spanish and Korean as their L1 and used a self-paced reading task with a stop-making-sense component. This entails the participants heard a sentence segment by segment and needed to indicate at what segment the sentence stopped making sense. In contrast to the other two groups of participants, the Korean L1 speakers showed effects of plausibility in the reading times in both the no constraint condition and in the island constraint condition. The absence of an interaction between plausibility and constraint type in the Korean L1 speakers casts doubt on whether they showed sensitivity to island constraints. This suggested a temporary filling of the gap even with relative clause islands, even though the Korean learners demonstrated knowledge that this was ungrammatical in their stop-making-sense judgements. The fundamental difference between Korean, Spanish and English is that the former has in situ wh-questions and therefore, this was considered an effect of language transfer. Nevertheless, Chinese also has in situ wh-questions but this was not found for the Chinese L1 English L2 learners in the Cunnings et al. study (2010).

In a second experiment, Felser et al (2012) manipulated the availability of a gap to be filled in no constraint and relative clause island constraints, as in (28a) and (28b) respectively, relative to sentences where the gap is empty as in (28c) and (28d);

(28a) Everyone liked the magazine that the hairdresser read quickly and yet extremely thoroughly about before going to the beauty salon.

(28b) Everyone liked the magazine that the hairdresser who read quickly and yet extremely thoroughly bought before going to the beauty salon.

(28c) Everyone liked the magazine that the hairdresser read articles with such strong conclusions about before going to the beauty salon.

(28d) Everyone liked the magazine that the hairdresser who read articles with such strong conclusions bought before going to the beauty salon.

In (28c) and (28d), the verb read is transitive, the presence of an overt object means there is no gap which can be filled. If the parser attempts to fill the gap, there will an interruption of processing at the point of the object and the need for syntactic reanalysis will result in an increase in reading times. Reading times were longer for the monolinguals in the filled gap condition relative to the empty gap when there was no constraint (i.e. (28c) compared to (28a)) but not for the relative clause island conditions, (28b) and (28d). This interaction between type of constraint and type of gap was not found for the L2 learners for either first-pass, reading time, regression-path time and total reading time until the spill-over region. This suggests that the non-natives made a delayed use of the syntactic information in real time processing which was interpreted as evidence for the Shallow Structure Hypothesis. One of the limitations of this study is the use of different lexical material at the critical region ("articles" vs. "quickly").

Using a self-paced reading task, Gibson and Warren showed reduced reading times in sentences with an intermediate gap such as (29a) at the point of the final gap, i.e. when reading the verb "pleased", and higher comprehension accuracy in comparison to sentences such as (29b) which contain a filler-gap dependency without intermediate movement.

(29a) The manager who the consultant claimed that the new proposal had pleased will hire five workers tomorrow.

(29b) The manager who the consultant's claim about the new proposal had pleased will hire five workers tomorrow.

This was interpreted as facilitation of processing as the filler was retrieved at the point of the intermediate gap and hence increased residual activation in memory. This effect was observed only in the sentences where there was extraction of the NP. A study with a similar design was carried out by Marinis, Roberts, Felser & Clahsen (2005) with L2 learners. Like in Gibson & Warren's study, there was a significant facilitation in terms of reaction times at the final gap site for sentences with filler-gap dependencies with an intermediate gap relative to those dependencies without an intermediate gap but only for the monolinguals. None of the groups of bilinguals showed this asymmetry and it was consequently speculated that the L2 learners did not utilise the intermediate gap in filler integration. In a follow up study, Pliatsikas & Marinis (2013c) examined differences in intermediate gap processing in L2 learners by type of exposure. Two groups of L1 Greek L2 learners of English were compared to monolingual English speakers; those with naturalistic exposure who had lived in the UK and those with only classroom exposure. All groups slowed down for sentences where there was extraction (wh-movement) relative to sentences where there was not. However, none of the two bilingual groups showed evidence of processing the intermediate gap at the intermediate gap position or immediately thereafter. The picture was different at the final segment analysed. The monolinguals and the L2 learners with naturalistic exposure showed faster reaction times at the final gap segment for sentences where there was an intermediate gap in comparison to sentences where there was not. The L2 learners with only classroom exposure, however, did not show this effect. This was interpreted as evidence that only bilinguals with naturalistic exposure process intermediate gaps and that those without it will show a less grammar-based ('shallow') processing.

It is also possible that the degree of proficiency impacts the processing of morphosyntax. Hopp showed that in a self-paced reading task with sentences with an embedded clause where the first NP is ambiguous (either subject or object, similarly as in object wh-questions) only the near native L2 learners' behaviour patterns with the monolinguals' whereas the advanced L2 learners' does not. The former two groups had longer reading times at disambiguating segments when it was clear that the ambiguous NP was an object and not an expected subject.

Processing morphosyntax has also been associated with cognitive capacities in addition to language proficiency. Numerous accounts have associated gap-filling or trace reactivation with demands on cognitive resources (see Chapter 1) with the main tenet being that previously heard content, i.e. the filler needs to be maintained in working memory in order to be integrated later on in the sentence, i.e. in the case of filler-gap dependencies, inserted into the gap. Dussias & Pinar (2010) carried out another self-paced reading study on long distance wh-questions in English in which they compared Chinese learners of English to monolinguals but split the groups by working memory (high vs low) based on a reading span task. The high span L2 learners performed similarly to their monolingual counterparts in that they showed effects of extraction (objects were harder) and plausibility described previously. Low span L2 learners only showed effects of extraction and did not utilise plausibility to disambiguate early on between subject and object questions.

Use of cues in real time processing in adult bilinguals/ L2 learners

One open question is to what extent bilingual adults utilise inflectional morphology such as case or number of either the verb or the NP to facilitate processing of wh-questions. As discussed in Chapter 2, Contemori et al. (2018) demonstrated that the number mismatch between the two NPs in which-questions can facilitate recovery from misinterpretation relative to sentences where the number of the two NPs does not as evidenced by a steeper increase in looks towards the picture corresponding to the sentence heard (see Chapter 2 for examples). The principal idea of this study was that the mismatch in number between the auxiliary verb and the first noun phrase in a sentence with a number mismatch would signal that the NP is not the subject due to agreement violation and that thus it would be easier to process, which was confirmed. Whilst the study was primarily about children, Contemori et al. also included monolingual adult controls. They found a faster reorientation of looks to the correct picture relative to the children but similarly to the children, this was facilitated by the number mismatch in the locally ambiguous object sentences. Similar effects of number mismatch have been attested for adults and children in other languages. Franck, Colonna & Rizzi (2015) showed faster reading times and higher comprehension accuracy in a self-paced reading task with object relative clauses in French when the number of the two NPs in the clause mismatched in comparison to when they matched, as in Contemori et al (2018). Alternatively, this could indirectly be accounted for under memory cue-based retrieval accounts under which the similarity in features acts as an interference (Gordon et al., 2001, 2002, 2003). If two NPs have the same features which match the relevant retrieval cues, then it becomes harder for the parser to retrieve the appropriate one[24].

[24] It should be noted that number itself is not a retrieval cue at the gap site of object-questions as the verb agrees with the subject. What could be happening is, if there is a mismatch in

Hopp (2017) used the same task as Contemori et al. (2018) with L2 learners of English with German as their native language and compared them to the monolingual adult data from Contemori et al.; the L2 learners had lower accuracy for the object questions and, like the children, had higher accuracy in the number mismatch condition for the object questions but not the subject questions (although the benefit was larger for the children; accuracy score objects-number match 62%, mismatch 88%, respective adult accuracy, 75% and 84%). The L2 learners were split into three groups by proficiency score in the baseline measures (intermediate, high-intermediate and advanced). All three groups showed an effect of structure; accuracy was higher for subject questions but only the high-intermediate group showed an interaction of structure by number match; mismatch benefitted the comprehension of object questions. The intermediate L2 learners showed an effect of match in the expected direction but this was not specific to object questions. The advanced L2 learners did not show this effect but this may be due to their performance reaching ceiling effects. For the gaze data (proportion of looks towards the picture which matches the sentence heard relative to the competitor in a five 400ms time windows), Hopp found more looks towards the target picture for subject questions relative to object which-questions. As with Contemori et al., Hopp found an initial decrease in looks to target for object questions relative to looks to the competitor picture which in both studies is explained as an initial misinterpretation of the question as a subject which-question. Group by structure and/or number match were found in the second and fourth time-windows. In all three groups of L2 learners there was an effect of structure, i.e. more looks to target in subject questions and either an effect of number match, i.e. more looks to target in mismatch condition, and/or a structure by number match interaction, i.e. the benefit of number mismatch was more prominent or exclusive to object which-questions. The advanced L2 learners showed limited effects of misinterpretation and

---

number, the only NP that agrees with the verb will not be a candidate for retrieval as it is already in subject position, thereby facilitating retrieval of the correct NP.

tended to assign a subject-question reading to an object question less than the other two groups, although looks to target remained around chance performance for longer than for the subject questions suggesting more prolonged indeterminateness and a later commitment to an object-interpretation of the object which-questions (Hopp, 2017:125, Figures 10 & 11). Visual inspection of looks to target picture for the incorrectly comprehended object question in Hopp's study shows that the intermediate L2 learners never increased beyond chance over time indicating a failure to reanalyse a sentence once garden-pathed although, even in this case, an effect of number match is reported. According to Hopp (2017), the advanced bilinguals reorient their looks to the correct picture earlier on than the other two groups and were still aided by the number mismatch. The high intermediate groups show a more protracted reorientation of looks which is speeded up by number mismatch as with the advanced L2 learners and monolinguals but the intermediate L2 learners showed an even longer reorientation of looks to the correct picture, which was aided by number mismatch only in the trials where the comprehension question was answered correctly.

In sum, existing research suggests that filler-gap dependencies are processed incrementally by both monolinguals and L2 learners. However, L2 learners have greater difficulty recovering from misinterpretation which may result from incremental processing. Both groups appear to be sensitive to semantics during real time processing; plausibility caused a slowdown at the position of the gap. Grammatical cues such as number match have been shown to be utilised by monolingual adults, but it remains an open question as to whether this holds for early bilinguals and L2 learners.

The current study

It was shown in Study 1A that the bilingual children showed similar qualitative patterns to the monolinguals but that there were differences in the time course of processing even though the bilingual children all scored within age-appropriate monolingual norms and that these children were comparable to the monolinguals in a series of language baseline tasks. On the other hand, research into bilingual adults has focused on late bilinguals who are usually classified as L2 learners. This means it is an open question as to whether such differences will be found in adult bilinguals[25] with a second or additional language. This study aims to answer this question. A further aim of this study is to bridge the gap in the literature in terms of comparison between morphosyntactic processing in bilingual native speakers of a language and L2 learners of the language. Most of the research in bilingual language processing has focused on differences between native and non-native speakers of a language with regards to morphosyntactic processing but there is very limited work on bilinguals who attain nativelike competence. Available evidence suggests that early bilinguals (e.g. Birdsong, 2001; Johnson & Newport, 1989) or bilinguals with naturalistic exposure (e.g. Dussias, 2003; Pliatsikas & Marinis, 2013c) who develop nativelike proficiency typically perform similarly to monolingual speakers of the language and show different patterns to those of L2 learners. However, the differences between the two types of bilinguals have not been examined in terms of the time course of locally ambiguous filler-gap dependencies.

This study builds on the work by Hopp (2017) in two more ways; it tests L2 learners who are living in the UK and compares native bilinguals to them and recruited participants from a variety of L1s. A further contribution in relation to both Hopp (2017) and Contemori et al. (2018) is that it controls for the number of the first noun phrase. In Study 1A, this was found to be significant for the children suggesting a psycholinguistic reality for one type of linguistic markedness. It remains, however, an open question as to whether this will be the case in the adults. A final contribution is that, in line with Contemori et al., time is used an

[25] We acknowledge that this is not a longitudinal study and we make no claims about potential developmental changes which may occur within the same individuals.

independent variable to model not just differences in looks between group overall or in specific time windows but differences in how looks to the correct picture change over time by group and condition.

The research questions are therefore, the following;

1. Do bilingual native speakers differ in their performance to monolinguals and L2 learners in processing which-questions in English on a range of measures reflecting the end result (accuracy, reaction times, total looks to target picture)?
2. Does the time course of processing in terms of garden path effects and recovery from them differ between native bilinguals, L2 learners and monolingual adults? Do the early native bilinguals perform similarly to monolingual natives or L2 learners?
3. Do native bilingual adults make use of morphological cues to facilitate processing of which-questions in the same way monolingual adults and adult L2 learners do?
4. Does the number of the first noun phrase have an impact on the processing of which-questions? If it does, is the effect the same for monolingual adults, bilingual native speakers and L2 learners.

Given the past research (e.g. Pliatsikas & Marinis, 2013b) in other areas and the degree of proficiency of the two bilingual groups, it is expected that the native bilinguals will cluster with the monolinguals rather than with the L2 learners on a number of measures; namely accuracy, reaction times, looks to the correct picture (intercept term) either overall or in the trajectory of change for looks to target (linear and quadratic terms). This is expected to be evidenced by an effect of group at the type of bilingual contrast (native vs L2 learner) but not at the type of native speaker contrast (monolingual vs. bilingual) on any of the previous measures.

It is expected that all groups will process language incrementally; therefore, it is expected that there will be an initial misinterpretation of object questions as subject ones in line with both the Active Filler Hypothesis and expectation-based models. This would also be consistent with expectation-based models under which the difficulty with object-questions is that they do not conform to the listeners' expectations of a subject question. Looks to the target picture will drop below chance and increase thereafter, reflecting the fact that the additional difficulty for object questions is the result of misinterpretation with the increase evidencing re-analysis. One possibility is that the drop will be less pronounced in the L2 learners with looks to both pictures remaining at about chance for longer if they are slower to build a syntactic structure and therefore, less likely to be garden-pathed. This would entail that both groups of native speakers are expected to show a greater drop below chance initially with a steeper increase. This is not expected to be the case for subject which-questions as looks will increase from the beginning and continue rising until a point. Given the results from Hopp (2017) who found these effects even in the intermediate L2 learners, the above is unlikely. Moreover, the self-paced listening studies on filler-gap dependencies with L2 learners have shown effects at the same critical regions as monolinguals. We therefore, expect all three groups to experience garden-path effects with the subject-object asymmetry in all experimental measures.

One further issue that can be addressed on the basis of previous studies is the course of recovery for the bilingual groups relative to the monolinguals. As will be discussed in more detail in Chapter 5, it has been shown that L2 learners experience garden-pathing as monolinguals do, but that the L2 learners have greater difficulty recovering from garden-paths (Pozzan & Trueswell, 2016; Jacob & Felser, 2016; for similar results for children vs. adults, see Contemori et al., 2018). Given the fact that processing appears to become more nativelike with an increase in proficiency (Hopp, 2017) and naturalistic exposure (Pliatsikas

& Marinis, 2013b), it is probable that the native bilinguals will show a pattern of recovery from misinterpretation similar to that of the monolinguals and faster than that of the L2 learners evidenced by a steeper curve.

If bilingual adults make use of the number mismatch in accordance to Relativized Minimality, there will be an interaction with structure and number match; in other words, the effect of number (mis)match should be present only in the object questions. This effect may be manifest in accuracy (higher for number mismatching object questions), reaction times or the gaze data. Looks to the correct picture may be more if the number of the two NPs does not match. An alternative measure of the effect of number mismatch is the steeper increase in looks to the correct picture for the mismatching questions relative to the matching ones. Relativized Minimality is not a theory which makes any explicit distinction between monolinguals and bilinguals, native or non-native speakers. The Shallow Structure hypothesis would predict that the L2 learners will make greater use of number mismatch to aid disambiguation as their system of parsing is less syntactic in nature. On the other hand, the effect of number mismatch is predicted to be the same in the native bilinguals and the monolinguals as the appropriate syntactic system is in place. This would also be consistent with the findings from Hopp (2017). If all groups make use of number (mis)match in the same way, then an interaction of group with structure and number or both is not predicted.

Most theories of language processing do not make explicit predictions about the effect of number of the first noun phrase, particularly with regards to bilingual processing. It is generally expected that marked forms are harder to acquire or process (Harley & Ritter, 2002; Haspelmath, 2005 but see White, 1987 for a series of features where this is not the case), perhaps stronger for the L2 learners for reasons of proficiency. Empirical evidence in support of markedness impacting processing of dependencies comes from a study on agreement attraction errors which showed more errors of attraction caused by plural nouns than by singular nouns (Wagers et al., 2009). It is unclear, nonetheless, what the Shallow Structure Hypothesis would predict but it is conceivable that it would not predict any effects for the monolinguals or the early/native bilinguals but that if any effects were observed they would be found in the late bilinguals/L2 learners. This would be empirically consistent with Hopp (2017) who found stronger effects on processing for the L2 learners with lower proficiency as well as Pliatsikas & Marinis (2013b) who showed less nativelike processing for L2 learners with classroom exposure relative to those with naturalistic exposure. The Interface hypothesis would not make any predictions about the effect of the number of the first NP as there the feature is purely morphosyntactic and there is no other cognitive domain with which syntax interfaces. Memory-based accounts also do not make any predictions as the number of the first NP will not burden memory resources or provide a retrieval cue on its own. This particular feature is also not relevant to expectation-based models under which wh-object questions are unexpected regardless of the number of the first NP. The current literature provides little insight into how the bilingual adults may perform specifically for the gaze data relative to the monolingual adults and the L2 learners as there are, to our knowledge no studies with adults using eye-tracking in this area of research. It is plausible that any effect in the gaze data will follow the direction of accuracy and/or reaction times. One possible difference is that the time course will be different, i.e. slower increase in looks towards target for the L2 learners relative to the monolingual and bilingual native speakers. This would suggest the L2 learners have greater difficulty with plural First NP than the other two groups.

3.2. Methods

3.2.1 Participants

3.2.1.1 Sample size and recruitment

A total of 107 adults were recruited for the purposes of this study and were included into three groups; monolingual speakers of English (N = 43, 32 women), native speakers of another language with English learned as a second or foreign language in a classroom setting (N = 29, 23 women), and adult bilinguals who were exposed to English either at home as children or as a majority language alongside another language (N = 35, 28 women). These three groups are henceforth referred to as L1 adults or monolinguals, L2 adults or L2 learners and 2L adults or bilinguals respectively. The monolinguals were mainly undergraduates studying psychology at the University of Reading (N = 33), while the remaining participants were post-graduate students in the same school. All but one participant had grown up in the UK – one participant did so in the US. All participants spoke English exclusively at home. In all three groups, a larger number of participants was female, reflecting the population of psychology students at the University of Reading.

Background information was collected from the adult participants through semi-structured interviews. This captured the background from the bilingual group and the L2 learners. Information asked as standard practice pertained to place of birth, languages spoken and to what extent, languages used at home and in education, age of first exposure to English and was modelled on the questionnaire used by Pliatsikas & Marinis (2013a, b). Participants were given an information sheet prior to the day of testing and had to complete a consent form before the testing commenced.

3.2.2.2 Age

Table 6 shows the participants' age by group.

Table 6 Age of adult participants

	Monolinguals	Bilinguals	L2 learners
N	43	35	29

Range (in years; months)	18-28	19-53	18-38
Mean (in years)	20.47	23.07	24.32
SD (in months)	2.38	7.33	6.33

The majority of participants were young adults. A one-way ANOVA with the factor age showed an effect of age ($F(2,93) = 3.994$, $p = .022$, $\eta^2_p = .082$). Pairwise comparisons with a Bonferroni post-hoc correction showed that the L2 group were older than the L1 group ($p = .021$). There were no other significant differences between any of the groups. The difference in age between the groups can be attributed to practicalities relevant to recruitment which is outlined in a subsequent section. Whereas recruitment for monolinguals was accomplished through an electronic system available exclusively to undergraduates (which meant they were more likely to be younger and more homogenous age-wise), this was not possible for the L2 learners and the bilinguals. Therefore, it was necessary that recruitment attempts were broadened. This meant that the groups were more heterogeneous in age and older in raw numbers. It is not expected that this difference in age will impact the results, as there is no literature-based expectation of developmental changes or cognitive decline in language and language processing abilities in people in the early to mid-20s.

3.2.1.3 Language profile of the bilingual adults

The L2 learners were exposed to English as a second language at school around the age of 8-11 years. All participants spoke another language at home and were educated in this other language throughout their life prior to their arrival in the UK. The L2 learners have lived in the UK for less than two years; therefore, their naturalistic exposure to English was limited. All participants reported English was a non-native language to them and that they were more proficient in the other language. Length of exposure was estimated at between 8.5 and 29 years (M = 14.91 years, SD = 6.33). As all L2 learners were students at the University of Reading, their language proficiency was of a minimum IELTS score of 6.5/9 or equivalent - e.g. CPE or high CAE score due to entry requirements. This is expected to be sufficient to attend English medium higher education at the University of Reading and equates to a proficient user of English under the classification of the Common European Framework (https://www.ielts.org/ielts-for-organisations/common-european-framework). The participants were from a mixed linguistic background with a total of 16 languages spoken between the 29 participants; there were five L1 German speakers, 3 L1 Malay, 2 L1 Bengali, Lithuanian, Cypriot Greek, Portuguese and Arabic and one participant with Italian, Bulgarian, Greek, Afrikaans, Japanese, Polish, Chinese, Dutch and Turkmen as their native language.

The native bilingual participants were a heterogeneous group of individuals who were exposed to English early on in a naturalistic setting but at the same time grew up with exposure to another language. The bilingual population varied in its linguistic experience in English but can be summed up in four partially overlapping categories. The largest group spoke French as their other language (N = 5), followed by Mandarin and German (N = 4 and 3 respectively). There were two speakers of Arabic, Bulgarian, Greek and Hindi and a single speaker of Bengali, Bosnian, Cantonese (alongside Mandarin), Creole English, Cypriot Greek, Ewe, Italian, Polish, Punjabi, Shona, Spanish, Tagalog, Turkish, Urdu and Yoruba.

Their defining feature is the early exposure to English (M = 1.29 years, SD = 2.37, range birth to 10 years[26], but all bar one were first exposed to English under the age of four years) and the naturalistic - and quantitatively more than in L2 learner settings - exposure to English either through the family and/or in education. The largest group of bilinguals were exposed to two languages at home (N = 15). All but four participants in this group were born in the UK and all except one were exposed to English from birth as it was one of their parents' native

[26] This participant was reportedly strongly dominant in English

languages. The second largest group (N = 11) was children of immigrants; they spoke a language other than English at home and were exposed to English as a majority language. These participants reported being first exposed to English in kindergarten (2-4 years approximately). One participant reported being exposed to English in primary school (10 years) but was reportedly dominant in English. The study also included two smaller groups of bilinguals. One could be classed as heritage speakers of English (N = 5). All but one reported speaking only English at home but grew up in countries where English was not spoken as a majority or second language. The last group was bilinguals who grew up in countries with a diaglossic situation (N = 4), namely Malaysia (2 participants), Haiti, and Zimbabwe. They reported using more than one language at home including English and reported using both languages outside home as well. There is a degree of overlap between these two groups of bilinguals. For example, one participant who is the child of immigrants also reported using English at home in addition to Tagalog, while four of the participants who were exposed to two languages at home may also fall under the label heritage speaker.

3.2.2 Design

The experimental design for this study was the same as for Study 1A. For more detail, see Chapter 2.

3.2.3 Materials

All materials for this study were the same as for Study 1A. For more detail, see Chapter 2.

3.2.4 Procedure

The procedure for the task was the same as the procedure for Study 1A with children. Adult participants completed this task in a single 60-90-minute session alongside a number of other experimental tasks. The order of testing was mostly constant with the adults with the task in Study 1B being undertaken about half-way through the testing session.

3.2.5 Analyses

The statistical analyses adopted were the same as for Study 1A. The only difference was that as there were three groups of participants, Helmert coding was used for the between-subjects comparison. The latter allows for a comparison between two groups and then between two subgroups within the one group; for the purposes of this study, these contrasts were used in two ways with an aim to tease apart the effects of bilingualism and nativeness and to allow a comparison between native bilinguals and the other two groups consistent with the research questions: one set of contrasts first compared monolinguals vs. bilinguals overall and then early bilinguals vs. late bilinguals/L2 learners. The second set of contrasts first compared native and non-native speakers first and subsequently, monolinguals vs. early bilinguals, who make up the group of native speakers. The results from models with the two different sets of Helmert contrasts are reported. By analysing the groups in these two ways, we are able to distinguish between effects of bilingualism on language processing from effects of non-native processing.

A substantial body of research has compared monolingual native speakers to L2 learners (see Cunnings, 2017 for a review). The aim of this study is not to examine the differences between native monolinguals and non-native L2 learners but to examine how the early or simultaneous bilinguals with naturalistic exposure will perform and to what extent they will pattern with the monolinguals or the L2 learners. There are two reasons for this; this contrast was not done in the previous analyses and it is not clear to what extent any differences between L2 learners and the other two groups can be attributed to limited proficiency of the former group. By showing that the L2 learners were equally as accurate as the monolinguals, it can be argued that any differences observed between the L2 learners and the other groups are not confounded by differences in proficiency.

3.3. Results

3.3.1 Accuracy

Table 7 summarises the average accuracy by group and condition.

Table 7 Accuracy as a percentage (SD) by group and condition

	Monolinguals	Bilinguals	L2 learners
Subject questions – Number match – First NP Singular	98.5% (12.3)	98.4% (10.1)	98.8% (11.0)
Subject questions – Number match – First NP Plural	99.0% (10.9)	99.0% (12.3)	97.5% (15.5)
Subject questions – Number match	98.7% (11.3)	98.7% (11.3)	98.2% (13.4)
Subject questions – Number mismatch – First NP Singular	98.5% (12.3)	99.1% (9.4)	96.6% (18.1)
Subject questions – Number mismatch – First NP Plural	98.1% (13.8)	97.8% (14.7)	96.5% (18.3)
Subject questions – Number mismatch	98.3% (13.0)	98.5% (12.1)	96.6% (18.2)
All subject questions	98.5% (12.1)	98.8% (11.1)	97.4% (15.9)
Object questions – Number match – First NP Singular	96.2% (19.3)	96.5% (18.4)	91.1% (28.5)
Object questions – Number match – First NP Plural	96.9% (17.3)	96.2% (19.2)	92.0% (27.2)
Object questions – Number match	96.5% (18.3)	96.4% (18.8)	91.6% (27.8)
Object questions – Number mismatch – First NP Singular	99.0% (9.8)	96.7% (17.9)	93.1% (25.4)
Object questions – Number mismatch – First NP Plural	98.0% (14.0)	96.1% (19.4)	90.3% (29.6)
Object questions – Number mismatch	98.5% (12.2)	96.4% (18.8)	91.6% (27.7)
All object questions	97.4% (15.8)	96.4% (18.7)	91.6% (27.8)
All questions	98.0% (14.1)	97.6% (15.4)	94.6% (22.7)

3.3.1.1 Monolinguals vs bilinguals; then bilinguals vs L2 learners

There was a significant main effect of structure ($\beta = -0.50$, SE = 0.11, $z = -4.63$, $p < .001$) and an interaction of structure by number match ($\beta = -0.21$, SE = 0.081, $z = 2.58$, $p = .010$). There was no effect of number of first NP. Accuracy was higher for subject questions than for object questions, although it was high for both structures and near ceiling for the subject which-questions (Figure 1). Accuracy was higher for the number mismatch for the object condition but not so for the subject condition. The effect, however, was very small. There was a trend for an effect of group in the second contrast, i.e. between bilinguals and L2 learners ($\beta = -0.32$, SE = 0.17, $z = -1.89$, $p = 0.059$) and a trend for an interaction between group and structure in the first contrast, i.e. monolinguals compared to bilinguals ($\beta = -0.26$, SE = 0.15, $z = -1.69$, $p = 0.092$). Accuracy trended to be higher for the bilinguals than for the L2 learners (see Table 7).

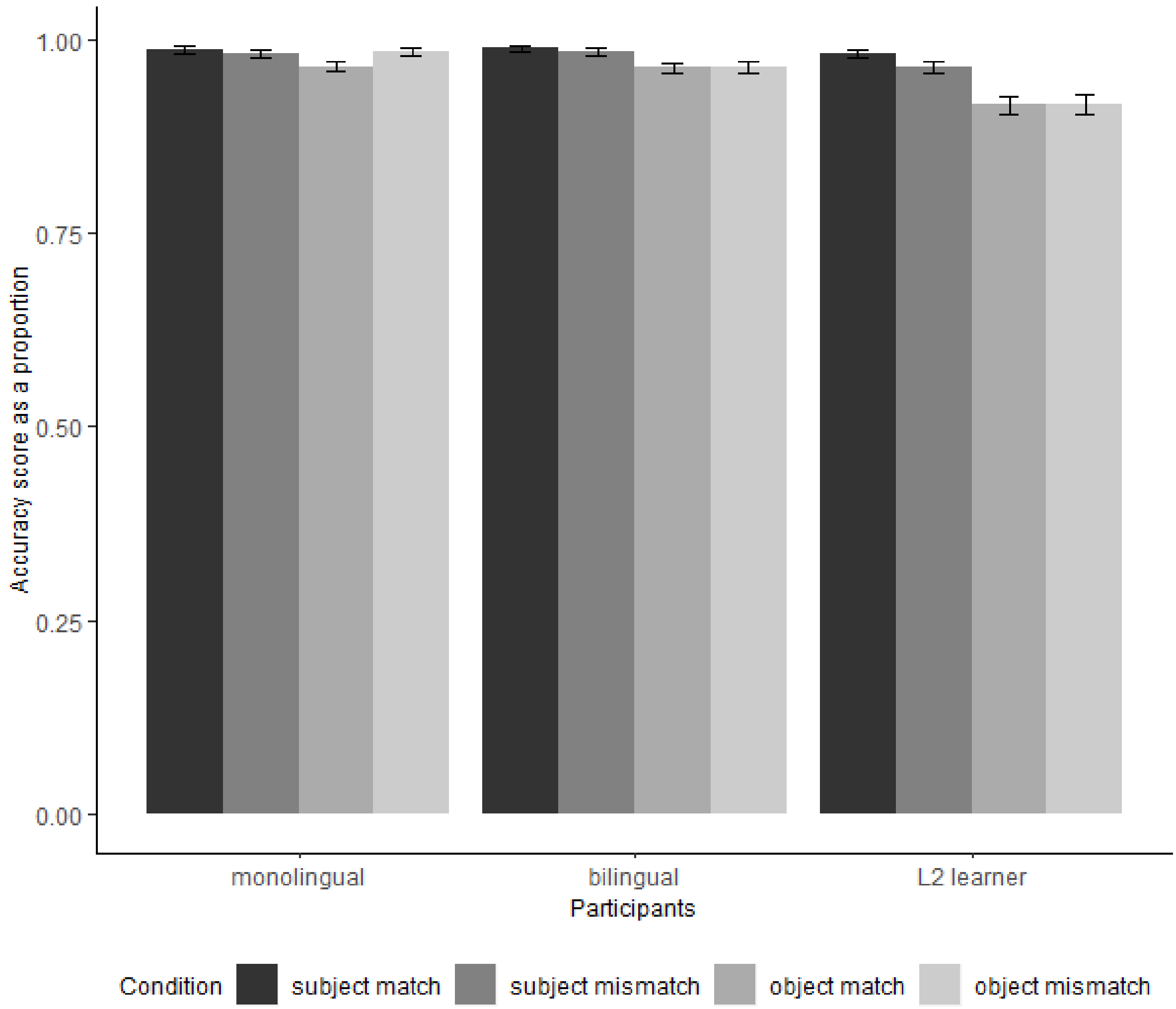

Figure 10 Mean accuracy as a ratio (SE) by group and condition

Given the aforementioned interaction, planned comparisons were undertaken for subject and object questions separately. A model fitted to the object questions only showed a main effect of group for both contrasts (monolinguals vs bilinguals; β = -0.52, SE = 0.23, z = -2.23, p = 0.026, native vs non-native bilinguals; β = -0.38, SE = 0.189, z = -2.01, p = 0.044), a trend for number match (β = 0.18, SE = 0.10, z = 1.83, p = 0.069) and a significant interaction of number match with the group in the first contrast (β = -0.31, SE = 0.15, z = -2.06, p = 0.040). Accuracy was higher for the monolingual group than for bilinguals and within the bilinguals, it was higher for the native bilinguals than for the L2 learners. Accuracy was marginally higher when the number of the two NPs was a mismatch than a match (95.7% and 94.9% respectively). The group by number interaction for the first contrast is due to the fact there was no difference in accuracy depending on number match for the two bilingual groups and a slight increase for the monolinguals. Further analyses of the monolingual and bilingual data separately showed a trend for number match in the monolinguals (β = 0.45, SE = 0.24, z = 1.88, p = 0.061) but no such effect in the bilinguals (β = 0.04, SE = 0.01, z = 0.37, p = 0.711). For the model for subject questions only, there were no main effects or interactions.

3.3.1.2 Native vs non-native speakers; then monolingual vs bilingual natives

The results for the alternative contrast (i.e. native vs non-native speakers, then monolingual vs. bilingual native speakers) reflect a similar image. There was a main effect of structure (β = -0.50, SE = 0.11, z = -4.63, p <.001) with accuracy across groups higher for subject questions than for object questions (98.2% vs. 95.3%). Again, there was an interaction of structure by number (β = 0.20, SE = 0.08, z = 2.58, p = 0.010). There was a main effect of group for the first contrast (β = -0.44, SE = 0.19, z = -2.34, p = 0.019) but not the second. Accuracy was higher for the native speakers than the non-natives (97.8% vs. 94.6%) but there were no differences between monolingual and bilingual native speakers. There were no other main effects or interactions which reached significance or trended.

To examine this interaction, the separate models were fit to the data by structure type. The model for the object questions showed an effect of group for the first contrast (β = -0.66, SE = 0.23, z = -2.90, p = 0.004), a main effect of number (β = 0.19, SE = 0.10, z = 1.96, p = 0.050) and a trend for a group by number interaction for the second contrast (β = -0.23, SE = 0.13, z = -1.75, p = 0.080). Accuracy was higher for the native speakers than for the non-natives at 96.9% and 91.6% respectively and marginally higher for the number mismatch condition (95.7% vs. 94.9%). There were no main effects or interactions in the model for the subject questions.

3.3.2 Reaction times

3.3.2.2 Monolinguals vs bilinguals; then native vs non-native bilinguals

There was a main effect of group in the contrast monolinguals vs bilinguals ($\beta = 292.66$, $SE = 96.32$, $t = 3.04$, $p = 0.003$) and a main effect of structure overall ($\beta = 87.83$, $SE = 20.84$, $t = 4.22$, $p < 0.001$). The bilinguals were slower than the monolinguals, although the native speaker bilinguals were not faster than the L2 learners ($\beta = 29.15$, $SE = 70.85$, $t = 0.41$, $p = 0.681$), while reaction times were faster for subject questions than for object questions. There was also a marginal group by structure interaction in the first contrast ($\beta = 50.49$, $SE = 28.63$, $t = 1.76$, $p = 0.081$) and a number match by first NP number interaction across groups ($\beta = -46.10$, $SE = 12.04$, $t = -3.83$, $p < 0.001$).

Models fitted to the monolingual and bilingual data separately showed a trend towards a difference in reaction times in the bilingual group by structure ($\beta = 105.76$, $SE = 55.60$, $t = 1.90$, $p = 0.076$) with slower reaction times for object questions (see Figure 2). This was not found in the monolingual data ($\beta = 39.57$, $SE = 36.05$, $t = 1.10$, $p = 0.297$). No other main effects or interactions were significant or trended in either model.

Table 8 and Figure 2 summarise the mean reaction times for accurate trials by group and condition.

Table 8 Mean reaction times in ms (SD) by group and condition

	Monolinguals	Bilinguals	L2 Learners
Subject questions – Number match – First NP Singular	3869 (828)	4207 (1014)	4115 (1194)

Subject questions – Number match – First NP Plural	3976 (781)	4373 (1182)	4328 (1428)
Subject questions – Number match	3872 (792)	4290 (1104)	4220 (1318)
Subject questions – Number mismatch – First NP Singular	3874 (712)	4189 (1125)	4252 (1496)
Subject questions – Number mismatch – First NP Plural	3869 (828)	4246 (995)	4188 (1149)
Subject questions – Number mismatch	3872 (765)	4214 (1069)	4224 (1352)
All subject questions	3889 (780)	4254 (1088)	4222 (1333)
Object questions – Number match – First NP Singular	3915 (879)	4224 (1116)	4242 (1582)
Object questions – Number match – First NP Plural	3915 (879)	4557 (1425)	4499 (1960)
Object questions – Number match	3972 (934)	4389 (1289)	4373 (1788)
Object questions – Number mismatch – First NP Singular	3923 (1020)	4371 (1361)	4433 (1773)
Object questions – Number mismatch – First NP Plural	4006 (1181)	4519 (1729)	4373 (1932)
Object questions – Number mismatch	3967 (1108)	4449 (1567)	4401 (1856)
All object questions	3969 (1018)	4417 (1423)	4386 (1819)
All questions	3928 (905)	4333 (1265)	4300 (1584)

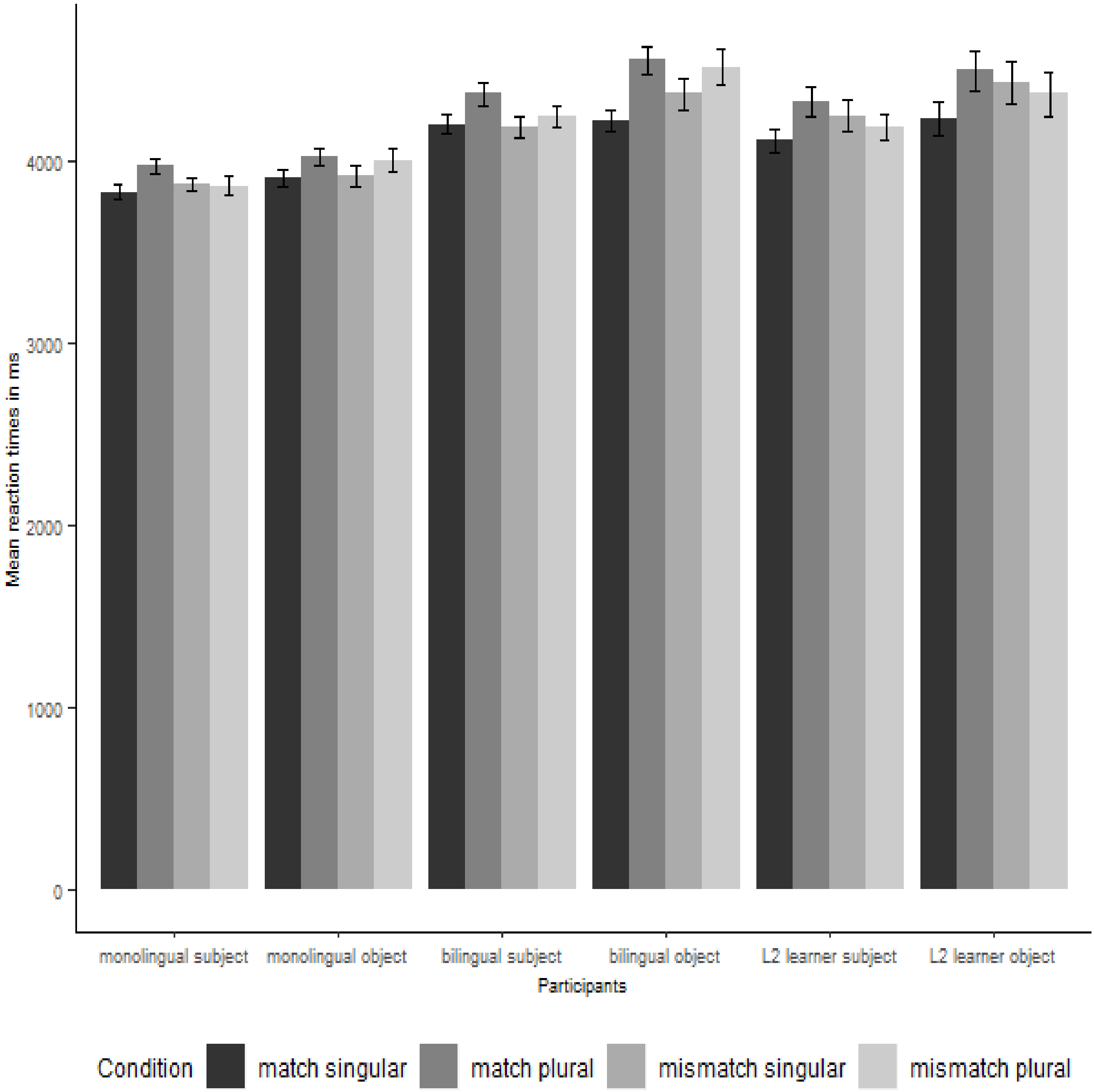

Figure 11 Mean reaction time in ms. (SE) by group and condition

In terms of the number by First NP interaction, number mismatch lead to shorter reaction times in the conditions where the first NP was in the plural but not in the singular. In light of the number match by First NP number interaction across groups, separate models were fitted to the trials where the first NP was singular and plural. There was a main effect of group for the first contrast (β = 264.42, SE = 89.20, t = 2.96, p = 0.004) and a trend for structure (β = 38.16, SE = 19.89, t = 1.92, p = 0.058) for sentences where the first NP was in the singular. There were no other effects or interactions including of number. For the trials with a plural first NP, there was again a main effect of group for the first contrast (β = 300.89, SE = 115.90, t = 2.60 p = 0.011), a main effect of number (β = -60.38, SE = 17.80, t = -3.39, p < 0.001) but also a trend for structure (β = 110.45, SE = 58.66, t = 1.88, p = 0.085). In the sentences with the first NP in the plural, reaction times were faster in the number mismatch condition than in the number match one.

3.5.2.2 Native vs non-native speakers; then monolingual vs bilingual natives

There was a main effect of group for both contrasts (for native vs non-native speakers; β = -170.65, SE = 85.43, t = -1.99, p = 0.047; for monolingual vs bilingual native speakers; β = 205.60, SE = 80.64, t = 2.55, p = 0.012) largely reflecting what was observed in the first comparison. Reaction times were faster for the native speakers than for the non-natives and faster for the monolingual relative to the bilingual natives. There was also a main effect of structure (β = 74.50, SE = 20.92, t = 3.56, p < 0.001) with reaction times being consistently slower for object than for subject questions. The following interactions were found to be significant; a structure by first NP (β = 24.46, SE = 11.90, t = 2.06, p = 0.040) and a number match by first NP number interaction (β = -30.23, SE = 12.11, t = -2.50, p = 0.013). The latter was also observed in the first analysis but the former one was not. Neither interaction was specific to one of the contrasts. As with the first analysis, the benefit of a singular first NP in the form of faster reaction times was stronger in the number match condition and in the object condition. Conversely, the processing cost (i.e. longer reaction times) found in the harder object question and number match conditions was more pronounced in the first NP plural condition. To address the two interactions of structure with the number of the first NP, separate models where fitted to the data based on the number of the first NP.

The model fitted to the trials with a plural first NP showed a main effect of group in the first contrast (native vs non-native speakers; $\beta = 273.28$, $SE = 86.97$, $t = 3.14$, $p = 0.002$) but not in the second contrast (monolingual vs bilingual natives; $\beta = 32.42$, $SE = 71.10$, $t = 0.46$, $p = 0.649$) and an effect of number match ($\beta = 45.63$, $SE = 16.24$, $t = 2.81$, $p = 0.005$; see Figure 6). For the questions with a singular first NP, there was a main effect of group in the contrast between native and non-native speakers ($\beta = 303.31$, $SE = 115.72$, $t = 2.62$, $p = 0.010$), a main effect of number ($\beta = -55.34$, $SE = 17.88$, $t = -3.09$, $p = 0.002$) and also a main effect of structure ($\beta = 105.37$, $SE = 26.03$, $t = 4.05$, $p < 0.001$).

3.3.3 Gaze data

3.5.3.1Monolinguals vs bilinguals; then native vs non-native bilinguals

In this section, we report main effects and interactions of the dependent variables on the intercept and the linear polynomial term. The former reflects effects that hold irrespective of time while the latter reflect variables which impact the change in participants' looks to the target picture over time. The effects of individual time terms are also reported as they reflect the nature of change in participants' looks to target over time. Figures 12 and 13 provide an overview of the effects by group.

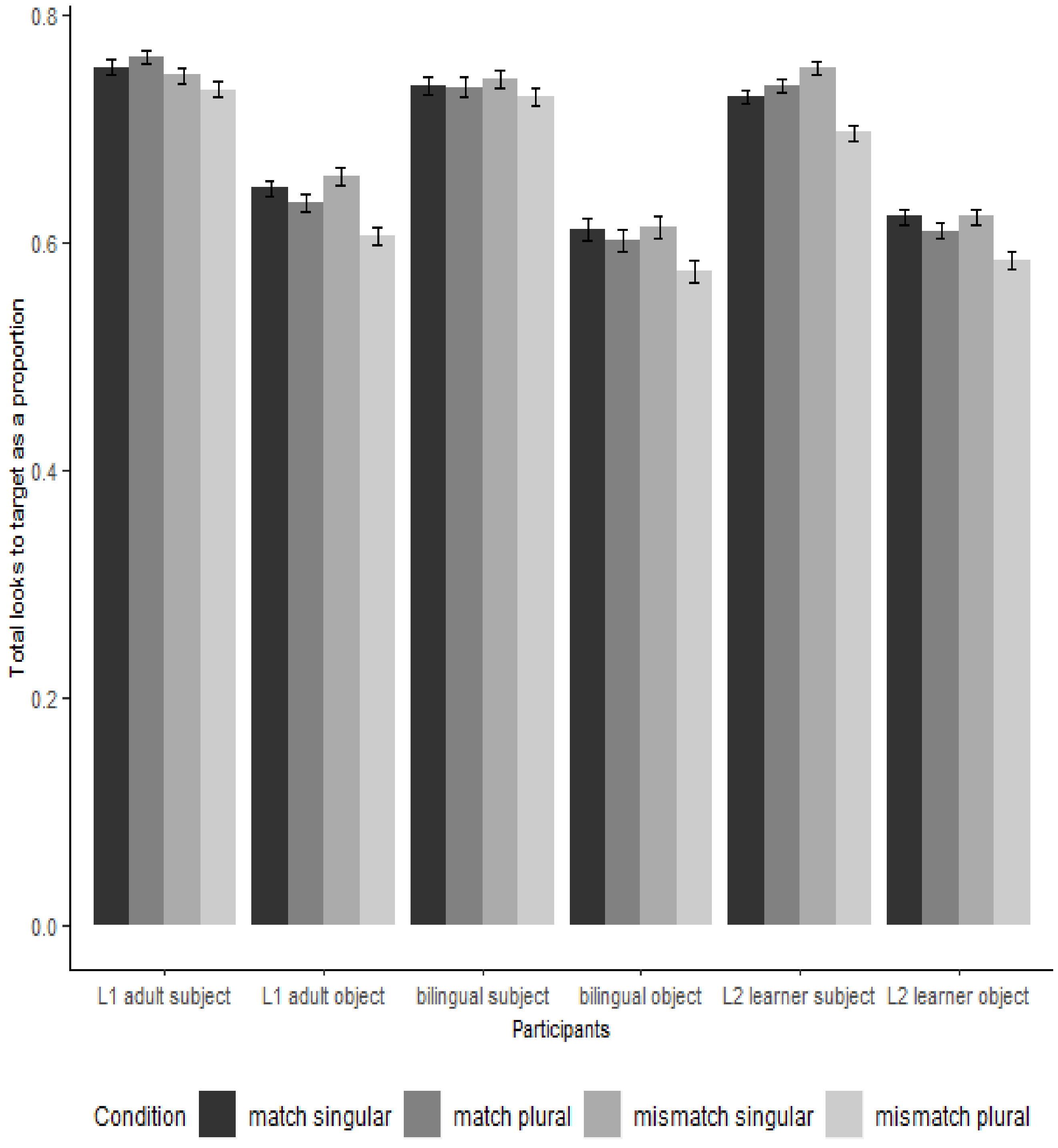

Figure 12 Total looks as a ratio (SE) by group and condition in the time window analysed (200-2,200 ms post onset of auxiliary)

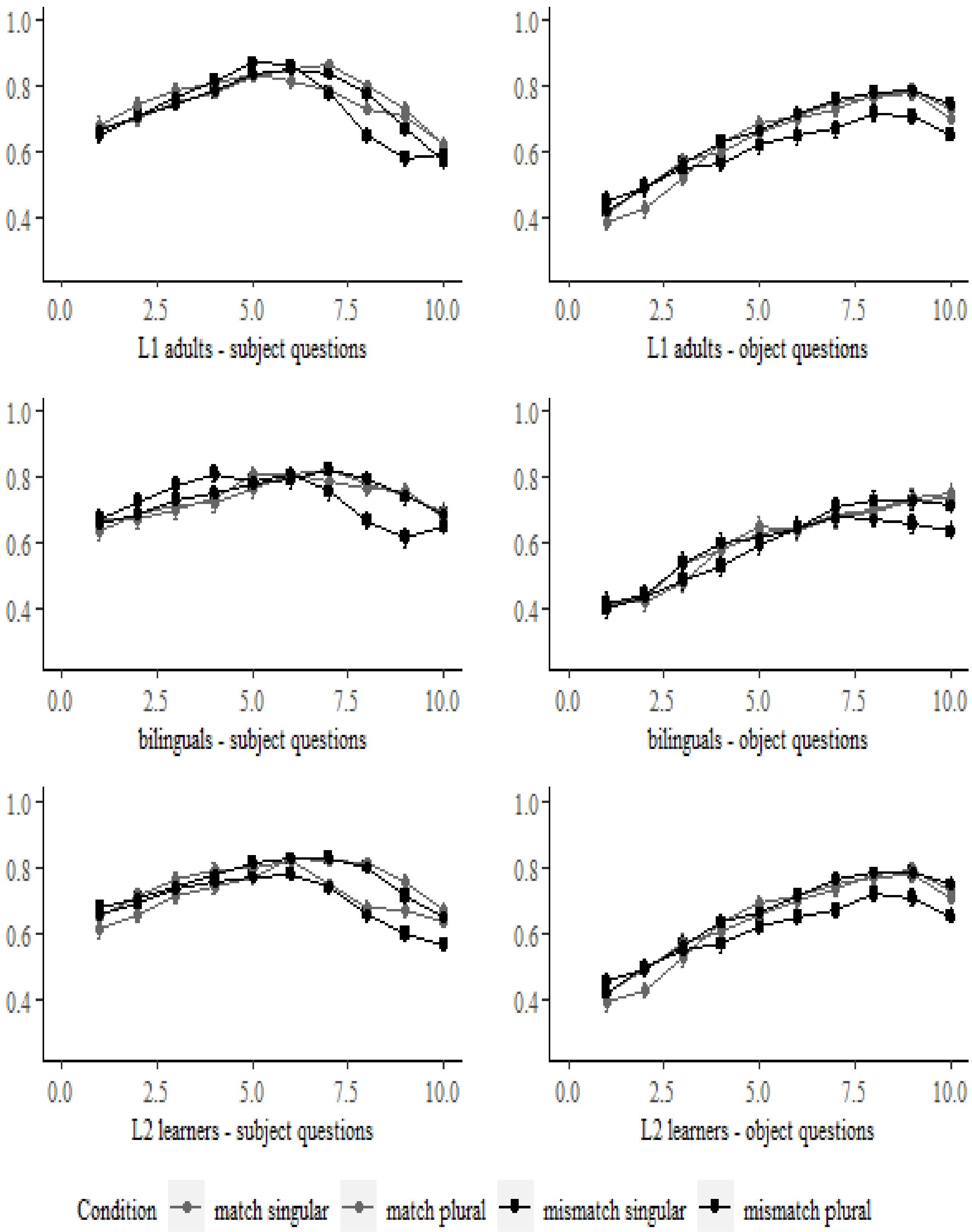

Figure 13 Looks to target as a ratio (SE) by group and condition over time in the time window analysed (200-2,200 ms post onset of auxiliary)

The only main effect on the intercept, i.e. impacted looks to target overall, was structure (β = -0.15, SE = 0.02, t = -9.12, p < 0.001). There were more looks to target when the sentence contained a subject question than an object question; this was the case for all three groups.

There was no main effect of number match (β = -0.02, SE = 0.02, t = -0.89, p = 0.395), number of first NP (β = -0.01, SE = 0.02, t = -0.82, p = 0.429) or group in either contrast (monolinguals vs. bilinguals; β = -0.03, SE = 0.03, t = -1.021, p = 0.310 and bilingual native vs non-native speakers; β = 0.02, SE = 0.02, t = 0.61, p = 0.542).

The linear, quadratic and cubic polynomial terms were significant predictors of looks to target (β = 42.90, SE = 1.17, t = 38.41, p < .0001 for linear term, β = -25.24, SE = 1.12, t = -22.58, p < .0001 for quadratic term and β = -8.66, SE = 1.12, t = -7.77, p < .0001 for quartic term). There was a weak trend towards statistical significance for the quartic term (β = 1.86, SE = 1.12, t = 1.66, p = 0.098). The above suggest that participants' looks to the target picture increased over time, which can be confirmed by visual inspection of the data (Figure 4), although not in a purely linear manner, i.e. the rate of the increase was changed over time.

There were the following significant effects and interactions on the linear term of the polynomial; (1) an effect of group for the first contrast on the linear term (i.e. monolinguals vs bilinguals; β = 3.46, SE = 1.5, t = 2.29, p = 0.022) but not for the second contrast (i.e. native bilinguals vs. non-native speakers; β = 0.90, SE = 1.42, t = 0.64, p = 0.525), (2) an effect of structure (β = 24.96, SE = 1.12, t = 22.35, p < 0.001), (3) an effect of number match (β = -7.50, SE = 1.12, t = -6.71, p < 0.001), (4) a group by structure interaction for the monolinguals vs. bilinguals contrast (β = -4.43, SE = 1.51, t = -2.93, p = 0.003), (5) a group by first NP number interaction for the native bilinguals vs. L2 learners contrast (β = 6.03, SE = 1.42, t = 4.25, p < 0.001), (6) a number match by first NP number interaction across groups (β = -8.54, SE = 1.17, t = -7.66, p < 0.001), (7) a group by number match by first NP number interaction for the bilinguals vs monolinguals contrast ($\beta = -3.02$, SE = 1.51, t = -2.00, p = 0.046) and (8) a structure by number match by first NP number interaction across groups ($\beta = 4.28$, SE = 1.12, t = 3.84, $p < 0.001$).

The effect of group on the linear term indicates different curves in the increase of looks to target for the monolinguals relative to both groups of bilinguals. Visual inspection of the data shows an earlier and steeper increase in looks to the correct picture for the monolinguals relative to both groups of bilinguals (Figure 15 previously). For the subject questions, the monolinguals reach an earlier peak and looks to the correct picture begin to decline earlier relative to the bilinguals from both groups. This can be interpreted as faster processing. This may account for the group by structure interaction in the monolinguals vs. bilinguals contrast.

The effect of structure on the linear term reflects this difference in the increase over time for each structure type; the processing of one is relatively straightforward while processing the other entails initial confusion. For all three groups, looks to target were already above chance for the subject questions in the earliest bin, i.e. 200ms after the auxiliary verb, and increased rapidly. They began to decline after peaking. This reflects the fact the participants have fully processed and understood the sentence they have heard. For objects, looks to the correct picture were below chance for all three groups, suggesting original interpretation as a subject question and they begin to increase. Looks to target start to plateau in the second half of the timespan analysed but the peak followed by a decline is not observed for any of the groups as was the case with the subject questions. Furthermore, the increase in looks to target over time is more protracted for the object questions indicating slower processing even after disambiguation. This pattern is, nonetheless, consistent across all groups. It is only in the last bins that looks begin to plateau and the peak and decline is not observed as it is with the subject questions.

The effect of number on the linear term is due to the steeper curve for number match relative to number mismatch. The effect of number on the linear term is captured in Figure 14.

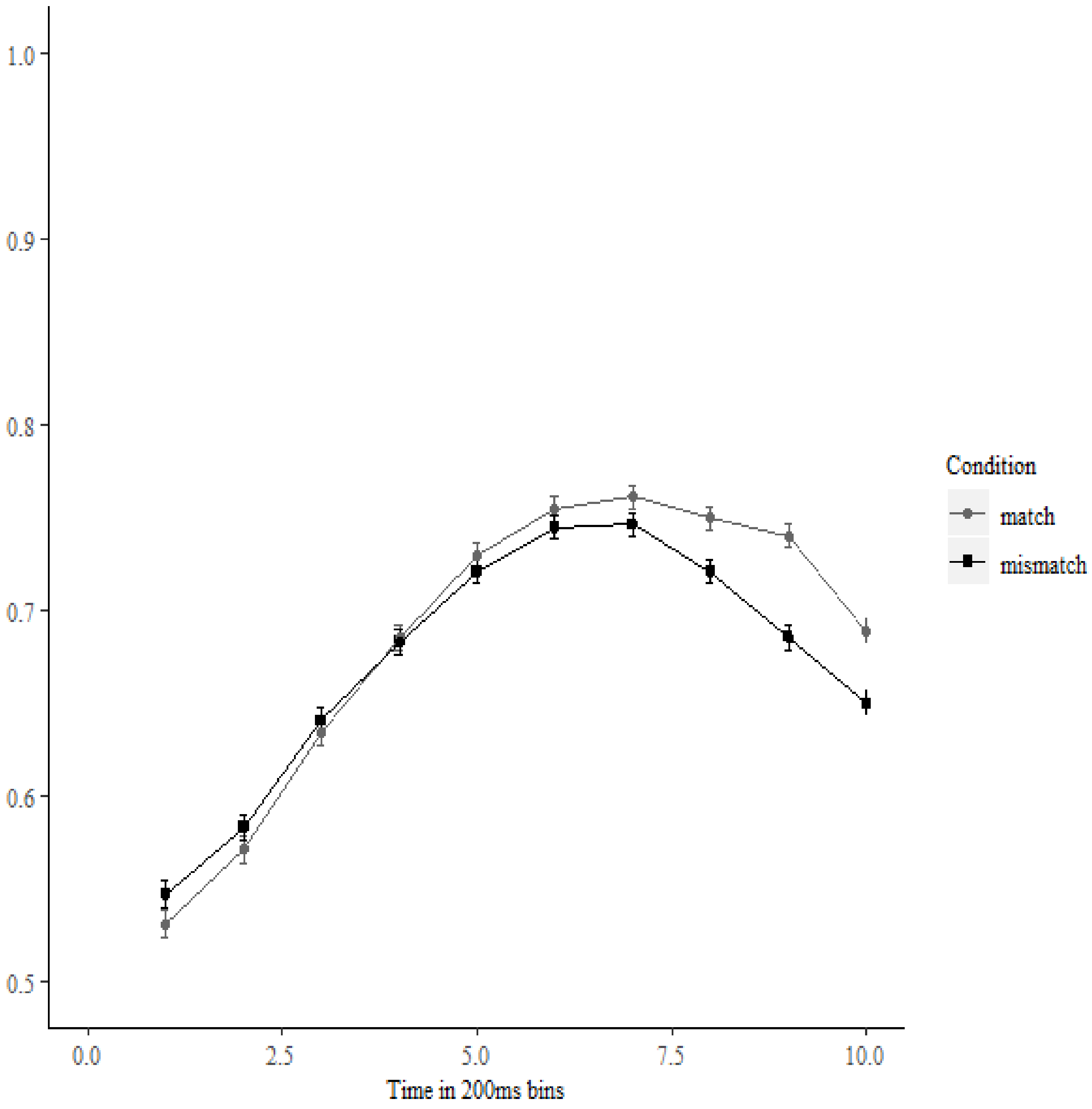

Figure 14 Looks to target over time by number (groups and other variables collapsed)

A steeper increase in looks to target for number matching trials relative to mismatching ones may appear counterintuitive but can be explained by the fact that there is an initial benefit to the number mismatch between the two NPs which disappears after about 500ms. After about

1000ms, the direction of the effect appears to be reversed. This may be due to the fact that looks to target peak and begin to decline earlier for the number mismatch condition reflecting a facilitative effect on processing; as participants have fully processed the sentence, they begin to look at both pictures more equally again as they did for the subject questions. The group by First NP interaction in the contrast between native bilinguals and L2 learners is due to the fact that the L2 learners have a steeper increase in looks towards the target but with a delayed onset in the First NP plural condition. In the First NP singular condition, however, looks to target are fewer for the L2 learners relative to the bilinguals but they increase in parallel as Figure 15 shows.

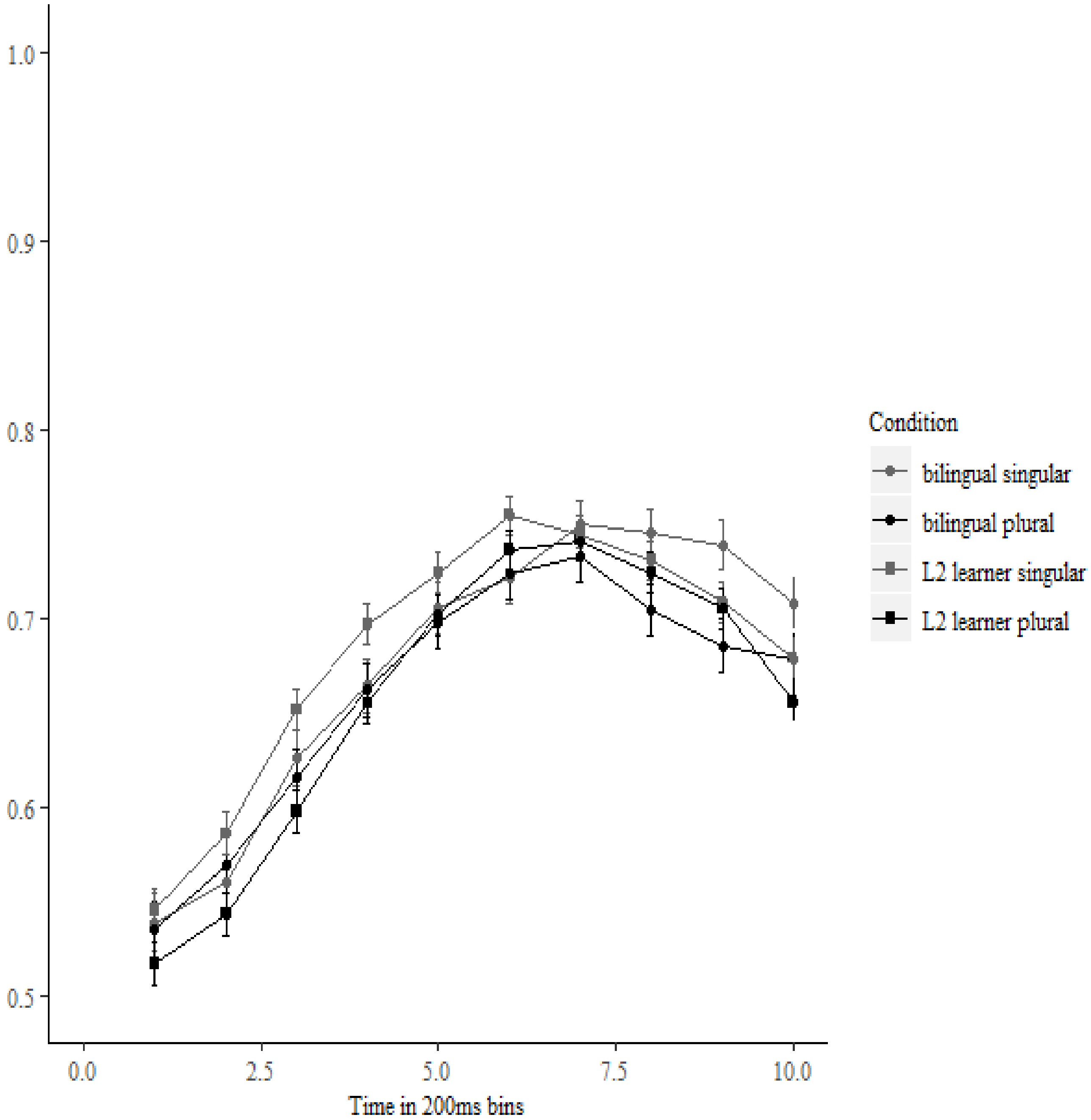

Figure 15 Looks to target over time by group (bilinguals and L2 learners only) and First NP

The number match by first NP number interaction is due to the different trajectory in change for the looks in the match condition when the first NP is plural (see Figure 18 below). This is expected to be the hardest condition; indeed, looks to target are the lowest among all four conditions across groups and show an initially protracted increase. The remaining conditions follow similar trajectories; this is particularly the case for the sentences with the first NP in the singular, although the looks to target peak somewhat higher when in the mismatch condition. The two three-way interactions are reflective of the aforementioned interactions.

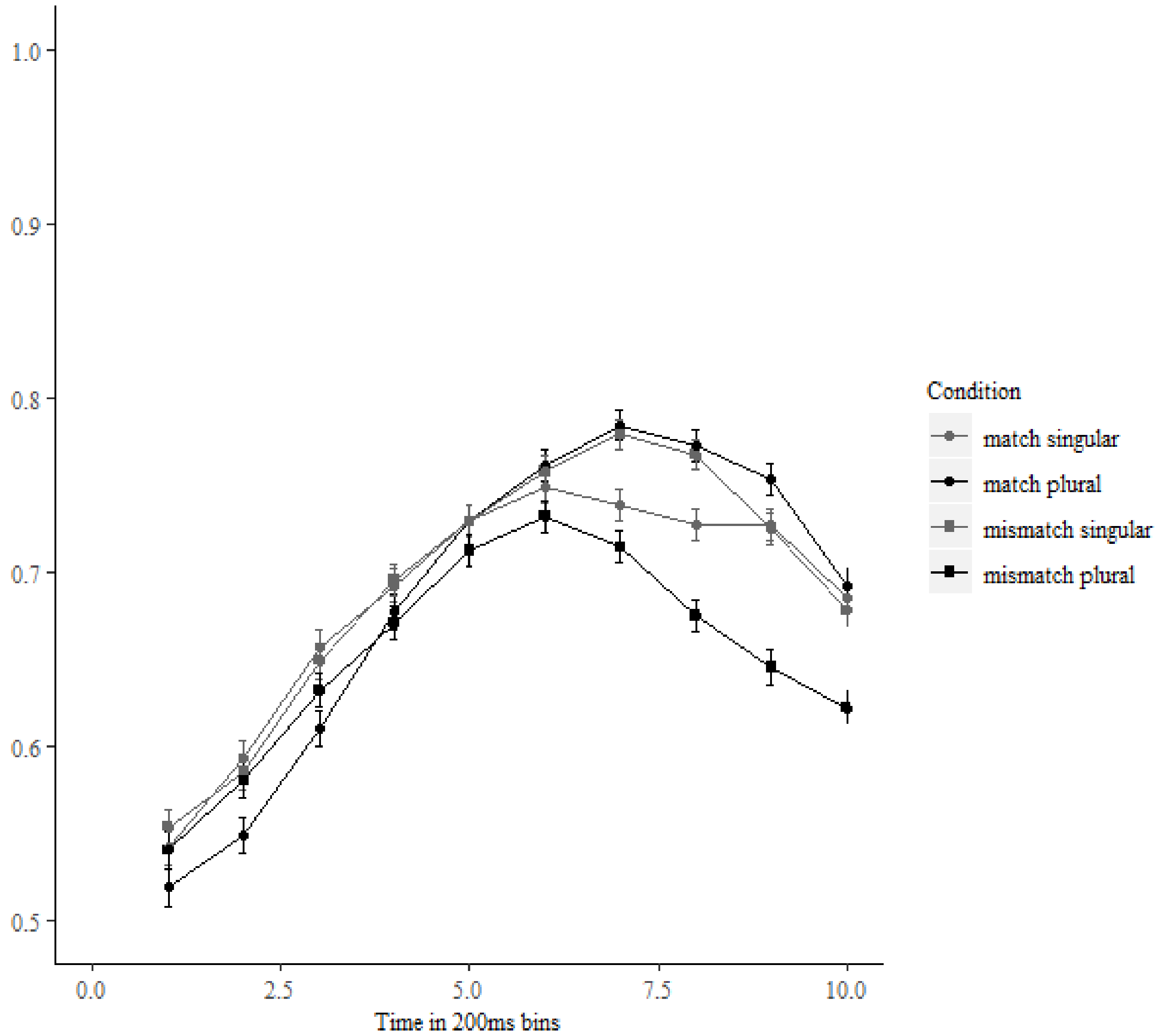

Figure 16 Looks to target over time by number match and First NP (all groups collapsed)

3.3.3.2 Native vs non-native speakers; then monolingual vs bilingual natives

The only main effect on the intercept term was that of structure (β = -0.16, SE = 0.02, t = -9.12, $p < .0001$). There were more looks to target overall when the sentence contained a subject question than when it contained an object question, as with the previous analysis. There was no effect of group for either contrast (native vs. non-native speakers; $\beta = 0.16$, $SE = 0.03$, $t = 0.06$, $p = 0.949$; monolingual vs. bilingual natives; -0.03, $SE = 0.03$, $t = -1.11$, $p = 0.270$). Similarly to the first analysis, there was no effect of number match ($\beta = -0.02$, $SE = 0.02$, $t = -0.89$, $p = 0.395$) or number of the first NP (-0.01, $SE = 0.02$, $t = -0.82$, $p = 0.429$). For the second contrast, the linear, quadratic and cubic polynomial terms were significant predictors of participants' looks to target ($\beta = 42.90$, $SE = 1.12$, $t = 38.41$, $p < .0001$, $\beta = -25.24$, $SE = 1.12$, $t = -22.58$, $p < .001$, $\beta = -8.66$, $SE = 1.12$, $t = -7.78$, $p < .001$). There was a trend for the quartic term, but it was not significant ($\beta = 1.86$, $SE = 1.12$, $t = 1.67$ $p = 0.096$). As with the first analysis, the significance of the three terms shows that looks to target increased with time for all groups and across all conditions but that this did not happen in a purely linear manner.

On the linear term, the following effects and interactions were found: (1) a main effect of structure ($\beta = 24.96$, $SE = 1.12$, $t = 22.35$, $p < .001$), (2) a main effect of number ($\beta = -7.50$, $SE = 1.12$, $t = -6.71$, $p < .0001$), (3) an interaction of group by structure for the second contrast (monolinguals vs bilinguals, $\beta = -4.69$, $SE = 1.43$, $t = -3.28$, $p = 0.001$), (4) a group by first NP number for the first contrast (native speakers vs non-natives, $\beta = 6.80$, $SE = 1.50$, $t = 4.54$, $p <. 0001$) and (5) a number match by first NP across groups ($\beta = -8.54$, $SE = 1.12$, $t = -7.66$, $p < .0001$).

These main effects and interactions of the second comparison mirror what was observed for the analysis where bilingual adults were clustered with L2 learners and compared to monolingual adults first and subsequently to L2 learners (effect of structure on intercept term; more looks to target for subject questions, effect of linear term; increase over time, effect of structure and number match on linear term across groups, likewise for number match by first NP number and three-way interaction).

3.4. Discussion

This study is a follow up to Study 1A but with adult participants. Using the same visual-world paradigm, we investigated the time course of subject and object which-question processing and how adult bilinguals use the feature of number mismatch between the two noun phrases to disambiguate faster.

RQ1: Do bilingual native speakers differ in their performance to monolinguals and L2learners in processing which-questions in English on a range of measures reflecting the end result (accuracy, reaction times, total looks to target picture)?

Accuracy scores were overall lower for object which-questions than for subject questions as they are more difficult to process because they require reanalysis. This is consistent with all studies discussed. All groups had very high scores which approached ceiling effects. The bilingual natives clustered with the monolinguals in terms of accuracy and had a higher accuracy score than the L2 learners. This is consistent with previous studies which have linked higher proficiency and earlier acquisition with more nativelike processing patterns (Dallas et al. 2013; Hopp, 2010, 2017; Pliatsikas & Marinis, 2013b). Yet this should not be seen as the result of the L2 learners not understanding the syntactic construction examined. Pairwise comparison between the L2 learners and the monolingual adults did not reveal a statistically significant difference for overall accuracy. Instead what is maybe the case, is that the L2 learners were less able to recover from the misinterpretation and to reanalyse the sentences they heard. Even the monolinguals do make errors, albeit very few. The group differences were driven by the object which-questions only; separate analyses by structure showed no group differences for subject questions but did for object questions. The L2 learners had lower accuracy scores than the native bilinguals in both contrast schemes used[27]. The bilinguals had lower accuracy than the monolinguals only in the first contrast scheme. The fact the second difference was less robust suggests the difference is smaller. This is potentially reflective of the fact that non-native speakers have greater difficulty reanalysing an ambiguous sentence where they have been garden-pathed (e.g. Pozzan & Trueswell, 2016 for garden-path sentences). The monolinguals showed a weaker subject-object asymmetry for accuracy which was driven by their marginally numerically higher scores in the object questions.

The reaction time data show a different picture. Reaction times were slower for object questions than for subject questions. This is, again, consistent with the literature and with the accuracy scores. The slowing down can be interpreted as the result of misinterpretation as the reanalysis would require additional time for the participants even when this is done correctly. It is the group differences that differ to the accuracy data. Reaction times were slower for the native bilinguals than for the monolinguals but not slower than the reaction times of the L2 learners. Only the bilingual participants (natives and L2 learners) showed an effect of structure whereas the effect was absent in the monolinguals. Given the previous works it is expected that the native bilinguals would cluster with the monolinguals instead of the L2 learners given the higher proficiency. It is unlikely that the differences between the native bilinguals and the monolinguals are due to differences in proficiency as there were no such effects in the accuracy data. One plausible explanation is the activation of the other language in both bilingual groups. It is known that both languages are active at least at the level of the lexicon in both balanced and (non)dominant bilinguals (e.g. Blumenfeld & Marian, 2007; Green, 1986, 1998; 1996; Grosjean, 1997, 1998; Li, 1996; Hartsuiker et al., 2004; Marian &

[27] Contrast scheme 1: Monolinguals vs. all bilinguals, then native bilinguals vs L2 learners – Contrast scheme 2: Native vs. non-native speakers, then monolingual natives vs. bilingual natives.

Spivey, 2003). It is possible than either this has an impact on syntactic processing or that syntactic structures are also active in real time language use.

The third metric relevant to overall processing capacity as an end result is the proportion of looks to the target picture relative to looks to the competitor picture. Overall, there were more looks to the correct picture for the subject which-questions than for the object which-questions. This suggests that participants had greater difficulty processing the object-which questions than the subject which-questions and looked at both pictures more until they had interpreted/disambiguated the sentences. This was the case for both previous studies which used the visual-world paradigm (Contemori et al., 2018; Hopp, 2017) and is expected given the available literature for English. This was consistent across groups and was found in both contrast schemes. There were no group differences between any groups in either contrast scheme. The bilinguals looked towards the correct picture as much as the L2 learners and as much as the monolinguals. The measure for the gaze data discussed in this subsection is that of the total amount of looks toward the correct picture in a time window of 200-2,000ms after the onset of the auxiliary verb in the sentence. This is a relatively large time window and as such may mask effects that have occurred at specific points within it. Therefore, it is more likely that only large and consistent differences will be statistically significant. This may account for the absence of group effects in the gaze data as total amount of looks. As effects at particular time points or differences in changes to looks over time do not speak to the groups' overall capacity to process language, they are discussed subsequently.

In sum, the bilinguals were as accurate overall as the monolinguals overall and more so than the L2 learners in processing which-questions. Furthermore, they looked to the target picture as much as the monolinguals when they heard the sentence. However, they selected the correct picture as slowly as the L2 learners. Moreover, this is across both subject and object which-questions aggregated. The bilinguals, unlike the monolinguals and like the L2 learners however, were more susceptible to the increased difficulty of object questions relative to subject questions.

The data from all three metrics do not suggest a qualitatively different pattern of processing, i.e. less structure-based in nature, than the monolinguals. The differences are predominantly ones of time course. As such they do not lend support to theoretical accounts which postulate qualitatively different processing between natives and non-natives (Clahsen & Felser 2006a, 2006b). Furthermore, the commonalities in patterns of performance between the native bilinguals and L2 learners do not suggest that differences in performance are purely due to proficiency as found in other studies (e.g. Pliatsikas & Marinis, 2013b). The overall numerically high accuracy scores and the absence of overall group effects for accuracy further supports this. A plausible explanation for this pattern is the activation of two languages in parallel during language processing in bilinguals which may require bilinguals (of both groups) to slow down. It is established that both languages are active in bilinguals even when only one language is used (e.g. Jared & Kroll, 2008; Marian & Spivey, 2003). Measures used to demonstrate this have been associated with speed (e.g. naming latencies in Jared & Kroll). While the majority of work related to bilingual language activation is based on lexical access, however, work demonstrating the possibility of cross-linguistic priming points to activation of a syntactic structure in both languages (Hartsuiker, Pickering & Veltkamp, 2004).

RQ2: Does the time course of processing differ between native bilinguals, L2 learners and monolingual adults? Is there evidence of incrementality in processing in bilingual adults?

The linear term was significant indicating that looks to the target picture increased over time following the auxiliary verb, but the significance of the other predictors suggests this did not happen in a purely linear manner. This is the case with both Contemori et al. (2018) and

Hopp (2017). There was an effect of group on the linear term when contrasting monolinguals and bilinguals but not when contrasting native bilinguals vs. L2 learners. This is consistent with the findings from the reaction times. The linear term reflects changes to looks over time, in this case increase; it means that the bilinguals and the L2 learners increased their looks to the correct picture more slowly than the monolinguals. If looks to target are assumed to be a proxy for processing, then the bilingual groups appear to be processing which-questions more slowly than the monolinguals. The result from the native bilinguals is unexpected given the previous literature which suggests that higher proficiency would lead to more nativelike processing patterns. The explanation provided for the reaction times may be applicable in case of the gaze data as well. Both languages are active during language production and comprehension and one language needs to be inhibited. Bilinguals have been found to be slower in tasks that involve vocabulary, and this has been related to the need to inhibit lexical information from another language. A structural equivalent may also be possible.

One should also keep in mind that previous studies have typically defined nativelike patterns of processing as the presence of effects (e.g. plausibility) that are also found in monolingual speakers. The effects of structure reported for the monolinguals are indeed attested for the bilinguals. The effect of structure on the linear term (and all other polynomial terms, for that matter) indicates a completely different time course of processing of object questions relative to subject which-questions. For subjects, looks at 200ms after the onset of the auxiliary verb are already above chance and rapidly increase until they peak. For the objects, they are below chance at the same time point and increase more slowly through the 2,000ms time window without peaking. The initially fewer looks to the target reflect an effect of garden pathing; participants have formed a structural representation with the linguistic material available and have interpreted the question as a subject one. In other words, all groups believed at the point of the Wh-phrase until the second NP that they were listening to a subject and not an object which-question.

The above finding is consistent both with the Active Filler Hypothesis (Clifton & Frazier, 1987) which predicts that the filler will be inserted into the first gap position possible, in this case the assumedly empty subject position. This has been attested in other studies with L2 learners as well (Williams, 2006; Williams et al., 2001; see also plausibility effects among others in Cunnings et al., 2010; Felser et al., 2012). It is also consistent with expectation-based models (Levy, 2008; Roland, Dick & Elman, 2007) which posit that input has guided speakers of a language to expect a subject question as these are more frequent. These two strands of accounts do not, however, make predictions to adjudicate between them that were tested in this study. One challenge to expectation-based accounts in this case, is that they are more input-driven than the Active Filler Hypothesis and the bilinguals and L2 learners have received different input to the monolinguals in terms of quantity and quality. This can be explained on the assumption that the bilinguals and L2 learners only require a quantity of input which amounts to a subset of their entire exposure to English to develop the same expectations as the monolinguals (although, it is unclear whether this is more or less than what the monolinguals would need). It is possible that the L1 of the bilinguals also exhibits an agent first preference. While this cannot be definitively excluded, the participants come from a wide variety of typologically different languages with no more than five participants having the same L1 in each group of bilinguals. Therefore, although properties of the L1 may have some impact on the results, it is unlikely that the effects observed in the bilingual groups are solely a result of the L1.

There is also an interaction of group by structure on the linear term. This indicates that the differences in change to looks between subject and object which-questions was not the same for the groups. In particular, the interaction was found in the monolingual vs. bilingual contrast and not in the native bilinguals vs. L2 learners contrast. This is consistent with what was found for the gaze data overall (intercept term) but also for the reaction times and is somewhat unexpected given the literature which suggests the bilinguals should be clustering with the monolinguals in terms of processing patterns rather than with the L2 learners. What this means is that for the object which-questions, the bilinguals showed a steeper but slightly delayed increase in looks to target than the monolinguals, i.e. they caught up. In the subject questions, the trajectory is similar but looks to target remain lower. It appears that both bilingual groups continue to explore both visual stimuli which results in more looks to the competitor and fewer looks to the target.

RQ3: Do native bilingual adults make use of morphological cues to facilitate processing of which-questions in the same way monolingual adults and adult L2 learners do?

The use of morphosyntactic cues to facilitate processing of which-questions was tested by manipulating the number of the two noun phrases so that they would be the same (match condition) or different (mismatch). Effects of number match were found on several measures. For accuracy, there was no main effect but an interaction with structure. There was a trend for accuracy to be higher in the object which-questions when the number between the two NPs did not match than when it was a match. This did not however reach significance. No such effect or trend was found for subject which-questions. This is consistent with Relativized Minimality (Rizzi, 1990) which predicts that number mismatch will facilitate the processing of object questions and not subject questions due to the effect of the intervener. The number match by group interaction in the first and second contrast schemes for the object which-questions suggests that groups made differential use of cues to facilitate processing. Number mismatch led to higher accuracy in the object questions for the monolinguals but not the bilinguals or the L2 learners. This could be interpreted as a qualitative difference between monolingual and bilingual morphosyntactic processing (i.e. shallow parsing) on its own but needs to be examined in conjunction with the other metrics. One should always keep in mind that accuracy is a measure of the end result of processing. As such, it does not capture the cognitive processes which may have taken place between exposure to the stimulus and the participants' response in the same way that gaze data does. Moreover, accuracy reflects behaviour after time to reflect (consciously or not). Therefore, it may be possible that these effects will not be observed in the accuracy as participants will have been able to overcome any difficulty incurred before clicking the correct picture. A further reason for which these differential effects may not reflect shallow parsing is that it's the monolinguals that are showing greater sensitivity to number match. But this should be the case if the participants have the appropriate syntax. On the other hand, under the Shallow Structure hypothesis, it is the parsing system of the L2 learners – late bilinguals – that is less syntactic in nature and would show greater sensitivity to number match. The interaction is in the opposite direction of what the Shallow Structure hypothesis would predict.

In terms of reaction times, no effect of number was found to be specific to any group of participants as evidenced by the absence of any group by number interaction or effect of group when analysing the data by group individually. There was, however, a number match by first NP number interaction. Reaction times were faster for the mismatch condition when the first NP was in the plural but not in the singular. This may be related to the fact that the questions with a plural first NP may be harder to process and, as such, the aid of a disambiguating number mismatch may be less redundant than in the easier singular first NP. This benefit appeared to be more pronounced for the native bilinguals and the L2 learners in the reaction times than for the monolinguals contrary to the accuracy data.

The gaze data also failed to show a benefit of mismatch for the object questions only. However, all groups seem to make the same use of the morphological cue in orienting their gaze to the correct picture. Looks to target were higher for the mismatch immediately after the auxiliary and shortly thereafter with looks to target rapidly converging. The results are partially consistent with Hopp (2017) in that the benefit of number mismatch was not specific to the object questions. In Hopp, however, the effects were found for the intermediate learners of English and not the advanced ones. In this study, they were found across groups.

The results are not consistent with Relativized Minimality as the benefit of number mismatch is not exclusive to object which-questions. Previous studies have habitually included only structures where there is an intervener with the potential to create competition during retrieval - e.g. Franck et al. (2015) only included object relative clauses. For this reason, it is also not explicitly predicted under memory cue-based retrieval accounts. It could be related to a process of rechecking which is triggered by the need to answer the comprehension question. Something like this is not inconsistent with all retrieval accounts (see Wagers, Lau & Phillips, 2009, who show that mismatch effects in reading times occur in the post verb region suggesting some form of checking of the initial encoding of constituents in the sentence). One should point out that not all studies find effects of number mismatch in grammatical sentences such as the ones the participants heard in the study (Wagers et al., 2009).

The results are partially consistent with the adult controls in Contemori et al. (2018). In both studies the number mismatch benefit disappeared in the gaze data after around 800ms post onset of the auxiliary. The difference is that in Contemori et al. it was present for object questions only. On the other hand, there were no effects of number match on accuracy scores in Contemori et al. with the adults scores reaching ceiling. This was not the case in this study where an effect was found for number match for object questions but did so in the monolingual control only. This study differs from Contemori et al. in that the stimuli were manipulated to include both singular and plural NPs; as such the items may have been more difficult which would account for the lower accuracy scores in the monolinguals.

RQ4: Does the number of the first noun phrase have an impact on the processing of which-questions? If it does, is the effect the same for monolingual adults, bilingual native speakers and L2 learners?

The evidence for a psycholinguistic reality for the linguistic markedness of plurality on the first NP is limited in this study. It emerges on two measures; the number match by first NP interaction for the reaction times where the effects of number mismatch are larger in the plural first NP. The second measure is the interaction of first NP number with group in the bilingual vs. L2 learner contrast on the linear term. This is due to the fact that L2 learners' looks to the target picture follow a different trajectory to the L2 learners' looks to target in the first NP singular condition and the bilinguals' looks to target in both conditions. They are initially fewer and increase more rapidly as they converge, indicating initial difficulty in interpretation which is only temporary. The effects are in the direction that would be expected given linguistic markedness as a theoretical concept. The sentences with the plural first NP are harder for the group one would expect to be less proficient and more likely to diverge from monolingual processing norms. The evidence for an impact based on number of first NP is relatively weak in this study. This, again, is not counterintuitive as markedness would more likely have a greater impact on processing in populations with a developing system and not where ceiling effects are more plausibly expected.

3.5. Conclusion

This study investigated the real time processing of which-question in monolingual adults, bilingual adults and L2 learners. All groups appear to be showing the same patterns of processing but sometimes less efficiently. All groups had greater difficulty with object-questions and appear to erroneously interpret them as subject questions until disambiguation. The adult native bilinguals appear to cluster with the L2 learners in terms of performance in that they were slower than the monolinguals in terms of reaction times and reoriented their looks towards the correct picture more slowly than the monolinguals. Yet in terms of accuracy, the native bilinguals clustered with the monolinguals. Given the previous observation, it is unlikely that the differences are due to proficiency but related to the activation of two languages during processing. Evidence from the object questions suggests a comparative difficulty in reanalysing the ambiguous sentences for bilinguals and particularly for the L2 learners. All groups appeared to make use of the number mismatch to disambiguate but did so for both subject and object which-questions in the gaze data and the reaction times. The effect of number mismatch was limited to the object questions and the monolinguals for accuracy. In all cases, the direction of the effect of number mismatch is in the anticipated direction. Native bilinguals appear to be as accurate as the monolinguals and more so than the L2 learners. Yet for speed they appeared to cluster with the L2 learners as evidenced by the particular effects of group found for the reaction times and the increase in looks to correct picture over time in the gaze data. They also differed from the monolinguals as, similarly to the L2 learners, they had greater difficulty with the object questions. This is not to say that the bilinguals were exactly the same as the L2 learners, as they did not show the same difficulty with plural first NP questions as the L2 learners did. In sum, the bilingual natives appear to be as proficient as the monolinguals yet slower than the monolinguals although all three groups show similar patterns of processing which is not predicted under the Shallow Structure hypothesis.

Chapter 4 Garden-path sentence processing in bilingual children

4.1. Introduction

Overview

The second linguistic feature examined in this PhD is temporarily ambiguous sentences. These are known as garden-path sentences. These are sentences where the interpretation is temporarily ambiguous in terms of syntactic structure but not vocabulary where the reader or listener is more likely to commit to an initial and ultimately incorrect syntactic structure. However, at some point in the sentence, maintaining the incorrect interpretation is no longer grammatical and the reader/listener needs to abandon the initial interpretation and rebuild the syntactic structure (for a more detailed description see Chapter 1). Despite the availability of ample research on how monolingual children process garden-path sentences, the dearth of the respective research for bilingual children reflects the scarcer body of works on how bilingual children process morphosyntax. This study aims to enhance our understanding of morphosyntactic processing in bilingual children by investigating garden-path effects induced by reduced relative clauses, as in previous studies, and whether or not bilingual children can use disambiguating cues to facilitate the processing of these. This chapter begins by outlining previous research on garden path sentence processing in children and discusses sources of information which have been shown to guide children's parsing. Subsequently it provides an overview of the methodology for Study 1B and outlines the findings. It concludes by discussing the findings in relation to the research questions and previous work.

Garden-path sentences in children

Garden-path sentences have been shown to be difficult for children to process in numerous studies and under varying experimental manipulations (Kidd, Stewart & Serratrice, 2011;

Snedeker & Trueswell, 2004; Trueswell, Sekerina, Hill & Logrip, 1999). What appears to be the core difficulty is not building a syntactic representation - this appears to be adultlike - but children's inability to build a second syntactic representation; in other words, revising one's interpretation of a sentence by abandoning the old one and building a new one. The first study to examine garden path effects in children was Trueswell et al. (1999). In this task, participants listened to sentences as in (30a) or (30b) while looking at four objects and needed to act out the instructions in the sentence. A camera recorded their gaze while they heard the sentence. Act out accuracy and looks over time were analysed. Trueswell et al. (1999) used temporarily ambiguous sentences as in (30a).

(30a) Put the frog on the napkin on the table

The prepositional phrase "on the napkin" is temporarily ambiguous as it may be a modifier to the noun phrase "the frog" or, more likely, the destination of the verb *put*. This is in contrast to (30b) where there is no ambiguity as the relative clause is unreduced;

(30b) Put the frog that is on the napkin on the table.

For both ambiguous and unambiguous sentences participants saw a picture of the entity to be moved (i.e. a frog on a napkin), the incorrect destination – what they initially misinterpreted as the destination (i.e. an empty napkin) and the correct destination – the destination in the correct interpretation of the sentence (i.e. the table), and a distractor object (either another frog or different animal).

Indeed, both adult and children participants appeared to be garden pathed in that there were more looks to the incorrect destination compared to control unambiguous sentences such as (30b). Likewise, for the subsequent act out task, child participants moved the frog to the incorrect destination (onto the empty napkin) more often in the ambiguous condition than in the unambiguous condition; children did so on about 60% of the trials in the ambiguous condition but almost never in the unambiguous one. This suggests that, although both adults and children were garden-pathed and initially misinterpreted the sentence, only the adults were able to revise the initial misinterpretation. The difference between adults and children was not exclusively in the effect of ambiguity but in a further manipulation. In this visual-world paradigm study, Trueswell et al manipulated the number of referents in the scene in order to test how participants integrate information available in language processing. In the two-referent scene, there were two pictures of the frog. One was on the napkin, the other one was not. It is expected that in the two-referent condition, the second frog will guide the listener to interpret the PP "on the napkin" as a modifier to the NP "the frog". In contrast, in the one-referent condition, there is only one picture of a frog in the scene and, as such, will not act as a disambiguating cue. Both groups misinterpreted the sentences in the one-referent condition but differed in the two-referent condition; the adults did not show effects of garden pathing in the two-referent condition, but the children did so as they were garden-pathed in the one-referent condition. This was interpreted as reflecting qualitatively similar processing patterns between adults and children (i.e. incrementality in real time processing, consequently garden-path effect) but who differ in the ability to integrate cues to disambiguate in their parsing.

In the initial Trueswell et al. study participants needed to revise sentences over a long distance in terms of both lexical items (linear distance) and a syntactically complex sentence with two prepositional phrases and a reduced embedded clause (structural distance). It is unclear what the impact of this additional complexity would have on processing, especially in children. In other words, it is unclear whether the children have difficulty revising misinterpretations overall or only in sentences with increased structural complexity. A more recent study to examine children's ability to revise syntactic representations is Kidd, Stewart

& Serratrice (2011) who addressed this issue. In Kidd et al. (2011), disambiguation was over a shorter distance with sentences containing a with-Noun PP as in (31).

(31) Cut the cake with the candle.

As with Trueswell et al. (1999), Kidd et al. investigated whether the children could utilise referential information. They tested whether five-year-old children can make use of the plausibility of the noun to determine in a similar manner to adult controls whether VP attachment (instrument interpretation) or NP attachment (modifier interpretation) was appropriate. In sentences such as (31), there is a strong bias based on the verb's semantics for an instrument interpretation of the PP, i.e. for VP attachment. Yet the meaning of the noun *candle* makes that implausible in (31). During this task, participants heard sentences such as (31) and looked at four pictures, two of which were of a cake (one with a candle, the other one without), a picture of a candle and a picture of another object which could be used as an instrument (e.g. a knife). Two measures of looks were analysed; looks to the NP with a PP modifier (cake with candle) and the implausible instrument (the candle in this example) as well as participants' responses in an act out task after hearing the sentence.

Two regions of interest were compared; the second NP (the candle) and the region immediately after the sentence. For the looks towards the complex NP, there was an increase in looks for the adult controls, in other words there were more looks to the complex NP picture in the region after the sentence than at the point of the second NP. This suggests that adults began to interpret the PP as a modifier to the NP rather than to the VP. This was not observed for the children and was interpreted as evidence that reanalysis is restricted to the adults. The same effect, or absence thereof, was found for looks to the implausible instrument. In terms of off-line responses in the act-out component, there were fewer actions which indicated NP attachment, which is consistent with the correct revised interpretation, for the children than for the adults (i,e., cut the cake which has the candle - 26% vs. 40% of responses respectively). This was interpreted as evidence that, whereas the adults entertained both interpretations and ultimately opted for the most plausible one, the children did not do so.

Difficulty revising erroneously interpreted temporary ambiguous sentences does not appear to be limited to English from the limited research available in other language. For example, strikingly similar findings have been found for typological different languages to English. Choi & Trueswell (2010) tested four-to-five-year-old L1 Korean children and compared them to adults using the same eye-tracking with act out paradigm. This study with Korean is interesting because Korean is a head-final language and the word order is hence the opposite of that in English. The verb appears at the end of the sentence and the prepositional phrase will appear before the noun phrase as in (32a) and (32b)

(32a) naypkhin-ey kaykwuli-lul nohu-sey-yo

napkin-Loc frog-Acc put-Hon-SE

"Put the frog on the napkin"

(32b) naypkhin-ey kaykwuli-lul cipu-sey-yo

napkin-Gen frog-Acc pick up-Hon-SE

"Pick up the frog on the napkin"

As can be seen in (32a) and (32b), temporary structural ambiguity is possible in Korean as the phrase "naypkhin-ey kaykwuli-lul (frog on the napkin)" is ambiguous until the verb. In (32a), the PP is an argument of the verb indicating destination of movement. In contrast, in (32b) the PP is a modifier to the noun. Although the word order is the reverse to that in English, the parsing process is assumed to be the same. The listener will build a syntactic representation based on available linguistic material and, if ambiguous, may need to revise it upon encountering incoming material. It is predicted that the parser will initially commit to an instrument interpretation of the ambiguous phrase either because this is structurally simpler (Minimal Attachment) or because the morpheme –ey is more common as the locative than as genitive (constraint-based models). Choi & Trueswell manipulated the verb (put vs. pickup) and the ambiguity of the sentences (ambiguous vs. unambiguous due to the presence of the equivalent “that is” in Korean). In the time windows after the verb region analysed, adult participants’ looks were more towards the destination (i.e. the napkin) for the verb *put* condition relative to *pick up*. This was particularly strong for the unambiguous condition. The children did not show any effect of verb type although there were more looks to the correct destination when the sentence was unambiguous than when it was ambiguous (for adults this was only found for the first out of the two post-verb regions). In terms of actions, the children produced adultlike responses in all but one condition, ambiguous pick up where the children produced an NP action on 46% of the trials (100% for adults). The results suggest that the children had difficulty revising their initial preferences even in the light of disambiguating evidence to the contrary which the adults do. This is in line with other studies but also indicates that the difficulty does not lie with the distributional properties of the verb. Instead, as the verb is the last phrase in the sentence, the misguided interpretation is built on the available NPs. This suggests the garden-path effect is due to cognitive strategies which cannot be revised. Nevertheless, six out of 16 children did not make any errors in this condition. It is Therefore, plausible that the ability to revise initially wrong syntactic interpretations is beginning to emerge in children.

It is possible that the inability to revise syntactic representations is an order of mention effect (Meroni & Crain, 2003). Under this account, children begin to plan a movement upon hearing the instruction put before hearing the entire sentence. This means they are unable to inhibit the original plan upon hearing new material. One study which has shown young children to be able to use referential context is Meroni & Crain (2003). Furthermore, they claimed that children incorrectly assumed that the frog that was not on the napkin was the intended referent to be moved. In this study, the two frogs in the two -referent condition were on different colour napkins and the participants were asked to close their eyes while listening to the instructions. Meroni & Crain found adultlike accuracy when the colour of the two napkins was different and claimed that previous methodological design prevented five-year-old children from inhibiting the original syntactic plan. However, for Meroni & Crain there was no comparison between one- and two-referent conditions and no comparison between the closed and open eyes condition.

Most studies available include children about five years old and are small scale in participant numbers. One larger scale study that examined garden-path sentences in children of various ages (five, eight and eleven years) is Weighall (2008). This study attempted to address some of the limitations of Meroni & Crain (2003); children took part in an act out task in the same paradigm as Trueswell et al. (1999), a similar condition but with the children having their eyes closed while listening, a pragmatic block condition where the two referents were placed against a different colour background (two napkins of a different colour) following (Meroni & Crain, 2003) and a condition which combines the previous two. The two older groups of children responded in a way that reflected modifier interpretation of a PP significantly more often that the five-year-olds. Overall, all groups showed lower accuracy for the two-referent condition suggesting that the children still experience lasting garden-path effects. The impact of referential content was different by presentation block as there was no effect of number of referents in the pragmatic block and the combinatory block, but this was consistent across age groups. This is in line with Trueswell et al. (1999) as the younger children were unable to use referential information to disambiguate early on. On the other hand, it does not replicate Meroni & Crain (2003) who found that even young children can reach adultlike accuracy with two referents under certain facilitatory conditions[28] and there was no overall difference between the conditions under which participants were exposed to the stimuli.

The ability to process garden-path sentences does not always align with children's ability to produce sentences with NPs which function as modifiers to other constituents. Hurewitz, Brown-Schmidt, Thorpe, Gleitman, & Trueswell, J. C. (2000) investigated both production (elicited production) and comprehension (act out) in children. In the production task, five-year-old children heard stories with content similar to the experimental items in previous studies and were given a question as in 33a and 33b.

(33a) Which frog went to Mrs Squid's house? (specific question)

(33b) What kind of cookies did the animals bake? (general question)

In the comprehension component, children were given instructions and had to act them out. Hurewitz et al. (2000) found a striking asymmetry between production (ca 70%) and comprehension (ca 30%). In a second study, Hurewitz et al. (2000) replicated the eye-tracking study from Trueswell et al. (1999) using the experimental items from the first study alongside a production task but also controlled for the visual stimuli. In one condition, the stimuli were as in Trueswell et al. (2000) but in the other there were physical attributes which differentiated between the two referents. The presence of perceptually salient stimuli resulted in an increased performance, in line with Meroni & Crain (2003) and Weighall (2008), but the asymmetry observed in Study 1A persisted.

There is a consensus that children get garden-pathed just like adults do. It is also fairly uncontroversial that children do not recover from garden-path effects with the same ease as adults. What is not agreed upon, is the reason for which children, and speakers of a language in general, experience these effects.

[28] Although it is unclear how much the presence of the second referent was driving the benefit, given they did not compare one- and two-referent conditions.

What guides parsing of garden-path sentences?

One way of accounting for the garden-path effects is known in the literature as Minimal Attachment (Frazier, 1987). This is postulated as a cognitive heuristic, a tendency for speakers to attach/integrate incoming material into the syntactic structure in a way that keeps the syntactic structure as simple as possible with no embedded phrases. Thus, the PP "on the napkin" is attached onto the VP. If, instead, the PP had been attached to the NP, this would have created an embedded clause within the main clause which will be harder to process. It is assumed the parser will attempt to build syntactic representations with minimal complexity unless needed. An alternative account is lexicalist in nature and posits that the PP "on the napkin" will be interpreted as the destination of the verb "put" because the specific verb, by nature of its meaning, always occurs with an NP as its destination. As both children and adults are assumed to be sensitive to statistical information about the distribution of words, i.e. what other words, phrases or structures will occur with them, they have come to build up an experience-based expectation that the verb in question will be accompanied by an NP with a specific syntactic position and semantic function. It is not possible to disentangle the two accounts based on the original study by Trueswell et al. (1999) as put was the only verb used.

This limitation was subsequently addressed in another study by Snedeker & Trueswell (2004). Following the same experimental design with the visual-world paradigm and acting out as in Trueswell et al. (1999), Snedeker & Trueswell manipulated the verb heard by the participants so that it would be either statistically not biased towards a particular interpretation as in (34) or that it was either biased towards an instrument bias as in (9) or a modifier to the noun bias as in (35).

(34) Feel the frog with the feather.

(35) Tickle the pig with the fan.

(35) Choose the cow with the stick.

As with Trueswell et al. (1999), Snedeker & Trueswell (2004) manipulated the referential context so that either 1 or 2 referents appeared in the visual stimuli. The sentences were structurally simpler as they did not contain a second prepositional phrase. The results suggested that both adults and children behaved in a way consistent with the verb's bias. In other words, actions involving the instrument and looks towards the instrument were significantly higher in the verbs with an instrument bias than those with a modifier bias or those with no bias. Crucially, only the adults showed an increase in correct act outs and looks to a correct picture in the two-referent condition. In the children, there was no effect of the referential context in the act out or the looks in the earlier time window – although there were some effects of referential context in the later time windows. This suggests that it is the lexical properties of the verb which bias listeners towards building certain syntactic representations even before all the necessary linguistic content is made available for this representation to be built. A second implication is that this lexical bias also holds in children with their only difference to adult parsing being that children do not appear to make use of the second referent to disambiguate.

Snedeker & Trueswell (2004) controlled for the semantics of the verb bias. However, other aspects of lexical semantics may also function as cues towards interpretation. One large-scale study which addressed this question and investigated the effects of semantics was Kidd & Bavin (2005). In this study, they manipulated verbal semantics (activity vs. perception), noun definiteness and preposition (*with* vs. *on*) in ambiguity resolution. In the first study, participants heard sentences as in 36a-d

(36a) The woman waved to the man with the flag (activity verb – definite noun)

(36b) The woman waved to a man with the flag (activity verb – indefinite noun)

(36c) The spy saw the policeman with the binoculars (perception verb – definite noun)

(36d) The spy saw a policeman with the binoculars (perception verb – indefinite noun)

Furthermore, this study tested children aged five, seven and nine years. Overall, the children selected a picture more often with an NP interpretation with a perception verb than an action verb. For perception verbs, there were fewer NP-attachment interpretations in the five-year olds than the other two groups regardless of the noun's definiteness. For action verbs, NP-attachment was more frequent in the seven-year-olds with indefinite nouns but that was the case with definite nouns in the nine-year-olds. There were fewer NP–attachment interpretations with the preposition *on* but the effect of definiteness was restricted to the oldest group. The above suggests that older children became more able to use the specificity conveyed by the definite article to aid disambiguation, but that interpretation is guided by verb type across ages. In a second study using picture selection and act out with three- and five-year-olds, Kidd & Bavin used nouns which were implausible with certain interpretations given the semantics as in 37a-c.

(37a) Tickle the hippo with the feather. (implausible NP-attachment, VP/instrument-bias)

(37b) Find the dog with the stick. (implausible VP-attachment, NP/modifier-bias)

(37c) Chop the tree with the leaves. (conflict-bias, bias changes in the sentence)

Overall, VP-attachment/instrumental interpretation was more common in the younger children and in the VP-bias condition relative to the other two conditions. It is in these other two where the younger children continued to show a preference for instrumental interpretation / VP-attachment relative to the older children. The fact that this happens in the conflict-bias condition suggests that the younger children are less able to revise their initial syntactic representations than the older children. This ability must, therefore, emerge with age.

In sum, the evidence so far available suggests that children process sentences in a qualitatively similar manner to adults in that they build syntactic representations in real time rapidly with the linguistic content available. Therefore, as with adults, syntactic misinterpretations are possible. Where children appear to differ is in their ability to use disambiguating contextual cues in real time to avoid garden pathing effects and in their ability to abandon their initial erroneous interpretations in favour of ones formed based on novel incoming material. This appears to be the case in more than one language and irrespective of the manipulations of the task and experimental stimuli. It is unknown, however, whether any of the above observations about sentence processing in children also hold true for bilingual children. For bilingual children, there is an emerging consensus that sentence processing is qualitatively akin to that of monolingual children and that the difference lies in bilingual children having greater difficulty in using linguistic cues to facilitate processing (see Chapter 1 for a review on the studies by Chondrogianni and Marinis). Furthermore, studies with bilingual children on sentence processing have used word-monitoring, self-paced listening or reading and have focused largely on detection of agreement violations in morphosyntax. With the exception of gender in Dutch, bilingual children have been shown in these studies to follow the same processing patterns as monolingual children. However, the existing literature leaves two gaps. The studies available have tested how bilingual children deal with ungrammaticality in their additional language when they encounter it. They do not address how bilingual children deal with ambiguity either when they encounter it or subsequently after encountering disambiguating material.

To date, the only study to examine how bilingual children process sentences with temporary structural ambiguity is Marinis & Saddy (2013)[29]. In this self-paced listening task with picture verification, participants heard active and passive sentences. These are ambiguous as it is not clear what the thematic role of the first noun phrase is until the verb phrase. As the parser is likely to assign an agent role to the first NP, the passive sentences are likely to be initially misinterpreted as active sentences. Consequently, reanalysis is expected once the sentence is disambiguated. The results show that both monolingual and bilingual children slowed down in the self-paced reading task at the critical segment (VP and second NP or by-NP) for sentences where the picture did not match the picture the participants were looking at. This suggests that both groups of children could utilise grammatical cues on the verb (-ed and –ing) as well as the by-NP to guide parsing. However, at the subsequent segment, the bilingual children continued to show longer reaction times for the passives relative to the actives. This asymmetry was absent in the monolinguals. This suggests that the bilingual children had difficulties with reanalysis of passive sentences which persisted beyond the point of disambiguation whereas the monolingual children did not show this difficulty.

To our knowledge, there are no studies available which explore bilingual children's ability to utilise referential context, i.e. one referent vs. two referents in the visual context, in the same way as Trueswell et al. (1999) did in their seminal study. While Trueswell et al. found that children, unlike adults were not able to use referential context to aid disambiguation, it should not be taken for granted that this will be the case for bilingual children. Research from bilingual adults, albeit focusing on native vs. non-native speakers, indicates a greater reliance on contextual cues than for monolingual adults (Pan & Felser, 2011; Pan, Schinke, Felser,

[29] Roesch & Chondrogianni (2016) investigates the use of cues to disambiguate by bilingual children but is actually an offline study.

2015). It is unclear whether this also holds for bilingual children and with more naturalistic exposure to the second language than in studies on L2 processing.

A second gap is methodological; studies on real time processing have relied on reaction times as a measure of difficulty while the word-by-word self-paced presentation of the stimuli is not the most natural. Moreover, there is limited work available with eye-tracking on language processing with children which could complement existing studies. It is, therefore, an open question as to how they will perform in a visual-world paradigm study for these garden-path sentences.

The current study

This study builds on existing work to investigate the processing of garden-path sentences in bilingual children. The motivation for this study lies in the open nature of the question about morphosyntactic processing in bilingual children. Bilingual children may differ from their monolingual peers in terms of productive language at any given age but also in their developmental trajectories (e.g. Hoff, 2013). The picture is much less clear for receptive language and even less so for real time processing. There is an emerging consensus that bilingual children process grammatical structures in a qualitatively similar manner as monolingual children (see Chapter 1 for a review). What remains unclear or appears subject to constraints to a greater degree than it is in monolinguals is their ability to integrate morphosyntactic information in real time for purposes of disambiguation and/or prediction (Lew-Williams, 2017; Lemmerth & Hopp, 2018; Marinis & Saddy, 2013). A further motivation for examining garden-path sentence processing in bilingual children is the available empirical evidence in bilingual adults (reviewed in Chapter 5). While bilingual adults show sensitivity to ambiguities and the need for them to be revised, they have been shown to be less able to abandon the original representations (e.g. Pozzan & Trueswell, 2016). It is currently an open question as to what the case with bilingual children is.

This study is a follow up to Trueswell et al. (1999) with older monolingual children and bilingual children, aged 8-11 years. The motivation for including older children in this study is that, given the potential for a lag in language development in bilingual children in the early years, to include bilingual children in the same age range as in Trueswell et al. - at least, as a first step – may have yielded results which are confounded between differences in bilingual development relative to monolinguals and differences in processing between the two groups. A second reason for doing this, is that Trueswell et al. show sensitivity to ambiguity in five-year-old children but not to referential context; it is an empirical question what the case in later years is and at what age children start to use this information in real time. Studies that have tested a large number of participants and grouped them by age (Kidd & Bavin, 2005; Weighall, 2008) have indicated that by age 8 developmental changes have already occurred in the processing of garden path sentences and as such have become more, although not completely, adultlike.

For the purposes of this study, ambiguity and referential content (number of referents in visual stimuli, i.e. one or two) were manipulated to examine if bilingual children experience garden-path effects and whether they can use referential content to disambiguate as they hear the sentence. The semantics of the verb were not manipulated as was the case with Snedecker & Trueswell (2004), as it is not the aim of this study to explore the source of garden-path effects (propensity towards building the simplest syntactic structure or driven by lexical semantics). Instead the verb *put* was used for all trials. Unlike other studies, there was no act out as a comprehension measure; instead participants had to click on a picture to answer a comprehension question. Additionally, this study does not use the imperative, but participants listened to declarative sentences which contained the answer to the comprehension question. There is no theoretical motivation to expect that this will impact garden-path sentences, in fact it may have somewhat more ecological validity. It can, however, test the possibility put forward in Meroni & Crain (2003) that the effects can be accounted for under an order of mention effect particularly applicable to instructions.

Our research questions are therefore:

1. Do bilingual children process temporarily structurally ambiguous sentences, in this case reduced relative clauses, in the same way as monolingual children?

In other words, do they opt for structurally simple and lexically biased VP attachment interpretations and experience garden-path effects? Are the magnitude and/or the time course of the garden path similar in both populations?

2. Do bilingual children recover from garden-path effects in the same way, i.e. equally successfully (or unsuccessfully) and over the same time course as monolingual children?
3. Do bilingual and monolingual children utilise number of referents as a disambiguating cue in interpreting garden-path sentences? If they do so, are there differences between the two groups as an end result or in the change of gaze data over time?
4. The research questions are addressed with a range of metrics, namely accuracy, reaction times and gaze data either aggregated in a critical region or increasing/decreasing over time within a specific time window.

Given the available literature it is expected that the bilingual children opt for VP attachment and be garden pathed. However, it is also expected that they will show sensitivity to the novel incoming material and revise their misinterpretation. The existing literature does not, however, make any explicit predictions about how rapidly they will alter their syntactic representation. In terms of referential context, it is likely that the monolingual children will be able to use the presence of two referents in the visual stimuli on the grounds that developmental changes between the ages of five and seven/eight have been shown with regards to this specific linguistic feature. It is less clear what will be the case with the bilingual children. In light of the existing literature on bilingual language processing, albeit for non-native adult speakers, it may be possible that this effect is enhanced relative to the monolinguals (Pan & Felser, 2011; Pan, Schinke, Felser, 2015). It is therefore, possible that the bilingual children will be more sensitive to referential content than the monolingual children and behave differently to the children in Trueswell et al. (1999). The gaze data obtained and how the looks to the (in)correct destination change over time after hearing a critical segment reveal the extent to which the participants are garden-pathed and their recovery from the wrong interpretation. The end accuracy of the response to the comprehension question and the speed needed to do so (reaction times) reflect the overall ability to recover from the garden path effects.

4.2. Methods

For the purposes of the second component of this PhD, a visual-word eye-tracking study is used as with the studies on which-questions. Participants looked at pictures while listening to a sentence and then needed to answer a question by clicking one of the pictures

4.2.1. Participants

The monolingual and bilingual participants who participated in this study are the same participants in the study on which-questions. For more details on their demographic characteristics and linguistic background, see Chapter 2. One should note that the bilingual children are of various linguistic backgrounds. If they differ in word order to English, then it is unclear how this will impact the processing of garden-path sentences and most importantly how the findings could be interpreted. To address this, we consulted the World Atlas of Language Structures (WALS) for the order of the relative clause relative to the noun modified to examine whether most children had a first language where a relative clause which followed the noun and therefore, a VP NP PP PP sequence as in the experimental sentences would be theoretically possible. All children except two (L1 Chinese and L1 Turkish, both highly proficient in English), spoke a language with a similar structure to English (noun followed by modifying relative clause, Dryer, 2013). This is more biased sample towards being similar to English in this respect than world’s languages overall (ratio of about 4:1) but speakers of these language do not make up sizeable communities in the UK where the children were recruited. Crucially, given the findings from Choi & Trueswell (2010), this should not impact their processing of English garden-paths.

4.2.2. Design

Participants looked at a series of four pictures and listened to a sentence at the same time. An example of the visual stimuli for a trial is shown in Figures 17 and 18.

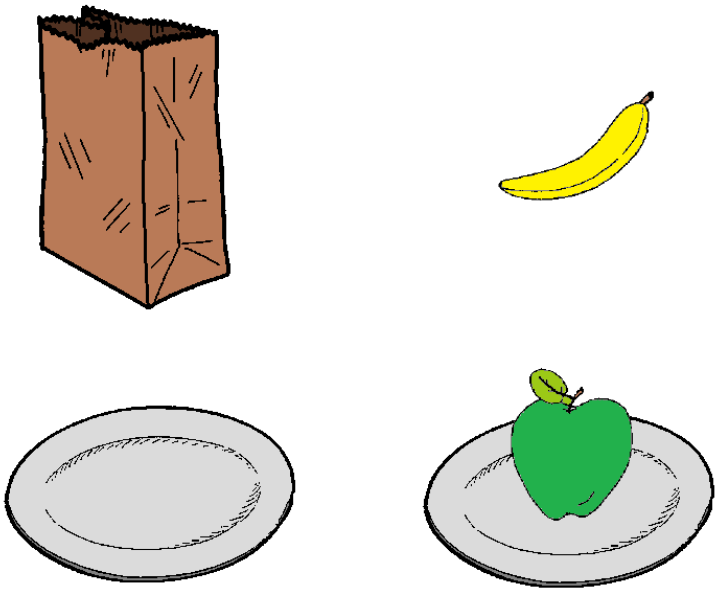

Figure 17 Sample visual stimuli for trial in 1-referent condition

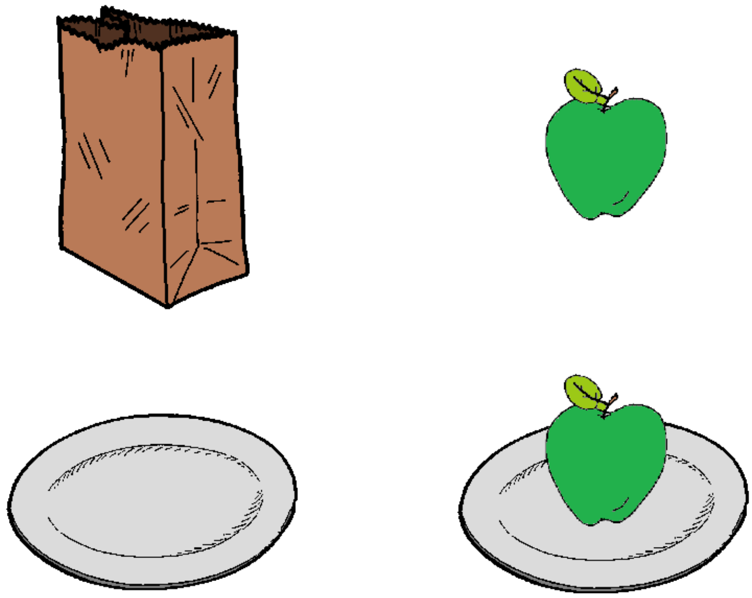

Figure 18 Sample visual stimuli for trial in 2-referent condition

In terms of content, the sentences involved moving a specific inanimate entity placed onto another entity towards a location as shown in (38a) and (38b).

(38a) Peter put the apple on the plate in the bag before going to school

(38b) Peter put the apple that's on the plate in the bag before going to school

Following the sentence, the participants heard a question relating to the sentence they just heard. For these sentences the relevant question was (39)

(39) Where did Peter put the apple?

To answer the question, the participant needed to click on one of the four pictures. The question was the same for both ambiguous and unambiguous sentences with the same lexical items across 1-referent and 2-referent conditions. The purpose of the comprehension question was to specifically tap on the correct interpretation of the correct ambiguity. This study differs from Trueswell et al. (1998) in that the ambiguous sentences were in the past tense rather than in the imperative. This was done this way as the study was not designed as an act out task which would not have been possible given that the eye-tracker used requires the participant to remain seated and avoid head movement. Moreover, E-prime does not have the facility to allow participants to move images by dragging them.

In the case of the study in question, the visual stimuli are pictures of single concrete entities which may be animate or inanimate but do not have internal event structure. The images used in this task correspond to different entities in the sentence heard and may have a thematic role or function as adjuncts. In fact, they are needed as the participants hear sentences which contain four entities which have a different semantic function. The four candidate images in this study were the object which needed to be moved (i.e. in the aforementioned example an apple), the original location which may be misconstrued as the destination (a plate), the location-destination (a bag), an image of the object at the original location (e.g. an apple on the plate) and/or an additional object (e.g. a banana for this item).

The audio stimuli were recorded by a male native speaker of English in a noise-proof sound booth. Intonational contours were not controlled for. The speaker read the sentences out several times at a slow rate with pauses between segments. The sentence with the fewest and least noticeable contours was used in the experimental task.

The between-subject comparisons are the same as with Study 1A in Chapter 2, namely monolingual vs. bilingual children. This study investigated whether there is greater difficulty in processing sentences with temporary syntactic ambiguity relative to those which do not, i.e. (38a) and (38b) respectively. Moreover, it manipulates whether the participants saw only the entity once or two times (henceforth 1-referent ambiguous and 1-referent unambiguous versus 2-referent ambiguous and 2-referent unambiguous respectively); i.e. whether they saw another semantically related object (e.g. the banana) or the apple again, although the auditory stimuli remained the same. This gives rise to a 2x2 experimental design. There were 16 experimental sentences, with participants seeing 4 trials per condition in a Latin-square design. All experimental materials are included in the appendices. Experimental sentences had 20 filler sentences dispersed among them. The relatively low number of trials per condition is due to the decision to use a larger number of fillers relative to the experimental trials (which in turn increases the length of the task) to avoid participants becoming familiar with the structure of the upcoming sentences and building a correct representation of the incoming sentence on the basis of previous exposure. A further reason for this is the fact that the length of the task would become excessive if the participants saw all items in all conditions. Accuracy, reaction times and looks towards the correct or incorrect destination image were obtained and used as dependent variables. It may be argued that 4 trials per condition is relatively limited. It is, however, the same number of trials as in Trueswell et al. (1999) as well as in other related studies (e.g Snedeker & Trueswell, 2003) but fewer than in Weighall (2008). However, Weighall manipulated mode of presentation; the number of trials per presentation mode is lower in Weighall than in the present study. The number of trials is discussed as a potential limitation in Chapter 6.

4.2.3. Materials

Sixteen inanimate objects were selected alongside 32 nouns, which were locations where these objects may be found or placed (e.g. table, desk etc.). Two statements were then written; one where the destination is originally ambiguous and prone to be misinterpreted as the location as in (38a) and one where it is not as in (38b). For each of these sets, four pictures were then created to depict: a) target i.e. the apple), b) the object's location without the object (Incorrect Destination, "ID", i.e. the plate), c) the destination of the object (Correct destination, "CD", i.e. a bag), d) a distractor which may have either been another apple in the

2-referent condition or something else in the 1-referent condition. The position of these objects varied across experimental sentences so that no type of image appeared in one position only. A question was then written for each of the sets about the identity of the object which was moved; this began with *where*; this was selected as it tapped revision of the temporary object

20 filler sentences were also included. All but one used a transitive verb and one used the ditransitive verb *give*. The subject of the sentence was always animate, and the object was always inanimate. A location was included in each sentence which may have been part of the verb's argument structure or not. This was either the destination or location. A question was also used similarly to the experimental items about the identity of the object moved. A similar comprehension question was created for the filler items as with the experimental ones. The visual stimuli for twelve trials contained four inanimate entities similarly to the images for the experimental trials. These either depicted the object or a semantically related competitor (mirroring the distractors), a potential location with or without on object close to it (target or ID/CD respectively). Additionally, a competitor location was also used. The distribution of each of these images varied across the filler trials. The images for the rest of the trials contained two animate entities (subject and competitor) and two inanimate ones (object and competitor). Two practise trials were constructed similarly to the fillers.

Four lists were created in a Latin-square design. Subsequently the same list was replicated but the order of the trials was altered to create 8 versions in total, so that each trial in an ambiguous condition, such as (9), occurred either before or after the same trial in the unambiguous condition when the number of referents they saw was the same number. Each version contained four trials for each condition while the filler and practise trials remained the same.

The ambiguous and unambiguous trials differed in the auditory stimulus the participants heard but the visual stimuli were the same across lists. For the ambiguous sentences, participants heard sentences as in (38a), whereas in the unambiguous conditions, they heard sentences such as (38b). On the other hand, the 1-referent and 2-referent conditions differed from one another in terms of visual but not auditory stimuli. For the 1-referent conditions, participants saw a novel but semantically related entity as a distractor (e.g. a banana), whereas in the 2-referent conditions they saw a distractor image which was the same entity in a different position was (e.g. apple but not on a plate). Filler trials occurred in the same positions across all eight lists and the target area of interest was the same for any given trial in all lists.

4.2.4. Procedure

The experimental procedure for this task is similar to the eye-tracking task with the which-questions. The participant sat in a quiet room wearing headphones at about 80cm from a laptop screen and a Tobii X-120 eye-tracker. Participants undertook a five-point calibration and began the task once the experimenter ensured the eye-tracker could reliably record gaze data. A fixation cross in the centre of the screen appeared before the onset of each trial which participants needed to fixate upon for 1000ms for the trial to begin. This was done for purposes of calibration in order to reduce the risk of track loss but also to ensure that participants continued to pay attention. Following this, the trial began. Participants heard a statement and saw four pictures of equal size in the four corners of the screen. Subsequently they heard the comprehension question and needed to click one of the pictures in order to answer the question. The position of the correct picture to answer the question was varied throughout the task but not across lists and was counterbalanced across conditions. The order of the stimuli was pseudorandomised so that no trials from the same condition occurred immediately after another one. All experimental trials were preceded and succeeded by fillers. Fillers typically occurred individually but some occurred adjacent to another one or two. Participants undertook the two practise trials prior to commencing the main task; the procedure was the same and no new calibration took place. E-prime recorded the participants' responses for each trial and their looks to each of the four images for each time frame which was captured 120 times per second.

4.2.5. Analyses

Mixed effects regression models were fitted for accuracy, reaction times and gaze data. The fixed effects structure included group as between-subjects comparison and ambiguity and (referential) context as within-subject independent variables. Random slopes for all main effects and their interactions were permitted. Random intercepts included subject, item (pairing of entity with the correct and incorrect destination) as well as item-ambiguity (defined as the pairing of lexical items within an ambiguous or an unambiguous sentence) and for the gaze data only, trials identity, defined as the unique pairing of participant by trial to capture non-independent data by participant over time. When the maximal model failed to converge, it was simplified by removing random slopes whose contribution to the variance was limited until it converged.

For accuracy, participants' responses were coded automatically in E-prime as either 1 (correct) when the picture selected corresponded to the correct destination or 0 when it did not. Accuracy data were analysed with a binomial mixed effects model. Data about which picture specifically was selected was not recorded and as such cannot be analysed. It is possible that the children selected the incorrect destination in the incorrect trials more often than they selected any other competitor picture.

For reaction times, inaccurate trials were not analysed and reported. The unambiguous sentences were longer than the ambiguous ones because of the disambiguating words "that's". As a result, the duration of the auditory stimuli in the unambiguous sentences was 500ms longer than in the ambiguous sentences. To eliminate the difference in length between the ambiguous and unambiguous trials, the length of auditory files was subtracted from the reaction times automatically logged in E-prime. By subtracting this time, the remaining time is a clearer reflection of the time the participants needed to process the sentence after it has been heard as a whole. A linear mixed effects model was fitted to the computed reaction times following the same rationale as for the accuracy data.

Gaze data were calculated and analysed in the following way. Four areas of interest (AOI) were defined a priori in E-prime capturing the left and right half top and bottom halves of the screen, corresponding to each picture presented in each trial. Eye-movement data were time locked to the onset of particular phrases: the incorrect destination ("on the plate"), the correct destination ("in the bag"), and the segment following disambiguation until the end of the sentence ("before going to school"), defined as the final region. The incorrect destination region is the earliest point in the sentence where effects of garden-pathing may be observed. The second region is perhaps more critical as it is the point at which disambiguation occurs. The final region was selected to potentially capture any delayed effects or any effects which may occur later on in the sentence as the result of syntactic reanalysis. It is expected that looks towards the incorrect destination picture will reflect the degree of syntactic misinterpretation/garden-pathing, in particular in the ambiguous region, while looks to the correct destination picture will reflect successful reanalysis, disambiguation and real time comprehension; the latter is anticipated in the region of disambiguation and thereafter. Both looks to the correct and incorrect destination were computed for all three temporal regions.

A window of 200ms was allowed for the time it takes to program a saccadic eye-movement and as such any movement prior to this would not reflect linguistic processing related to the stimulus (Matin, Shao & Boff, 1993). Eye-movements were analysed for a period of up to 1,600 milliseconds (200-1800ms post phrase onset) for each region. The time period examined was divided into equal bins of 200ms.

For each bin, two measures were computed; the proportion of looks towards the incorrect destination (i.e. "the plate") and the correct destination ("the bag") relative to the total looks recorded to any of the four pictures. Trials where looks to none of the pictures exceeded 70% were removed from further analyses. This may have been due to participant looking away or due to track loss but was treated in the same way by the Tobii eye-tracker (namely "out"). As a practice, this is consistent with other studies, for example Atkinson, Wagers, Lidz, Phillips & Omaki, (2018) – although the adopted threshold in that study was 65%.

It may be argued that looks to the correct and incorrect destination are the mirror image of one another. This is not completely the case. In Studies 1A and 1B where participants needed to look at two pictures, this was indeed the case. An increase in looks to the target entailed a decrease in look to the competitor, once the looks coded as "out were removed, and vice versa. In this study, participants saw four pictures. Thus, an increase in looks towards any given picture does not directly translate into a similar in magnitude reduction in looks to any other one. As such, the two measures are only partially dependent on one another. Moreover, the two measures evidence difficulties in processing differently and are valuable at different time points. Looks to the correct destination may be taken to reflect overall processing difficulty and indicate the participants' ability to reinterpret the ambiguous sentence once they have heard the disambiguating segment. Looks to the correct destination are, nevertheless, not informative of the presence or not of garden path effect; the fact that a participant may have difficulty correctly interpreting a sentence may not always be due to the fact that the sentence has been misinterpreted in real time. Looks to the incorrect destination, in particular at the point of the ambiguity – although in subsequent segments as well - would give a clearer insight into whether garden-path effects are indeed observed. The use of looks to the incorrect destination is also valuable in light of the fact that E-prime files were not informative as to the picture selected by the participant for the inaccurate trials.

As with Studies 1A and 1B, the analyses were conducted using logistic mixed effects for the accuracy data and linear mixed effects for the reaction times and the gaze data (Baayen, Davidson & Bates, 2008) implemented in the lme package in R (Bates, Maechler, Bolker & Walker, 2015, version 3.5-0). The fixed effects structure was the same as for the accuracy data and reaction. The random effects structure included random intercepts and slopes for subject, lexical item, item-ambiguity but also trial id defined as a unique pairing of subject and trial for which there were multiple observations. Separate analyses were carried out for looks to ID and CD as the measures are not fully independent from one another and were not compared between them. Each region was analysed individually.

A growth curve model was fitted to access changes over time (Mirman, Dixon & Magnuson, 2008; Mirman, 2014) using power polynomials as with Studies 1A and 1B (see Chapter 2 for more details about the nature of this analysis and its advantages). The process was the same as with the previous studies.

Sum coding (-1, 1) was used for between subject variables (monolingual vs. bilingual) and fixed main effects of ambiguity (ambiguous vs. unambiguous sentences) and referential context (one or two referents in visual stimuli). Time was scaled, and a fourth order orthogonal polynomial was fitted in the range of bins in order to conduct the growth curve analysis. The maximal model which converged was then compared to the same model with lower order polynomials. The reports from the best fitting model are always reported. The dependent variable for the eye-tracking was the empirical logit transformation of the aforementioned ratio in order to reduce skew in the data (Barr, 2008). Weights were added to each observation based on the reciprocal of the variance (i.e. 1/weights).

4.3. Results

4.3.1 Accuracy

There was a main effect of ambiguity ($\beta = 0.789$, $SE = 0.11$, $z = 7.113$, $p < 0.001$) but no effect of referential context ($\beta = -0.047$, $SE = 0.10$, $z = -0.473$, $p = 0.636$) or group ($\beta = -0.020$, $SE = 0.24$, $z = -0.083$, $p = 0.934$) and no interactions between variables. Accuracy was lower for the ambiguous condition than for the unambiguous condition. Figure 19 shows the raw accuracy data.

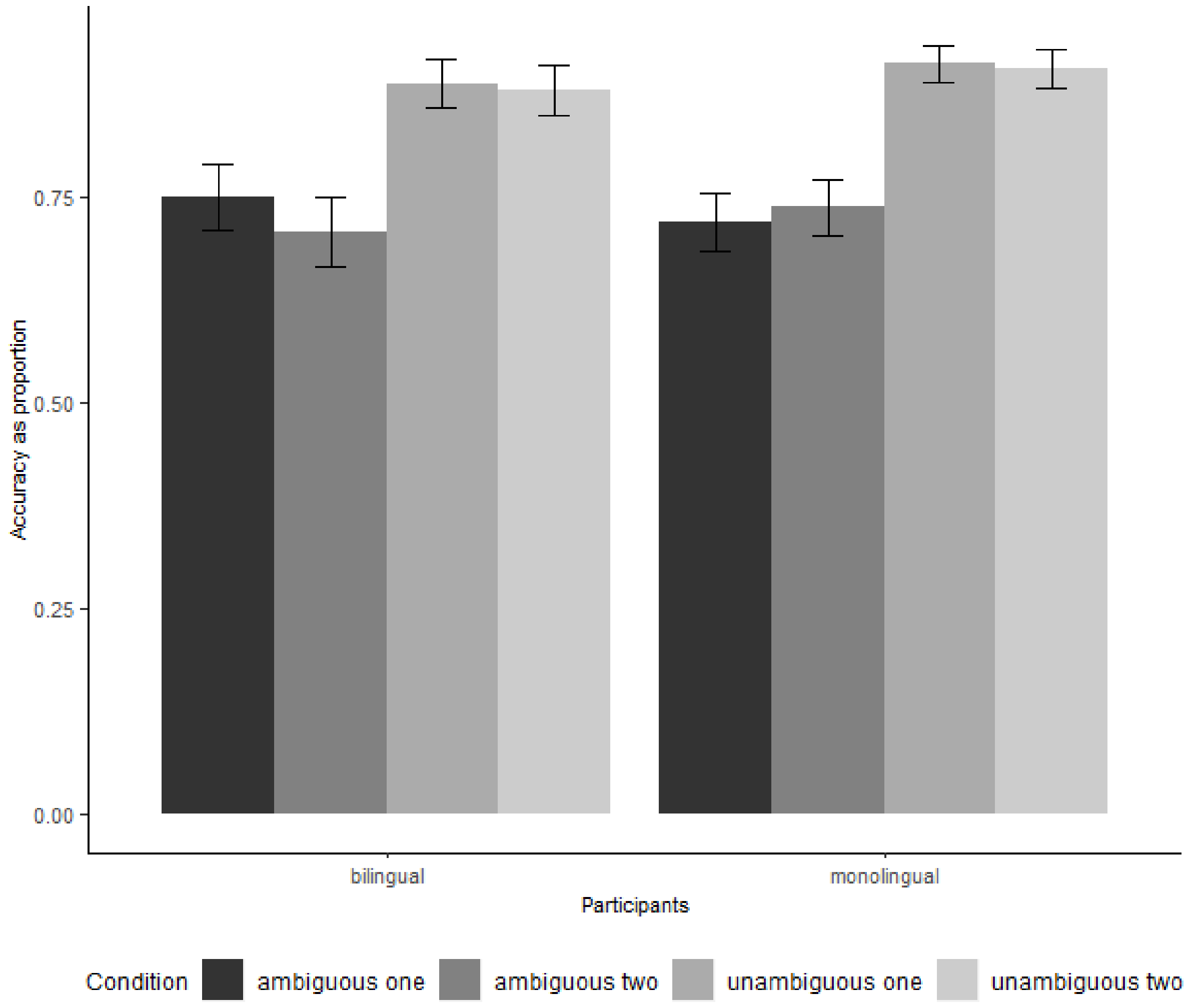

Figure 19 Accuracy as ratio (SE) by group and condition

4.3.2 Reaction Times

Figure 20 presents an overview of the reaction time data.

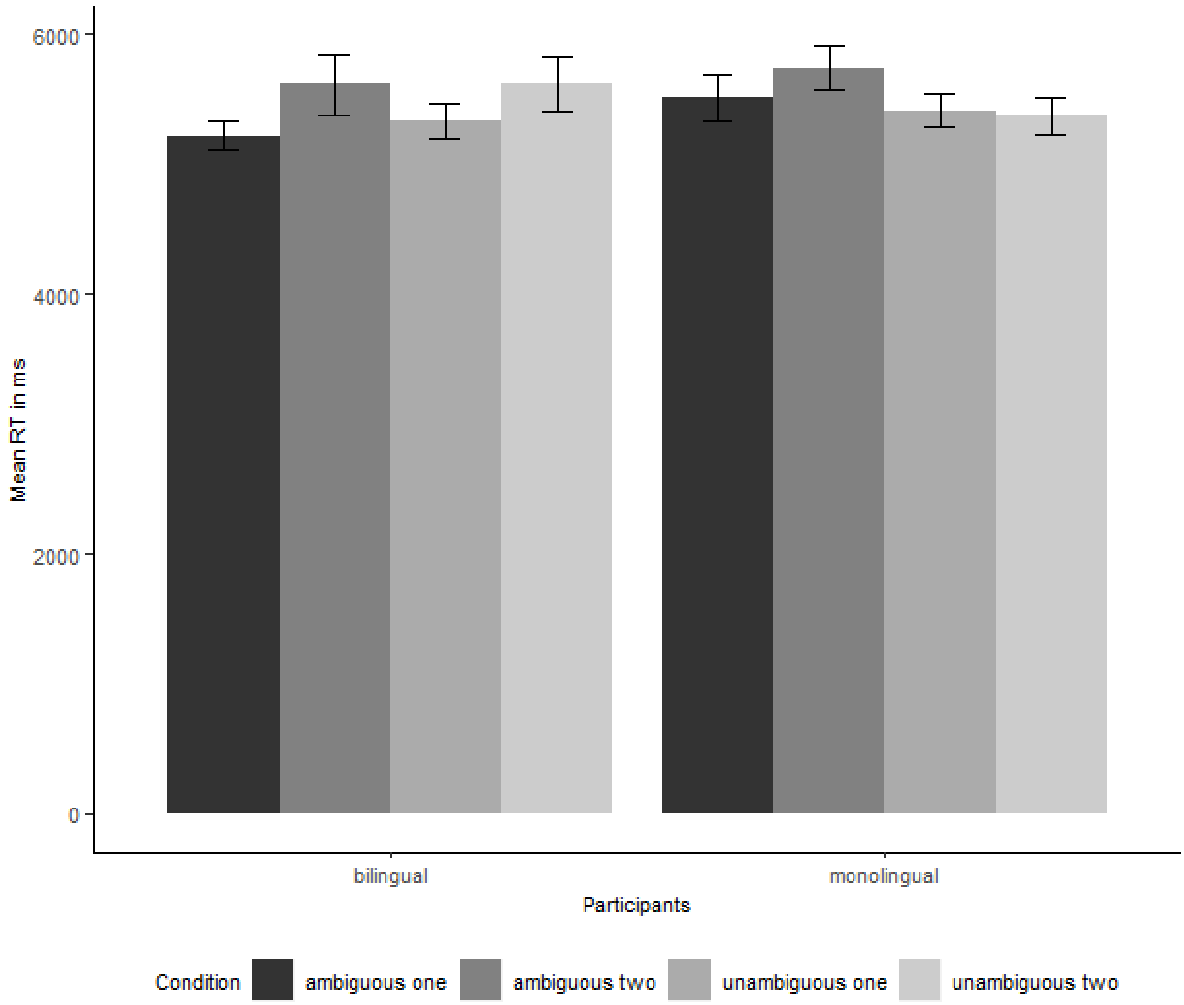

Figure 20 Reaction times in ms (SE) by group and condition

There was no effect of group (β = -5.21, SE = 116.39, t = -0.045, p = 0.964) or ambiguity (β = -63.98, SE = 58.48, t = -1.219, p = 0.223) but there was a trend for an effect of referential context (β = 107.48, SE = 58.94, t = 1.824, p = 0.074). Reaction times were faster for the one-referent context than for the two-referent by about 200ms across most conditions. There were no significant interactions.

4.3.3 Gaze data

ID region ("on the plate")

Looks to Incorrect Destination (i.e. the plate)

Figures 21 & 22 present an overview of children's looks to the incorrect destination (the plate) when they heard the first PP (on the plate) by group and condition, overall in the time window and over time within it.

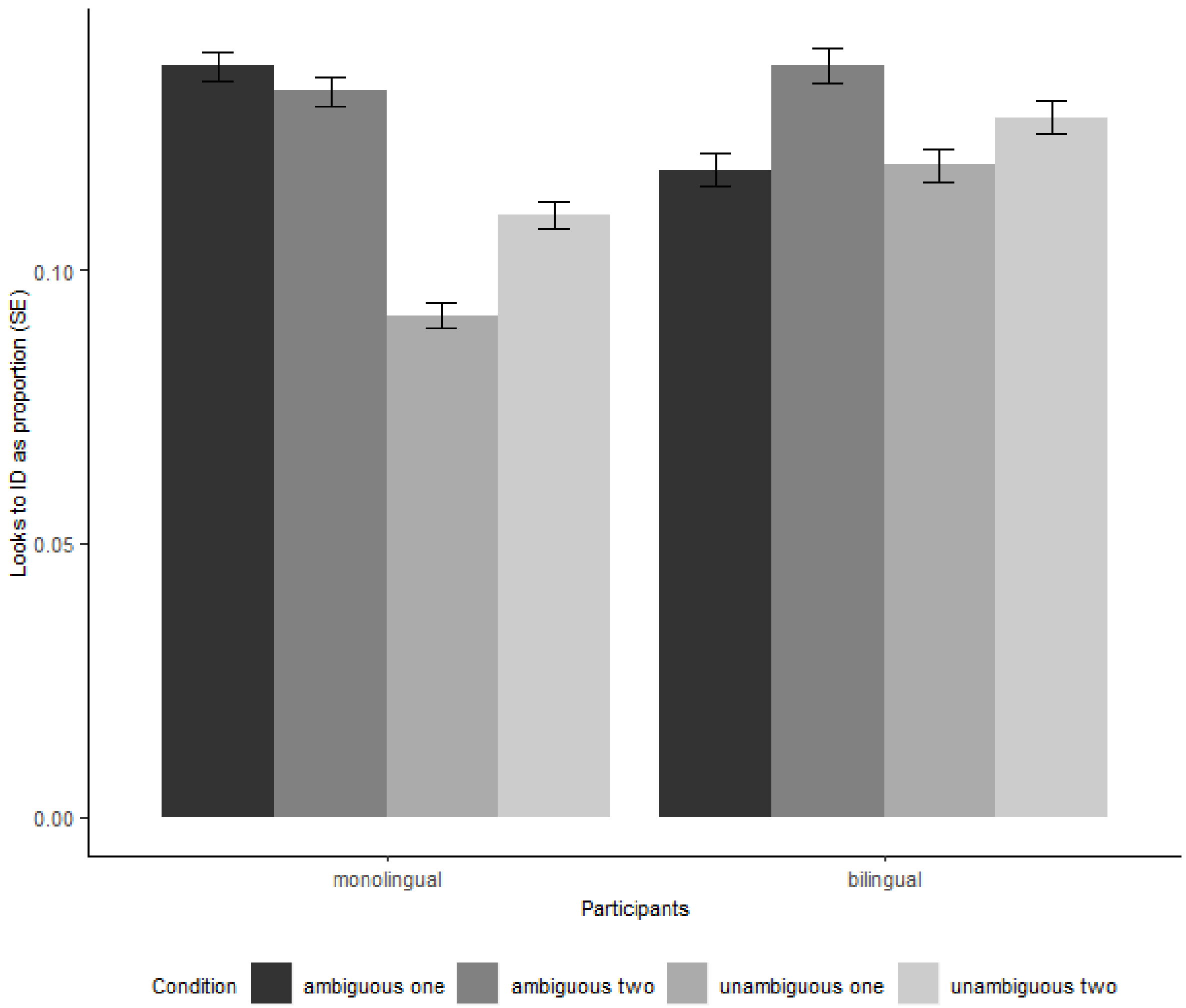

Figure 21 Looks to the incorrect destination (the plate) in the ID region (while the participants heard "on the plate") by group and condition overall

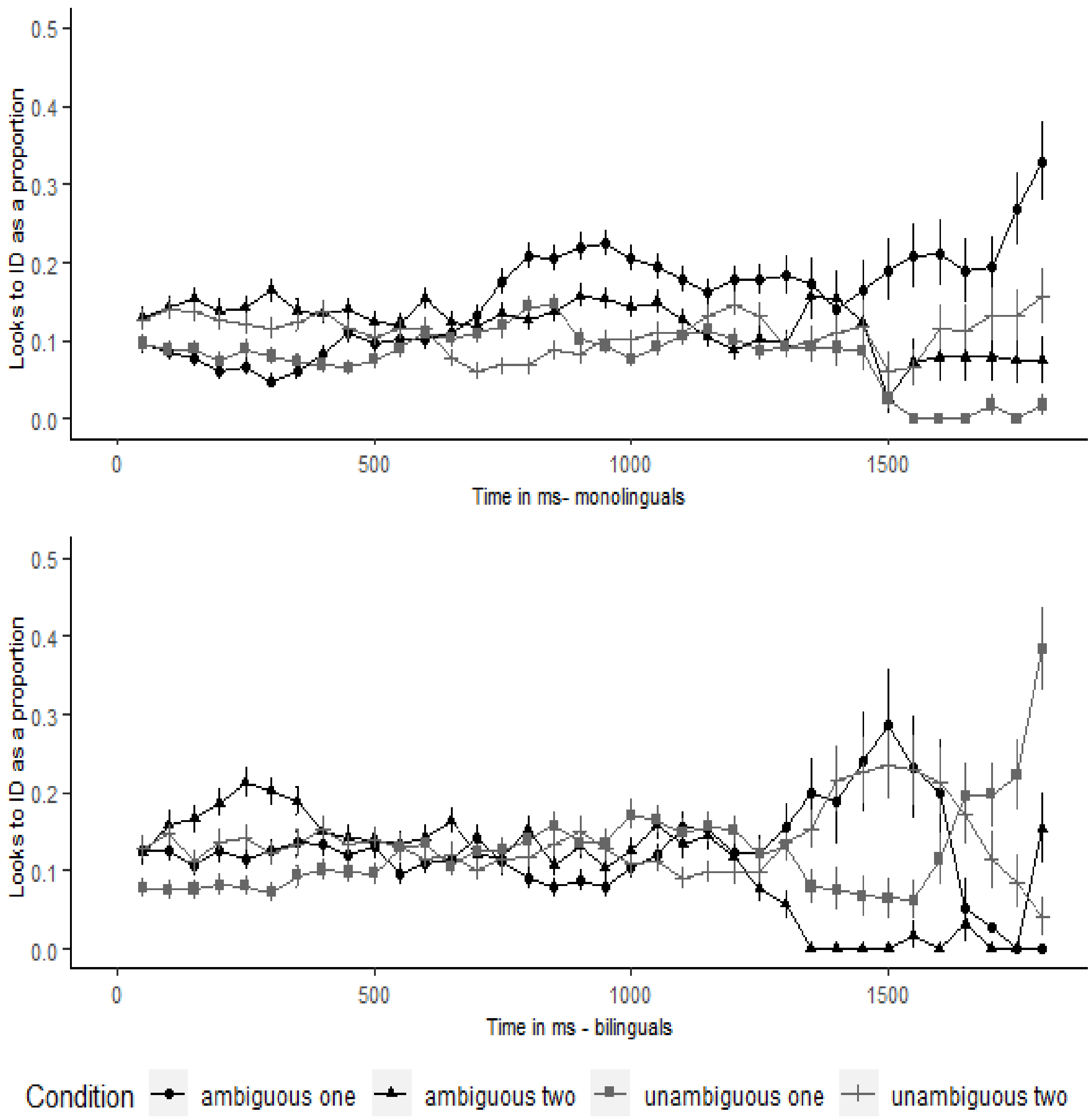

Figure 22 Looks to the incorrect destination (the plate) in the ID region (while the participants heard "on the plate") by group and condition over time

There was no main effect of group (β = 0.010, SE = 0.02, t = 0.434, p = 0.666), ambiguity (β = -0.019, SE = 0.02, t = -1.040, p = 0.299) or referential context (β = 0.007, SE = 0.02, t = 0.402, p = 0.688) and no interactions on the intercept term for looks towards the image depicting the incorrect destination in the ID region. Although numerically both groups showed fewer looks towards the incorrect destination in the unambiguous condition, as expected, only the monolingual children looked more towards the incorrect destination in the ambiguous one referent condition relative to the 2-referent condition. In both populations, however, looks to the incorrect destination were few (Figure 22). The polynomial terms were also not significant (β =-0.028, SE = 0.52, t = -0.053, p = 0.958 for the linear term; β = -0.727, SE = 0.5, t = -1.406, p = 0.160 for quadratic term).

On the linear term, there was a main effect of group (β = -1.140, SE = 0.52, t = -2.176, p = 0.030) and an interaction of group by ambiguity (β = 1.364, SE = 0.52, t = 2.603, p = 0.009).

The group by ambiguity interaction is driven by an increase in looks towards the incorrect destination in the monolingual children; the latter is most pronounced in the 1-referent condition. It is for this condition that looks to the incorrect destination are most steadily increasing about 600ms post onset of the first PP. In fact, about 1,000ms into the time window looks towards the incorrect destination are numerically about double for the ambiguous 1-referent condition than for the other three even though there was no significant effect of context or interaction with context overall. For the bilinguals, the pattern is much less clear.

The pattern observed in the monolinguals can be interpreted as evidence for garden path effects; the children are progressively committing to an interpretation consistent with the prepositional phrase modifying the verb prior to disambiguation and moving away from this interpretation when there is no ambiguity. Numerically, this is most strongly observed in the 1-referent condition where there is referential context to aid processing but the effect of context or the interaction with ambiguity did not reach significance. It occurs for a less protracted period of time for the 2-referent condition where the presence of the second target object (apple on the plate) guides the parser towards an NP-modifying interpretation of the ambiguous PP.

In contrast, the bilinguals showed a more protracted pattern of fluctuation and variability than that seen in the monolinguals which lasts until about 1,200ms post onset. Following this, looks to the incorrect destination increase for the ambiguous 1-referent condition but also for the unambiguous 2-referent one as well. Moreover, looks to the incorrect destination drop to nearly none for the ambiguous 2-referent condition and the unambiguous 1-referent condition. The increase in looks towards the incorrect destination for the unambiguous 2-referent conditions is theoretically unmotivated and readily interpretable.

Looks to Correct Destination (i.e. the bag)

Figures 23 & 24 present an overview of children’s looks to the correct destination (the bag) when they heard the first PP (on the plate) by group and condition, overall in the time window and over time within in.

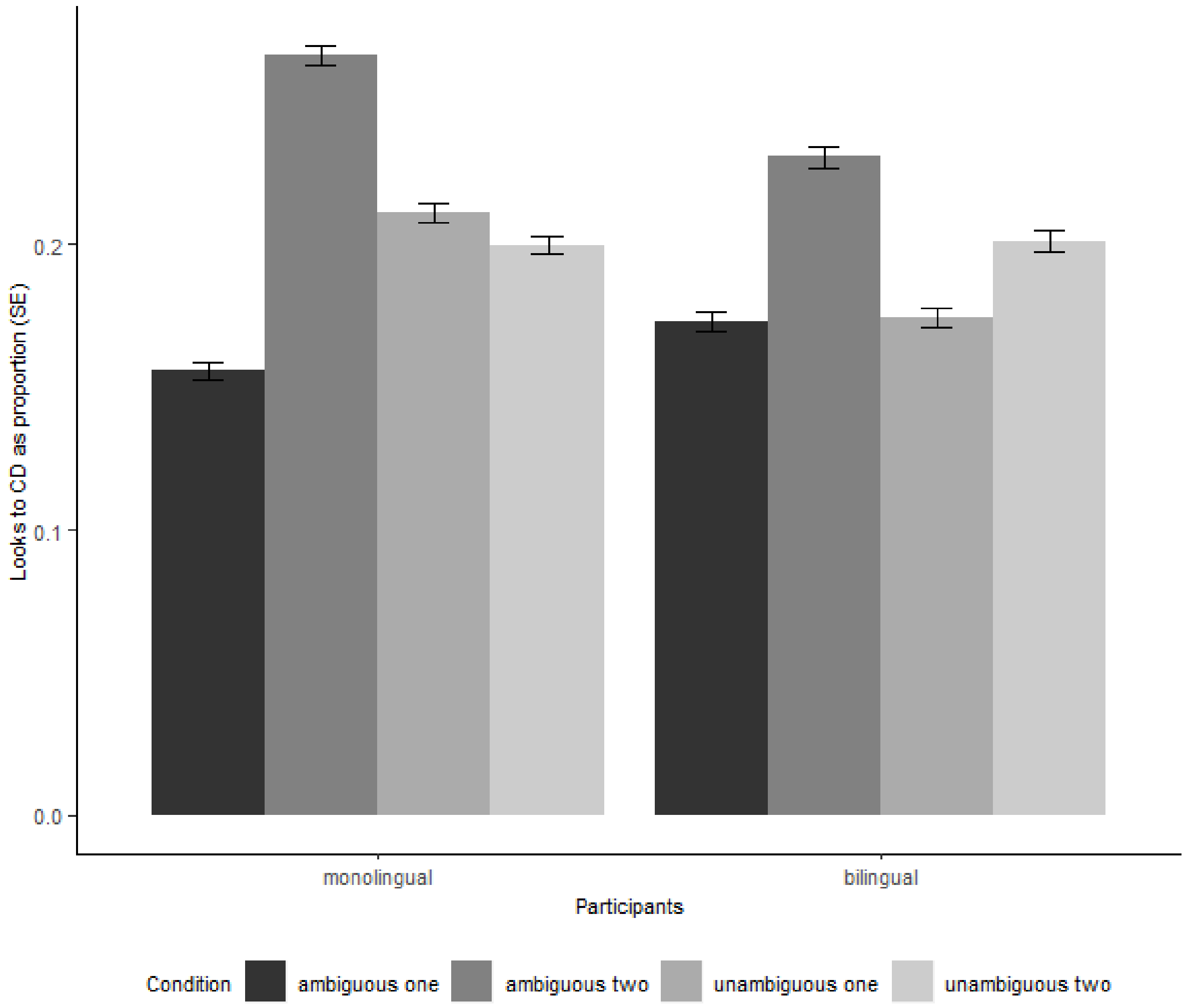

Figure 23 Looks to the correct destination (the bag) in the ID region (while the participants heard “on the plate”) by group and condition overall

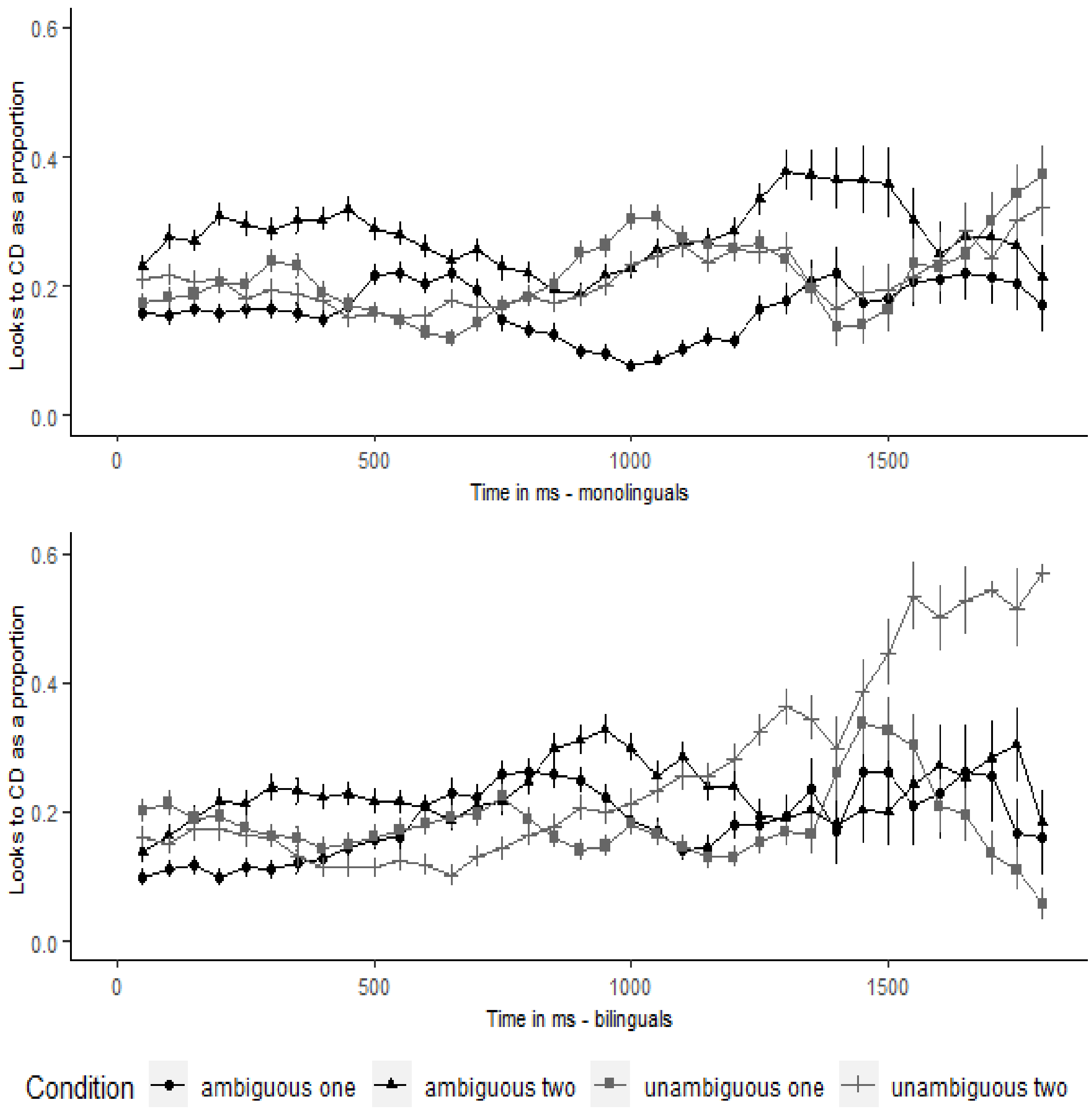

Figure 24 Looks to the correct destination (the bag) in the ID region (while the participants heard "on the plate") by group and condition over time

There were no main effects or interactions on the intercept term. There was a trend for the linear term towards significance (β = 1.127, SE = 0.67, t = 1.683, p = 0.093). There were no interactions on the linear term. It is not expected that there will be any effects on the looks towards the correct destination, in the given example, the plate, as at this point the participants have not heard the respective disambiguating PP yet. However, for both groups looks to the correct destination were more for the 2-referent ambiguous condition than for the respective 1-referent condition. The only main effect on the linear term is the main effect of ambiguity (β = 4.667, SE = 0.67, t = 6.963, p < 0.001). The effect of ambiguity on the linear term is due to the progressive increase in looks to the correct destination in the unambiguous condition after about 500ms (see Figure 25) which does not occur for the ambiguous sentences.

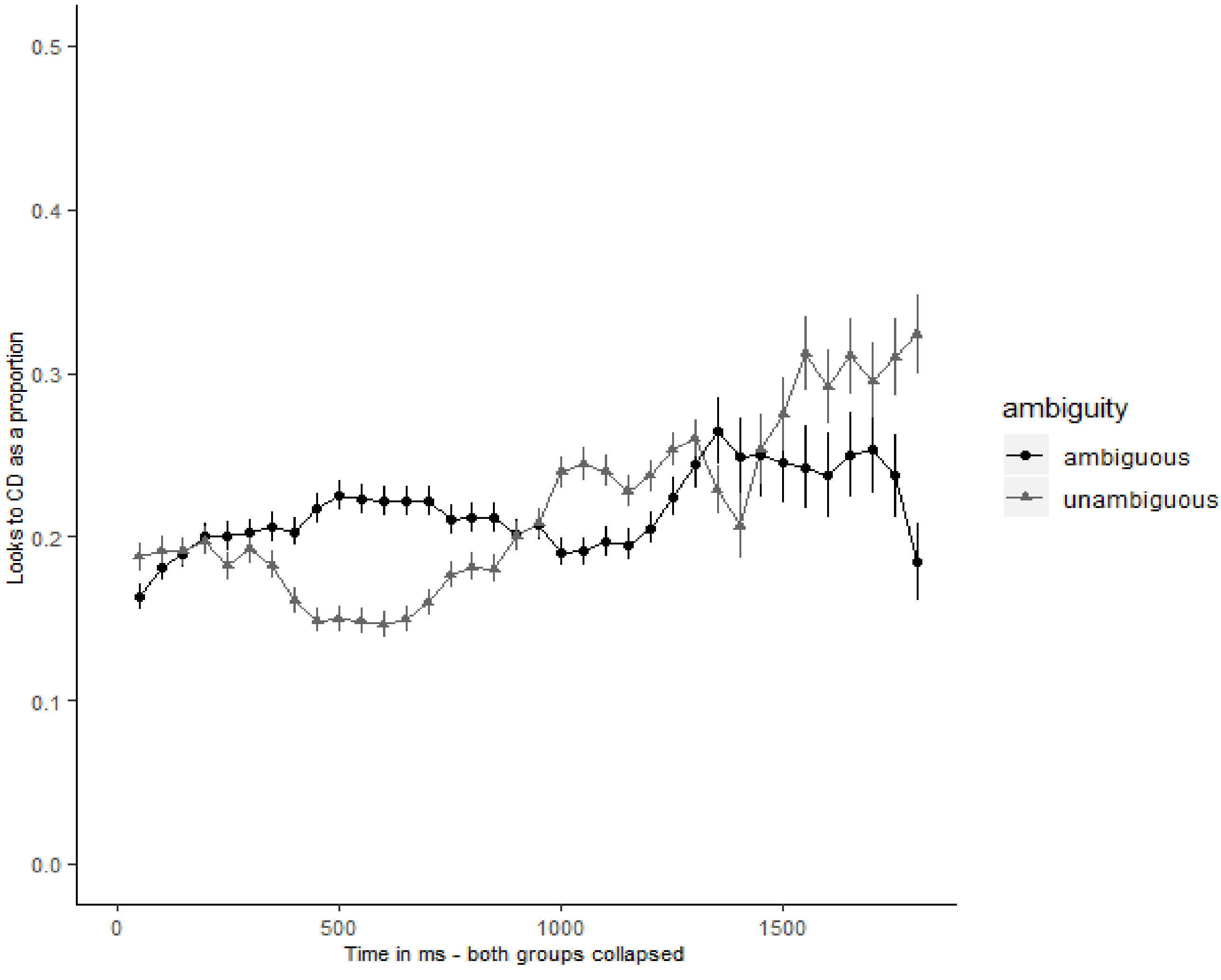

Figure 25 Looks to correct destination (the plate) in the ID region ("on the plate") by ambiguity - both groups collapsed

CD region ("in the bag")

Looks to Incorrect Destination

Figures 26 & 27 present an overview of children's looks to the incorrect destination (the plate) when they heard the second PP (on the plate) by group and condition, overall in the time window and over time within this window.

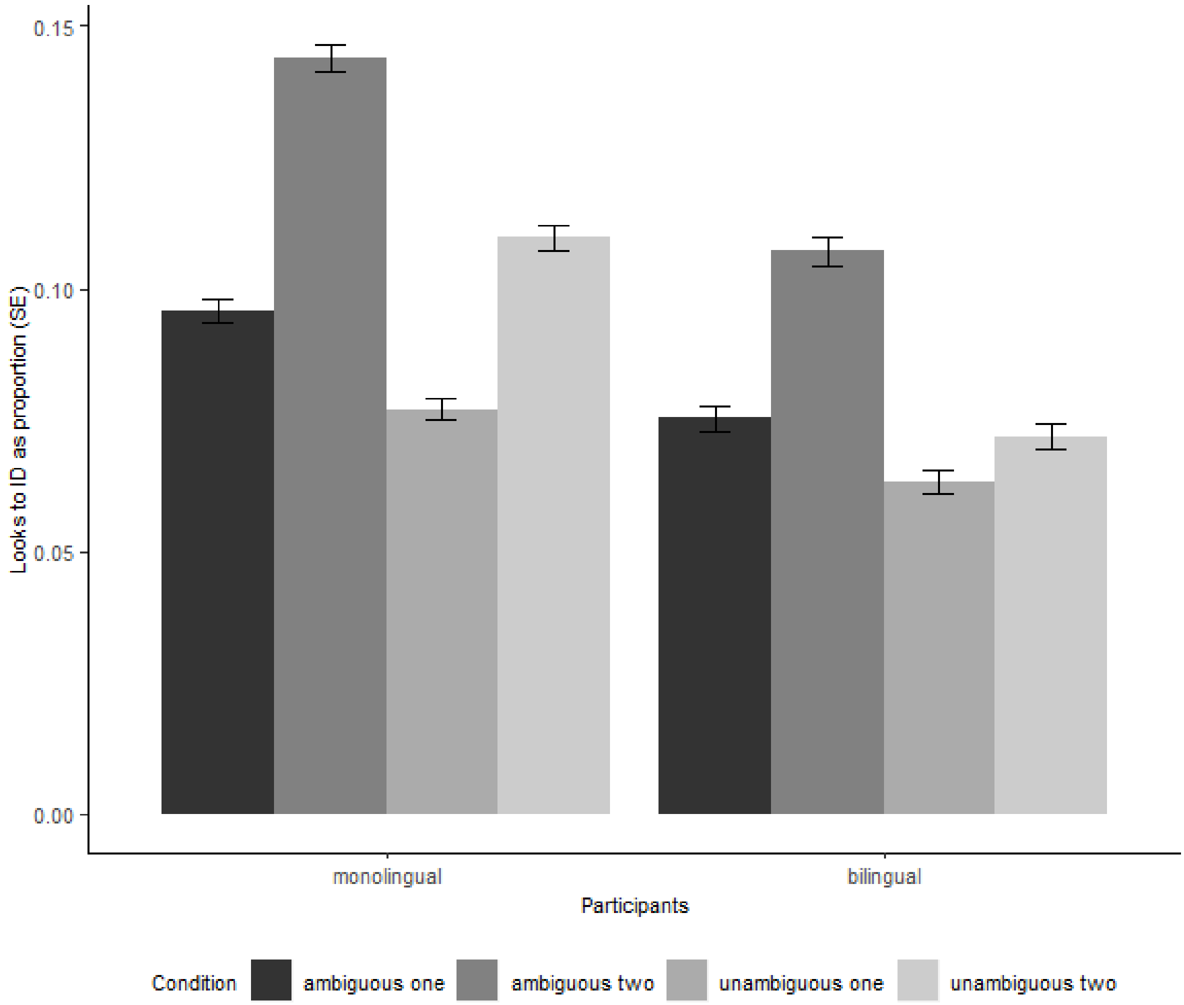

Figure 26 Looks to the incorrect destination (the plate) in the CD region (while the participants heard "in the bag") by group and condition overall

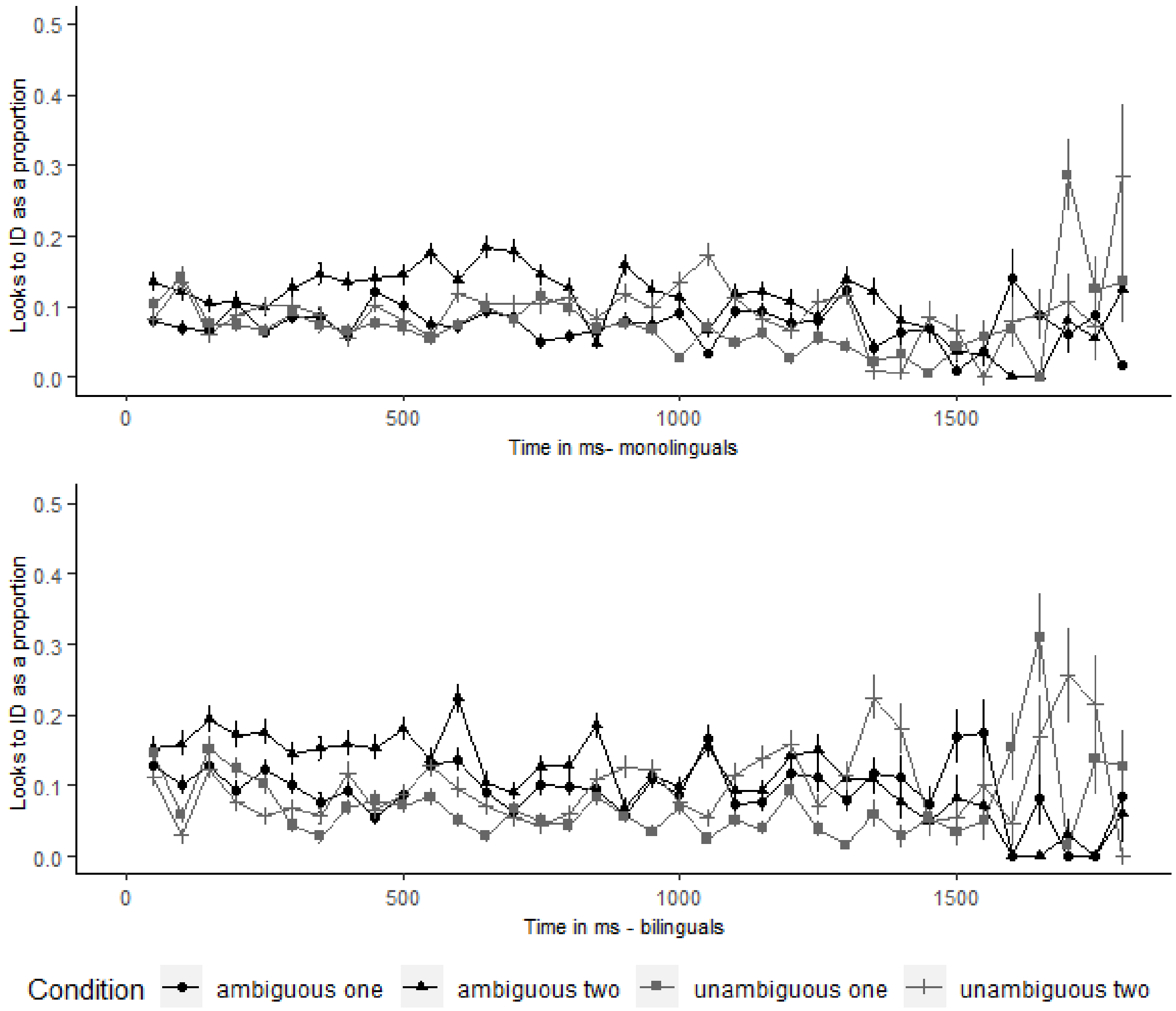

Figure 27 Looks to the incorrect destination (the plate) in the CD region (while the participants heard "in the bag") by group and condition over time

There were no main effects or interactions on the intercept term. The quadratic term was significant (β = 0.982, SE = 0.45, t = 2.189, p = 0.029) but the linear and cubic term were not (β = -0.419, SE = 0.45, t = -0.923, p = 0.356 and β = 0.683, SE = 0.44, t = 1.551, p = 0.121 respectively) nor where there any main effects or interactions on the linear term. Numerically looks to the incorrect destination were more for the ambiguous conditions than the unambiguous conditions aggregated over the time window in the monolinguals, but the bilinguals showed more looks to the incorrect destination for the ambiguous conditions in the first 500ms of the time window in question. The motivation for examining looks towards the incorrect destination is to determine if any garden-path effects have lingered or belatedly manifested in the second time window analysed. There was no evidence of garden-pathing at this segment. However, absence of evidence should not be equated with evidence of absence.

Looks to Correct Destination

Figures 28 & 29 present an overview of children's looks to the correct destination (the bag) when they heard the second PP (on the plate) by group and condition, overall in the time window and over time within in.

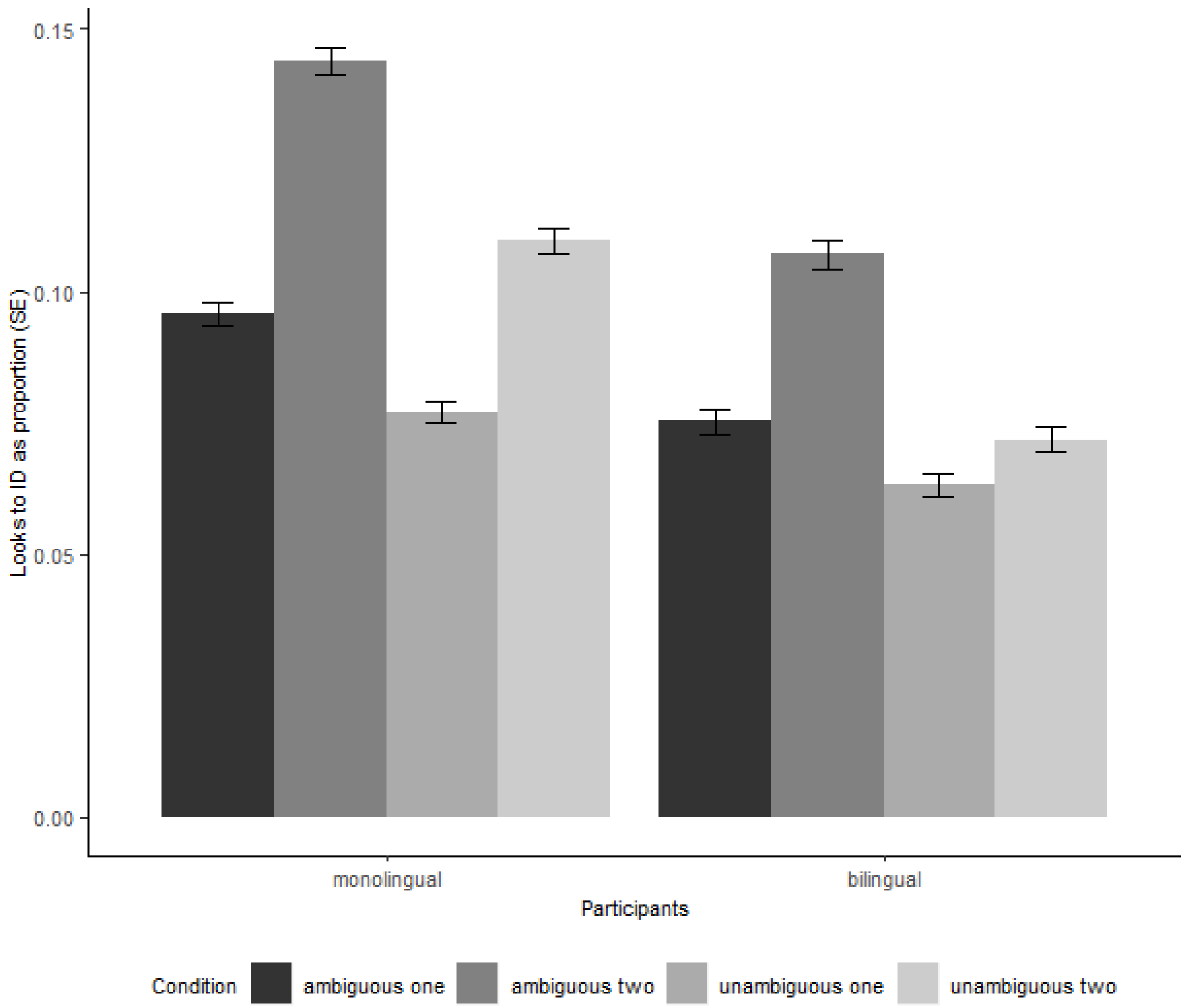

Figure 28 Looks to the correct destination (the bag) in the CD region (while the participants heard "in the bag") by group and condition overall

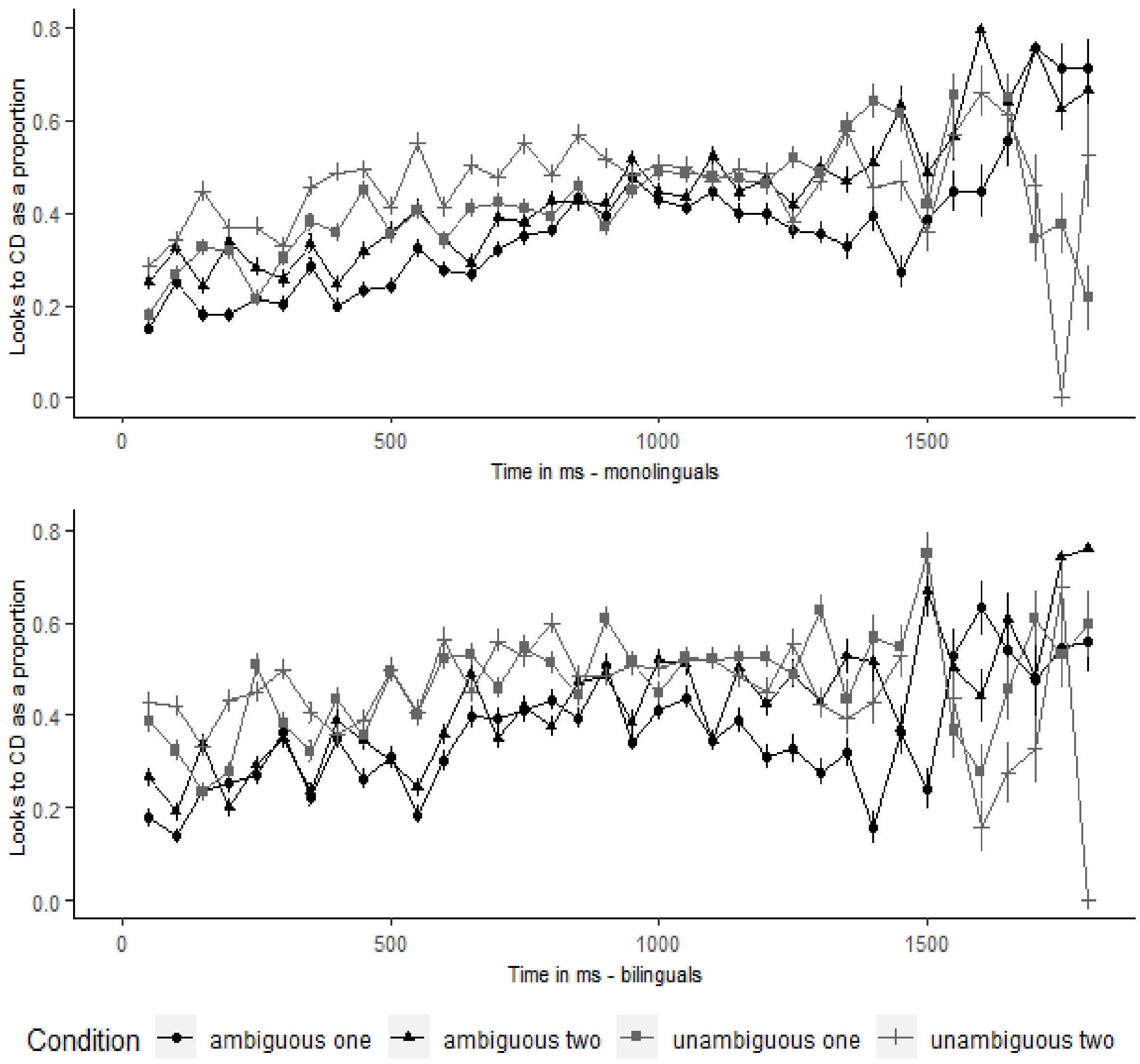

Figure 29 Looks to the correct destination (the bag) in the CD region (while the participants heard "in the bag") by group and condition over time

There was an effect of ambiguity (β = 0.085, SE = 0.03, t = 3.054, p = 0.003) on looks to the correct destination in the CD region but no main effect of group (β = 0.022, SE = 0.04, t = 0.608, p = 0.545) or referential context (β = 0.044, SE = 0.03, t = 1.393, p = 0.184) on the intercept term. There were no interactions. Looks towards the correct destination were consistently fewer for ambiguous sentences than unambiguous ones for both groups and both referential contexts. This suggests that the ambiguous sentences were harder to process.

The polynomial terms were significant predictors of the children's looks to the correct destination ($\beta = 3.630$, SE = 0.80, $t = 4.542$, $p < 0.001$ for the linear term and $\beta = -5.133$, SE = 0.79, $t = -6.491$, $p < 0.001$). There were no significant main effects on the linear or quadratic term; there was only a trend for a main effect of ambiguity on the linear term ($\beta = -1.425$, SE = 0.798, $t = -1.785$, $p = 0.074$). There was only one significant interaction on the linear term (and none on the quadratic term); the ambiguity by context interaction ($\beta = -2.211$, SE = 0.80, $t = -2.769$, $p = 0.006$). This was driven by the overall benefit that the presence of two referents in the ambiguous condition conferred (Figure 30).

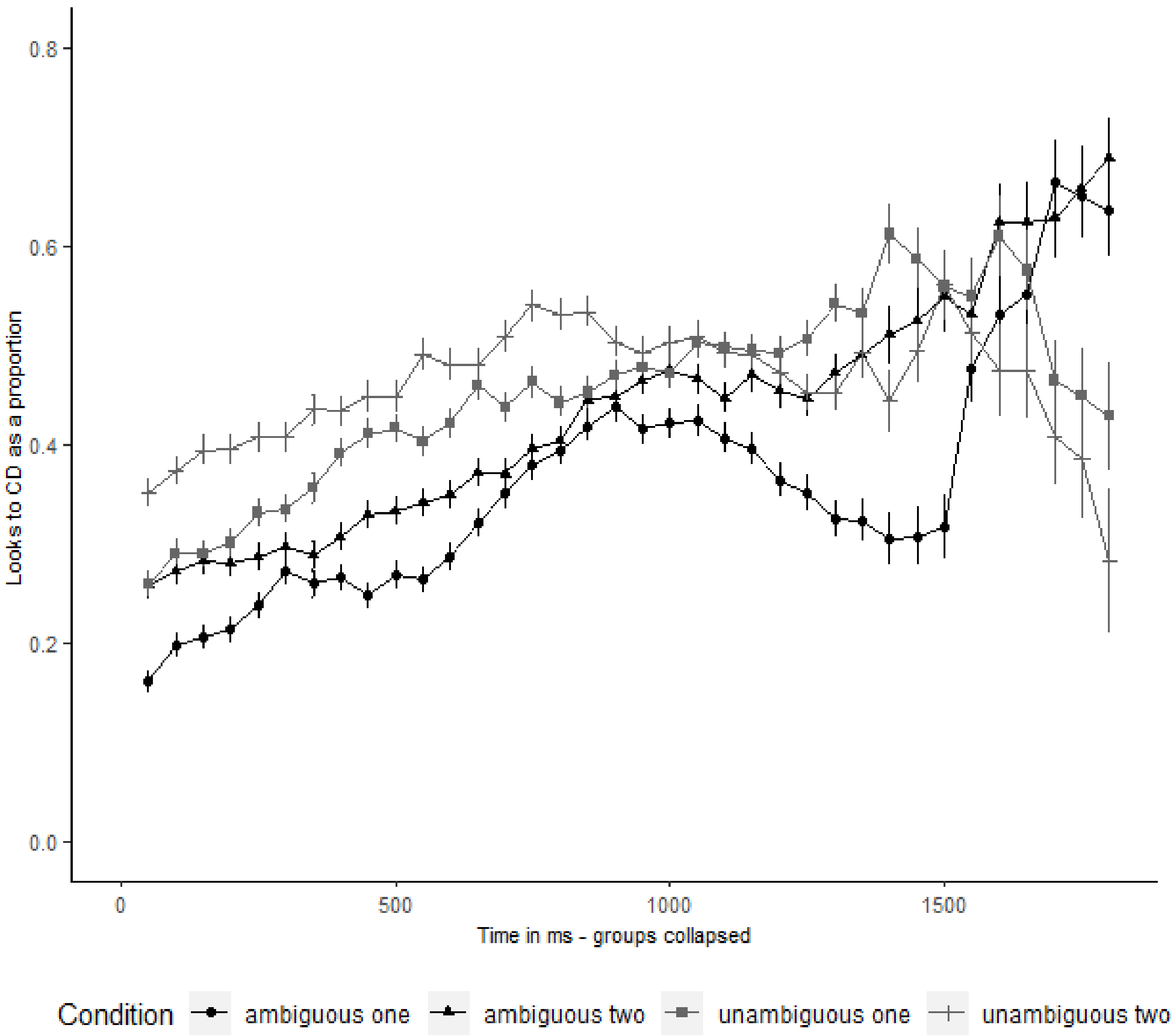

Figure 30 Looks to correct destination (the bag) in the CD region ("in the bag") by condition over time - all groups collapsed

More specifically, looks towards the correct picture increased in both ambiguous conditions over time, even though they remained overall consistently fewer than for the unambiguous condition, this increase was sustained only for the two-referent condition. For the one-referent ambiguous condition, there was a drop in looks to the correct destination after about 1,000 ms and a rapid increase after 1,500 ms. A plausible interpretation for the latter is that this decline is due to participants looking towards other pictures to confirm the interpretation formed. For the unambiguous sentences, looks towards the correct destination are consistently higher than for the ambiguous sentences but increase. The presence of two referents leads to more looks towards the correct destination until about 1,000ms after which the looks seem to peak and fluctuate. This suggests that the children are using the context.

FINAL region ("before going to school")

Looks to Incorrect Destination

Figures 32 & 33 present an overview of children's looks to the incorrect destination (the plate) when they heard the second PP (on the plate) by group and condition, overall in the time window and over time within in.

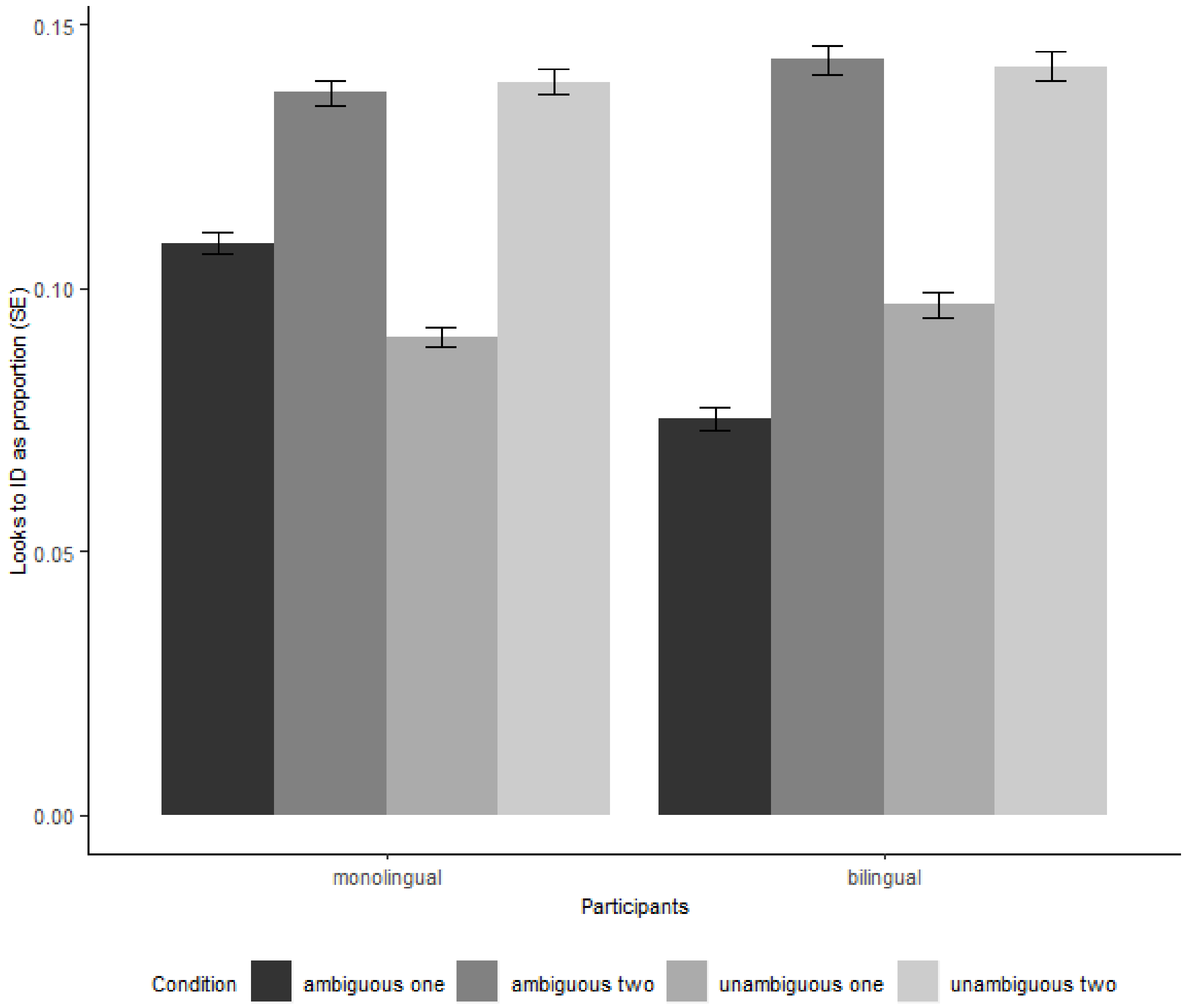

Figure 31 Looks to incorrect destination (the plate) in the final region ("before going to school) by group and condition overall

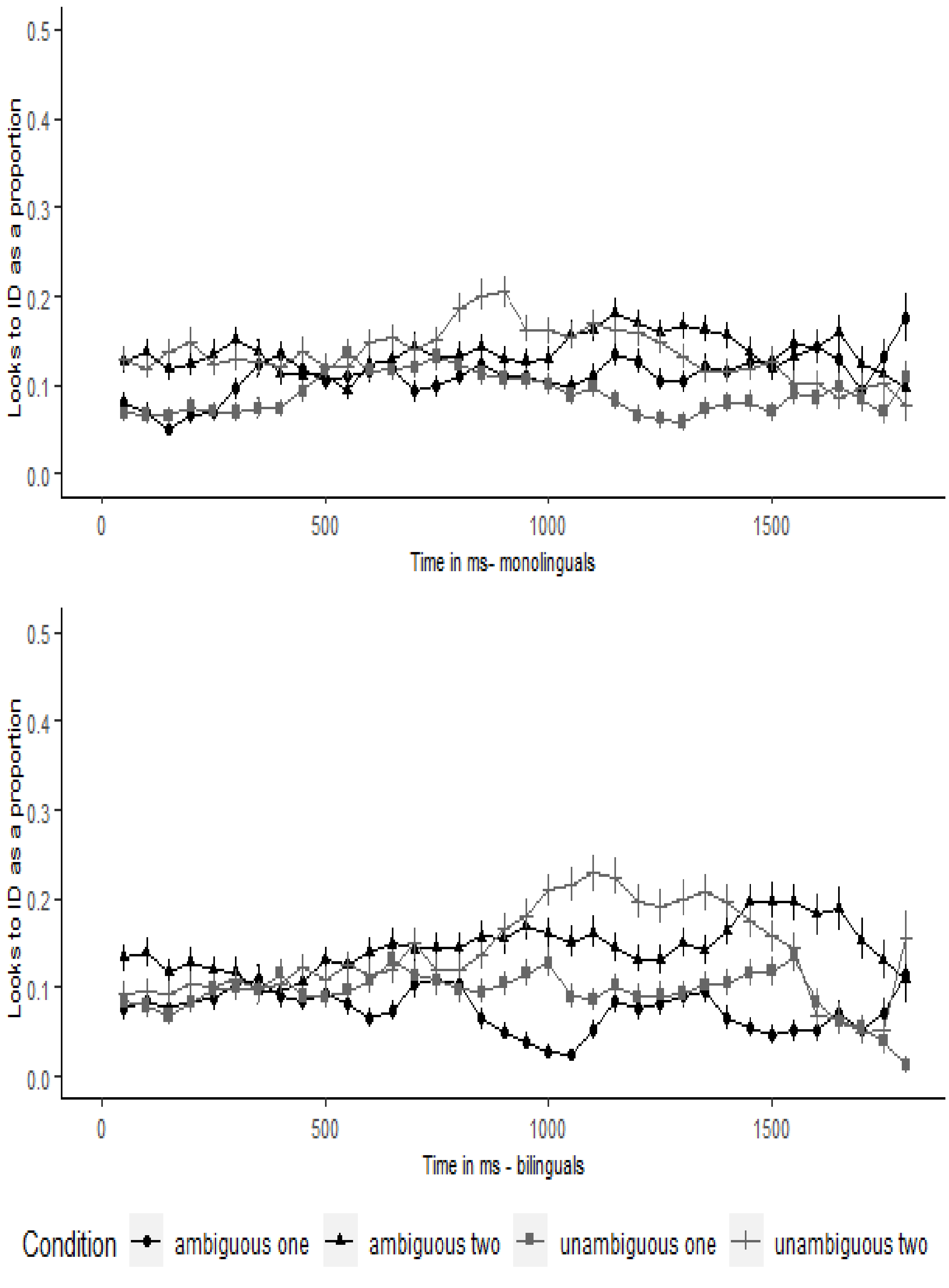

Figure 32 Looks to incorrect destination (the plate) in the final region ("before going to school") over time by group and condition

There was a main effect of referential context on the intercept term for the looks to the incorrect destination in the final region of the sentence ($\beta = 0.053$, SE = 0.02, t = 2.811, p = 0.013). There were more looks to the incorrect destination in the 2-referent condition than the 1-referent condition' for both groups (Figure 15). There was no effect of group or ambiguity or any interaction on the intercept term.

The linear and quadratic terms were also not significant. There were no significant main effects on the linear term but there was a trend for an effect of context on the linear term (β = 0.986, SE = 0.51, t = 1.937, p = 0.053). Two interactions reached statistical significance on the linear term; the interaction of group with ambiguity (β = 1.330, SE = 0.51, t = 2.614, p = 0.010) and the interaction of group with context (β = 1.228, SE = 0.51, t = 2.414, p = 0.016). The interactions on the linear term reflect differences in change over time for the two groups. These two interactions are driven by the differential impact of ambiguity and referential content on how the gaze data changes over time for each group. For the monolinguals, looks to the incorrect destination are more for the 1-referent unambiguous condition than in the other three throughout the window. For the bilinguals, there are more looks towards the incorrect destination in the 2-referent conditions and the unambiguous about 800ms the time into the time window but not before that. The patterns are difficult to interpret as they are theoretically unmotivated.

Looks to Correct Destination

Figures 33 & 34 present an overview of children's looks to the correct destination (the bag) when they heard the second PP (on the plate) by group and condition, overall in the time window and over time within it.

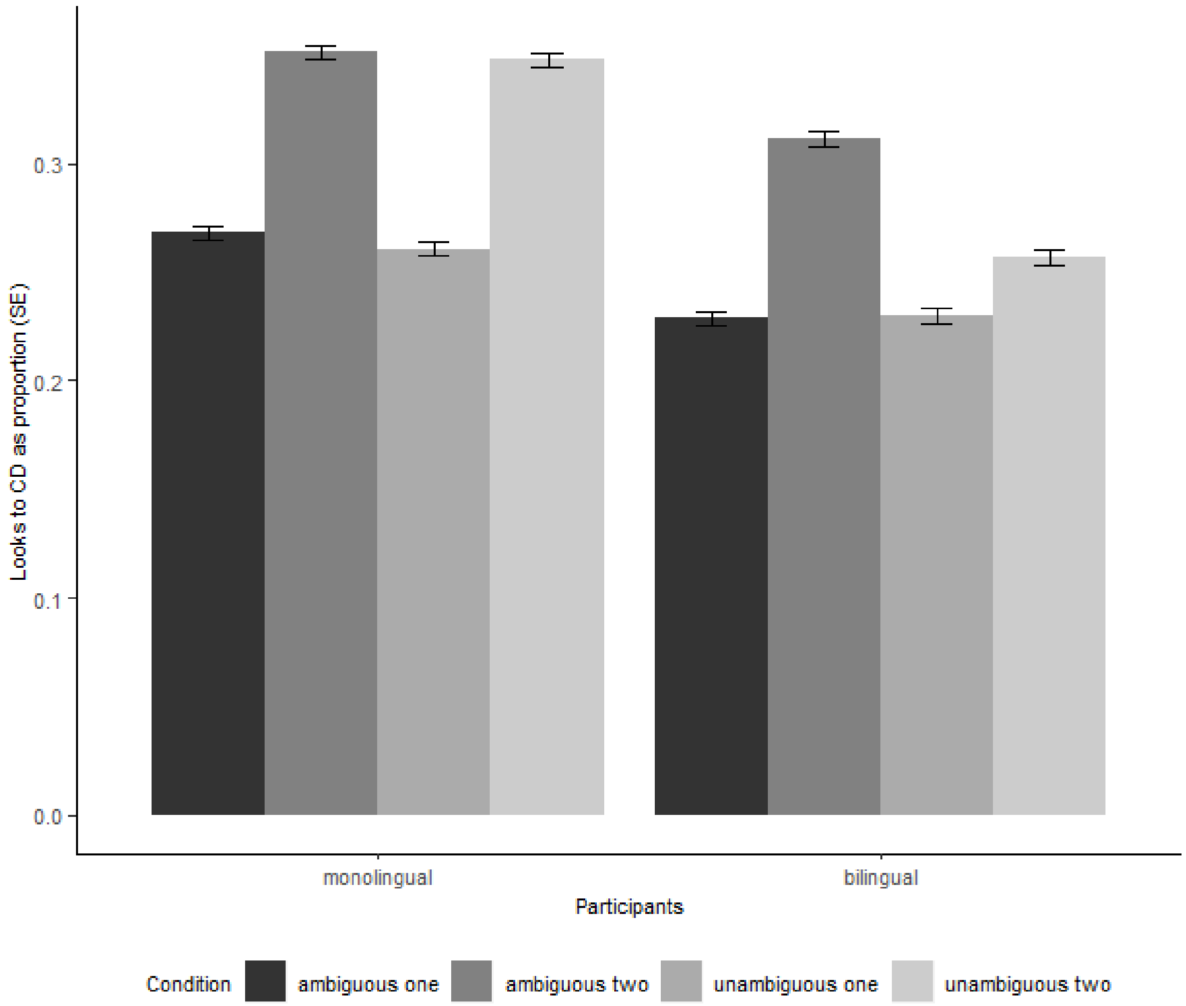

Figure 33 Looks to correct destination (the bag) in the final region ("before going to school) by group and condition overall

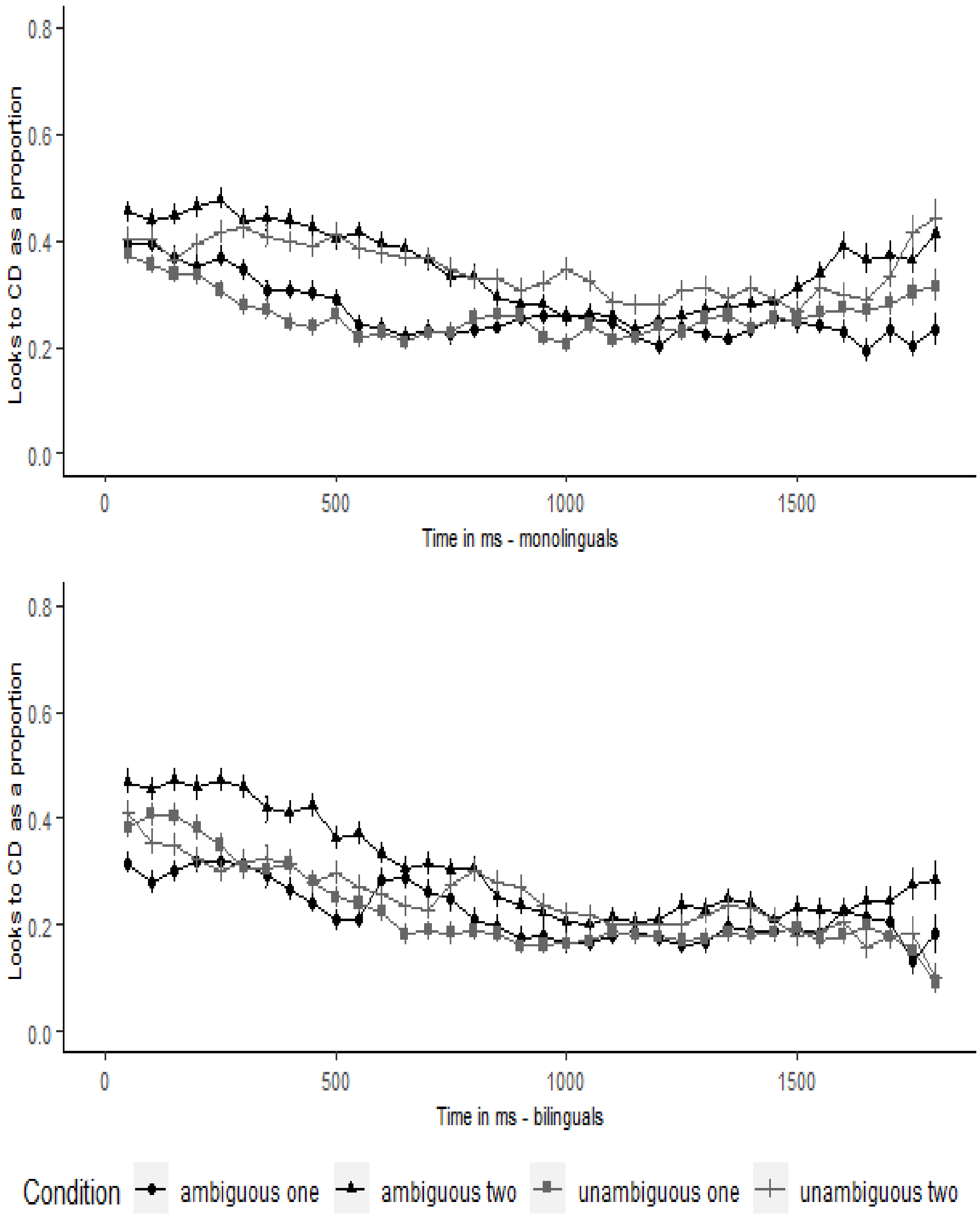

Figure 34 Looks to correct destination (the bag) in the final region ("before going to school") over time by group and condition

There was a main effect of referential context on the looks to the target in the final region ($\beta = 0.071$, $SE = 0.03$, $t = 2.518$, $p = 0.025$) but not of group or ambiguity. There were more looks towards the correct destination in the 2-referent condition than in the 1-referent condition. There were no interactions on the intercept term despite the numerically smaller benefit of referential content in the ambiguous condition for the bilinguals relative to the monolinguals. There were more looks to the correct destination for the two-referent condition (Figure 34).

The linear and quadratic terms were both significant ($\beta = -4.842$, $SE = 0.67$, $t = -7.209$, $p < 0.001$ and $\beta = 3.892$, $SE = 0.67$, $t = 5.841$, $p < 0.001$ for linear and quadratic term respectively). There was an effect of both ambiguity ($\beta = 1.934$, $SE = 0.67$, $t = 2.881$, $p = 0.004$) and context ($\beta = -1.651$, $SE = 0.67$, $t = -2.458$, $p = 0.014$) on the linear term but not of group. Looks to the correct destination also declined with time but did so more steeply for the ambiguous sentences than for the unambiguous ones.

On the surface, this is counterintuitive; it may simply reflect that the ambiguous sentences were harder to process, and that participants spent more time looking at the correct destination for confirmation. Therefore, at the start of the time window, there were more looks towards the correct destination in the ambiguous condition, but these progressively converged for both groups. As they were certain earlier on for the unambiguous sentences, looks were reoriented to the alternative pictures. In terms of the effect of context on the linear term, looks to the correct destination remained consistently higher for the two-referent conditions until about 1,000 ms into the window. This benefit is predicted in the literature but only for the ambiguous condition.

The only significant interaction on the linear term was the group by context by ambiguity interaction ($\beta = 1.654$, $SE = 0.67$, $t = 2.463$, $p = 0.014$). The three-way interaction reflects differential changes for each condition over time for the two groups (Figure 34). It should be noted that looks to the correct picture are higher for the 1-referent ambiguous condition relative to the other three conditions for the bilinguals and these looks decline more slowly. There was however no interaction of group with any of the other variables on the linear term.

For the monolinguals, looks to the correct destination declined more slowly for the 2-referent conditions; this effect was stronger in the ambiguous sentences but only for the first 500ms (see Figure 34). For the bilinguals, there was a slower decline in looks to the correct destination in the 2-referent condition only for the ambiguous sentences. One way of tentatively interpreting this finding is that referential context to some extent facilitates processing but in a rather delayed manner in real time processing. A visual inspection of the respective looks over the time window in question (Figure 35) provides weak support for this.

4.4. Discussion

This study constitutes a follow up to Trueswell et al. (1999) with bilingual children who are older than the children in the original study. The main finding was that both bilingual and monolingual children had greater difficulty with ambiguous sentences than with unambiguous ones even after they had heard the complete sentence (as evidenced by the difference in accuracy scores), suggesting difficulty in reanalysing. The two groups did not differ in terms of accuracy or reaction times, but the monolinguals showed effects of garden-pathing somewhat more clearly in the gaze data. Neither group appears to consistently utilise number of referents in the visual stimuli to aid disambiguation. This section addresses each research question individually and contrasts the results from this study with those from other seminal studies based on each research question separately.

RQ1: Do bilingual children experience garden-path effects?

The fact that both groups of children experience garden-path effects is manifested on a range of metrics obtained. The effect of ambiguity on comprehension accuracy for both monolingual and bilingual children suggests that the bilingual children experienced greater difficulty with ambiguous sentences than unambiguous ones as did the monolinguals. The most plausible explanation for this is that they have experienced garden-pathing during processing. Gaze data from each individual segment and how it changes over time also provides evidence of garden-pathing. Two measures can be interpreted to support the occurrence of garden-path effects; looks to the incorrect destination (i.e. the plate) when the participants hear the ambiguous segment and looks to the correct destination (i.e. the bag) when they hear the disambiguating phrase ("in the bag").

In the ambiguous region, differences between the two groups are observed. This is evidenced by the interaction of group by ambiguity on the linear term. More specifically, only the monolinguals showed an increase in looks towards the incorrect destination in the visual stimuli (i.e. the plate) for the ambiguous sentences, which was numerically stronger in the 1-referent condition, suggesting garden-pathing in real time. This was not the case for the bilingual children. It may be plausible that these effects may be observed in the subsequent region, where participants heard the disambiguating material (i.e. "in the bag"). However, there were no effects of ambiguity for the looks towards the incorrect destination (the plate) at this later point for either group.

Looks to the correct destination also support the notion of garden-path effects and the two groups appear to converge in terms of eye-movement. There were overall significantly fewer looks towards the correct destination in the ambiguous condition than in the unambiguous condition in the CD region, i.e. once the participants heard the second segment (the disambiguating PP "in the bag"). There was no effect of group or interaction with ambiguity on either the intercept or linear term. This suggests that both groups experienced the same degree of difficulty with garden path sentences at this segment. Visual inspection of the data shows consistently fewer looks to the correct destination in the ambiguous condition but that they increase for both ambiguous and unambiguous sentences. The above observations are to be expected; although participants have heard the disambiguating cue, the initial interpretation has not yet been abandoned and as such, participants will still spend some time still looking at the incorrect destination picture. Therefore, it makes sense that looks towards the correct destination (the bag) will increase over time. The fact that there is not a group by ambiguity interaction on the intercept term or the linear and quadratic terms indicates that this overall effect applies to both groups in terms of quantity of looks aggregated over the entire segments but also that looks towards the correct destination increase similarly over time for both groups. In other words, once given the appropriate cue, both groups realise in the same time window that the NP-modifying interpretation of the PP is the correct one. This is the earliest segment in the sentence where disambiguation is possible and the fact that there is an increase in looks to the correct picture for both groups in the same way (evidenced by the absence of a group interaction with other variables) suggests that the bilingual children are able to integrate incoming information rapidly. In sum, given the presence of ambiguity effects at this point, this does not support the notion that garden path effects are absent in bilingual children but rather that they were simply not manifested in the previous region.

It should be noted that the absence of effects in the ID region for the monolingual children is partially inconsistent with previous studies (e.g. Snedeker & Trueswell, 2004; Trueswell et al. 1999) which have shown an effect of ambiguity for looks to the incorrect destination in the ID region (more aggregate looks to the plate for ambiguous sentences). This is not replicated in the present study for the monolingual children as there is no equivalent effect on the intercept term. The relevant conclusions in this study are reached on the basis of the group by ambiguity interaction on the linear term, which reflects differences in the change (increase or decrease) over time and are consistent with other studies finding garden path effects in the eye-tracking data. Age is a possible explanation for this difference as the children in this study are older than those in other studies. The challenge to this explanation is that adults have also been shown to experience garden-pathing in eye-tracking tasks (e.g. Pozzan & Trueswell, 2016; Trueswell et al. 1999).

The studies reviewed on garden-path effects with eye-tracking have so far reported the total looking time towards the correct or incorrect destination over a given time window; their main finding has been more looks towards the correct destination in the unambiguous condition. The current study replicates previous studies in that this effect is found on the intercept term for looks to the correct destination in the CD region. Previous studies have plotted looks towards a given picture but analysed the total number of looks by trials within a given time window (e.g. Trueswell et al., 1999 for children and adults; Pozzan & Trueswell, 2016 for adults only, Snedeker & Trueswell, 2004 for children). Previous studies have not analysed how looks towards a picture increase or decrease in the critical time region for garden-path sentence but have provided inferential statistics mainly for the overall proportions of looks within a time window. One study which has adopted this type of analysis is Atkinson et al. (2018), but it does so for ambiguous wh-questions. The effects on the linear terms observed in this study for looks towards the correct destination in the CD region can be interpreted as being consistent with results from previous studies. The clearest evidence in support of real time garden-path effects in bilingual children consistent with those observed in monolingual children also comes from participants' looks towards the correct destination, i.e. the bag.

RQ2: Do bilingual children recover from and reanalyse garden-path sentences?

The effect of ambiguity on comprehension accuracy suggests that the bilingual children did not fully recover from the garden path effects. This is consistent with a range of other studies for either comprehension questions or act out tasks in both adults and monolingual children which have shown lower accuracy scores or VP attachment-consistent act outs (e.g. Hurewitz et al., 2000; Kidd & Bavin, 2005; Choi & Trueswell, 2010; Trueswell et al. 1999, among others) and have been interpreted as such. This effect of ambiguity is attested in the literature regardless of any further manipulations in the stimuli and experimental design. The absence, however, of any group effects or an interaction of group with ambiguity indicates that the preservation of the original misanalysis is not unique to the bilinguals. The speed at which participants selected a picture to answer a question does not provide any insight as to whether bilingual children have difficulty revising misinterpretations of temporarily ambiguous sentences.

Accuracy scores are in line with some studies (e.g. Kidd et al., 2011; although scores in that study were higher in the unambiguous condition than in this study respectively) including the original garden-path study using eye-tracking with children (Trueswell et al. 1999). Accuracy scores in this study are noticeably higher than in Hurewitz et al. (2000) where it is about 30% for the comprehension / act out task. A possible reason for this is that, in the Hurewitz et al. study, questions which required the selection of a specific entity (i.e. *which questions*) out of a subset or to assign an attribute to an entity (*What kind* …?) were used. The questions may have been more difficult as they are more complex semantically (for a discussion on this as well as empirical evidence, albeit from impaired participants, see Salis & Edwards, 2008). In this study, all comprehension questions were wh-questions with the word *where*. This additional complexity is absent and, therefore, the questions in this study may have been easier.

One should always keep in mind that, even for ambiguous sentences, accuracy scores are relatively high (70-75%) and that both bilingual and monolingual children do not actually always give correct responses even for unambiguous trials (accuracy about 80-85%). What this shows is that, while there is a consistently increased likelihood of an error occurring with ambiguous sentences, it is not the case that the bilingual children do not have the capacity to revise initial syntactic representations or to correctly interpret temporarily structurally ambiguous sentences.

The original study by Trueswell et al. (1999) explicitly reported an increased likelihood of participants performing an action that was consistent with a VP interpretation of the sentence (i.e., the ambiguous PP was interpreted as a modifier to the VP; i.e. the ambiguous PP was interpreted as the destination of the action verb). In other words, they moved the frog onto the plate as they maintained the initial interpretation that the PP was a modifier to the verb rather than a modifier to the noun (for similar results, see Kidd & Bavin, 2005). Unfortunately, E-prime only recorded whether the participant's response was accurate or not but did not record the specific picture selected when the trial was inaccurate. This means it is possible to establish that the participants had persistent difficulty with sentences with temporary ambiguity and for sentences where there is an expected need to revise one's interpretation. It does not allow one to claim with certainty that it is the misinterpretation of a temporarily ambiguous NP as attached to a VP that lingers by relying on the accuracy scores alone. If E-prime had recorded the exact image selected as an answer to the comprehension question, it is possible that the majority of inaccurate responses involved the selection of the incorrect destination as the answer or that this selection was made more often for the ambiguous condition than the unambiguous condition.

The difficulty in recovery and reanalysis for ambiguous sentences relative to unambiguous ones is evidenced by the effects of ambiguity on accuracy only. There is no evidence of difficulty in the reaction times. There is no specific reason for which an effect of ambiguity should be observed in both accuracy measures and reaction times. In fact, given speed-accuracy trade off, it is reasonable to expect that effects will be observed in one of the metrics only. Previous studies which have used act out (e.g. Snedeker & Trueswell, 2004; Trueswell et al., 1999) have not recorded the time participants need to undertake the action but have shown lower accuracy for ambiguous sentences in the act out. Studies on garden path effects (e.g. Papangeli & Marinis, 2010) which include reaction time as a metric have used self-paced reading, but these are reaction times to words or segments of a sentence and not the comprehension question. In this study, ambiguous sentences usually had longer reaction times at the segment where there is disambiguation.

RQ3: Do bilingual children use referential context for disambiguation?

This study did not show conclusively that monolingual and bilingual children used referential context, i.e. information in the visual stimuli to disambiguate. The comprehension accuracy showed no effect of referential context and no interaction of context with ambiguity which would be expected if the presence of the second referent had a disambiguating function in the garden-path sentences.

It is perhaps in the accuracy data that the absence of an effect of referential context is best demonstrated. In terms of gaze data, the picture is less clear. There were no effects of referential context for looks either on the intercept or linear terms to the incorrect destination in the ID region where the effect of the referential context is expected given its predicted function. This means that neither group looked to the plate less when there were two referents in the visual stimuli when they heard the ambiguous PP. For looks to the correct destination, the effect is expected to be in the opposite direction with the presence of two referents leading to more looks towards the correct destination. This, again, was not observed in the ID region for either group. Seeing an apple on a plate whilst hearing the phrase “on the plate” simply did not guide the children towards an NP-modifying interpretation of the ambiguous phrase. This is similar to other studies which have shown no effects of referential context in children (Snedeker & Trueswell, 2004; Trueswell et al., 19999). This is inconsistent with Meroni & Crain (2003) who showed that children were indeed able to utilise referential context when the perceptual salience of the contrast in the visual stimuli facilitated a contrast between them (e.g. two napkins with a different colour for the example of the frog on the napkin). These results were however obtained for act out accuracy; as such, participants had time to consider their response before executing their action. It is unclear whether the children would be able to integrate this information fast enough in real time as captured by the visual-world paradigm. A limitation of the Meroni & Crain study is that there is no comparison between the one- and two-referent conditions as in this study and Trueswell et al. (1999) or Snedeker & Trueswell (2004).

While the ID region is probably the most critical region for testing the effect of referential context as a disambiguating cue, it is still possible to function as such later in the sentence. Effects of referential context are absent in the CD region as well. The effect of referential context emerges at the end of the sentence; in the final region (i.e. when the participants heard the phrase "before going to school"), where there is a main effect of referential content for looks towards the correct destination (the bag), with more looks to the correct destination in the two-referent condition than in the one-referent one. This effect, however, is not limited not the ambiguous condition only. It is assumed that the second referent would help disambiguate in the ambiguous condition and that there would be little need for this in the unambiguous condition. This effect is not only manifest on the intercept term (looks overall) in this region but also on the linear term (change over time); looks to the correct destination decline more slowly for the two-referent condition.

In sum, the data do not show a strong effect of referential context in aiding disambiguation for the ambiguous sentences only. Moreover, it does not appear present at the critical segment. Some effects of referential context are observed in the final segment, but these are not limited to the ambiguous condition contrary to expectations.

4.5. Conclusion

This study examined how bilingual children process temporarily structurally ambiguous sentences which involve (reduced) relative clauses relative to monolingual children. Existing work has shown that monolingual children build syntactic representations and interpretations of sentences in real time and are sensitive to garden-path effects. In this sense, children are qualitatively like adults. The predominant difference established in the literature is that children appear to have greater difficulty in revising their original interpretations and using contextual information during processing. Much less is known about sentence processing in bilingual children. Using a similar paradigm as Trueswell et al. (1999), we further examined whether the bilingual children utilised the referential context to disambiguate. Our findings suggest that bilingual children build syntactic representations incrementally as they had greater difficulty with temporarily ambiguous sentences than sentences without initial structural ambiguity. In addition to this they appeared to have difficulty revising initial misinterpretations. This is consistent with the monolingual children in the study and broadly in line with previous work on monolingual children. Bilingual children did not appear to make use of referential context at the point of ambiguity, and neither did the monolinguals (consistent with Snedeker & Trueswell, 2004; Trueswell et al., 1999). The main point of divergence between monolinguals and bilinguals is that, for the gaze data, the effects in bilinguals do not emerge at the time windows of interest but later although they are in the same direction. Consistent with previous works on processing of grammatical violations in bilingual children (Chondrogianni & Marinis, 2012; Chondrogianni et al., 2015; Marinis & Saddy, 2013) which have showed generally slower reaction times for bilingual children, this may be taken to mean that processing was slower in the bilinguals rather than that garden path effects were absent in the earliest segments of the sentence.

Chapter 5 Garden-path sentence processing in bilingual adults

5.1. Introduction

Overview

This chapter presents the final study in this thesis. The study expands the existing body of literature on garden-path sentences in bilinguals, which has focused on differences between monolingual natives and L2 learners with late and classroom-based exposure to English. Its contribution lies in that it examines processing in bilinguals with early and naturalistic exposure to the second language and who have attained nativelike proficiency. It compares this group of native bilinguals to monolinguals and late bilinguals / L2 learners. The chapter begins by reviewing the literature on locally ambiguous sentence processing in bilingual adults. It then presents the results from this study and discusses them. Information on the participants and the experimental design is available in chapters 3 and 4 respectively.

Processing of garden-path sentences in bilingual adults

Research in sentence processing in bilinguals has focused on adults and more specifically on those who acquired the second language late (usually aged six and older; for a review, see Cunnings, 2017). These bilinguals are often considered L2 learners and the debate has been framed by the contrast between native and non-native language processing. Work on early bilinguals or adults who, as children had an additional language, is still exceedingly rare. There are several studies which have examined garden-path effects in bilinguals, i.e. L2 learners of English in recent years some of which have used eye-tracking. Some studies (e.g. Pozzan & Trueswell, 2016) test PP attachment in reduced vs. unreduced relative clauses as those used in Trueswell et al. (1999) and in the study in Chapter 3 of this thesis. There are exemplified again in (40a) and (40b).

(40a) Put the apple on the plate in the bag (reduced relative clause; temporarily ambiguous)

(40b) Put the apple which is on the plate in the bag (unreduced relative clause; unambiguous)

In terms of other studies (Jacob & Felser, 2016; Roberts & Felser, 2011), several use sentences with subject-object ambiguities, i.e. sentences where an NP is initially misinterpreted (usually as a subject instead of an object) as in (41);

(41) While Mary fed the baby slept.

In their study with Italian L2 learners of English, Pozzan & Trueswell (2016) replicated the experimental design in the seminal Trueswell et al. (1999) study on garden-path sentences where participants heard sentences as in (40a) and (40b) whilst looking at pictures. Sentence ambiguity was crossed with number of referents in visual stimuli (1 vs. 2). In terms of accuracy of act outs, there was also a main effect of group, ambiguity and referential context, measured by the participants' movement of the object towards the incorrect destination. There were overall more incorrect actions suggesting misinterpretation in the L2 learners than native speakers, the ambiguous sentences relative to the unambiguous ones and the one-referent context relative to the two-referent context. Pairwise comparisons by ambiguity revealed an effect of referential context only for the ambiguous sentences (higher accuracy for the two-referent condition). The effect of group was significant in both ambiguous and unambiguous sentences but was larger in the ambiguous sentences. This suggests that the L2 learners had particular difficulty in revising the initial misinterpretation. In terms of the effect of referential context on the act out accuracy, the L2 learners had higher act out accuracy in the two-referent context relative to the one-referent context in the ambiguous sentences but the difference was much smaller for the unambiguous sentences. Overall, the difference in accuracy between monolinguals and L2 learners was more pronounced in the 1-referent context. There were minimal errors in the act out by the monolingual adults in Pozzan &

Trueswell and these were made exclusively for the ambiguous sentences. This was also the case in the Trueswell et al. study when comparing monolingual adults to children. The fact that these effects are found in the accuracy data, in addition to the gaze data which captures real time processing of the sentence, suggests that the difficulty in processing is not limited to the disambiguating segment but persists at least until the end of the sentence.

For looks to the incorrect destination, i.e. looks to the plate, in the region between the ambiguous PP until the onset of the disambiguating NP (the bag), Pozzan and Trueswell found there was no main effect of group but a main effect of ambiguity, referential context and an interaction of ambiguity with group and with referential context. Both groups looked at the incorrect destination more in the ambiguous condition than in the unambiguous one suggesting a garden-path effect. The absence of a main effect of group suggests that both groups look towards the incorrect destination equally, suggesting a similar effect of initial misinterpretation. The main effect of referential context indicates more looks towards the incorrect destination in the one-referent context where the second referent is not present. The interaction of ambiguity by referential context is due to the fact that the benefit conferred by the presence of the second referent was larger in the ambiguous sentences where it would have a disambiguating function. The group by ambiguity interaction suggests larger effects of ambiguity in the L2 learners. Taken together, the results from the gaze data in Pozzan & Trueswell suggest that both groups misinterpret the ambiguous PP as a modifier to the verb rather than the noun phrase in the one referent conditions, replicating Trueswell et al for the adult group. It also suggests that both could utilise referential information to aid disambiguation consistent with the adults but in contrast to the children in the aforementioned study.

Pozzan & Trueswell (2016) further modelled the likelihood of an erroneous act out on the amount of looks towards the incorrect destination. They found that looks to the incorrect destination predicted the chance of a wrong act out; more looks to the incorrect destination increased the chance of an erroneous act which was overall higher in the L2 learners than in the monolinguals. However, they also report an interaction of group with ambiguity as a predictor of a wrong act out in the same model. This suggests that, even when looks towards incorrect destination are accounted for, ambiguity impacts the ability of participants to revise an initially erroneous representation depending on whether one is a native speaker or an L2 learner. This was interpreted as evidence for the fact that the L2 learners were less able to revise interpretations following garden path effects during listening. They found that looks to the incorrect destination predicted the chance of a wrong act out.

Similar findings have been obtained for monolingual adults in Slattery, Sturt, Christianson, Yoshida & Ferreira (2013) who showed that garden-path effects impact the processing of subsequent sentences as measured by longer reading times. Taken together, there is evidence that the source of difficulty lies in the persistence of the initial misinterpretation.

The parser has been shown, however, to make use of other non-structural cues beyond the visual context during processing. For example, some studies have shown that recovery from garden-path effects is affected by the plausibility of interpretation differently in L2 learners and monolingual native speakers. In a study on subject-object asymmetries, Roberts & Felser (2011) carried out a self-paced reading study with comprehension questions in advanced Greek L2 leaners of English. Roberts & Felser controlled the plausibility of each interpretation so that for some sentences there were weak garden path effects as in (42) and for others there were strong garden path effects as in (43), where substantial revisions of syntactic structure occurred whilst controlling for plausibility

(42) The inspector warned the boss/crimes would destroy very many lives.

(43) While the band played the song/beer pleased all the customers.

For accuracy, Roberts & Felser (2011) found an interaction of group with plausibility for the weak garden path sentences; accuracy was lower for plausible sentences for the L2 learners, but this was not true for the monolinguals. The monolinguals had high accuracy in both conditions. The additional difficulty for the L2 learners indicates that they were not able to revise the ambiguous sentence when they needed to as often as the native speakers. Roberts & Felser found that there was a slowdown for implausible sentences relative to plausible ones for reading times at the disambiguating segment and thereafter for the weak garden path sentences. The effect, however, was reversed in the latter segments analysed, indicating easier reanalysis in sentences with implausible direct objects. Participants had difficulty processing the sentence when the implausibility rendered the initial interpretation unlikely, but this implausibility facilitated the switching from the old to the new interpretation. This was stronger, more prolonged and occurred earlier in the L2 learners than in the monolingual native speakers, reflecting a greater use of semantics by the L2 learners than the native speakers. Similar effects of plausibility were found for the L2 learners but not for the monolinguals in the strong garden path sentences. This entails that both groups showed garden path effects but that the L2 learners were more sensitive to plausibility in the recovery from these effects. Strong garden path sentences had lower accuracy scores than the weak garden path ones, suggesting that revision in L2 learners is less likely to be successful when more demanding.

Hopp tested processing and reanalysis of subject-object ambiguities in L2 English in native German speakers using an eye-tracking study with reading. Hopp (2015) used sentences with subject-object ambiguities as in (44a) and compared them to sentences without ambiguity due to punctuation, verb intransitivity, semantic implausibility or case-marked pronoun as in (6b-e respectively);

(44a) When the girl was playing the piano made some funny noises.

(44b) When the girl was praying, the boy made some funny noises.

(44c) When the girl was praying the boy made some funny noises.

(44d) When the girl was playing the boy made some funny noises.

(44e) When the girl was playing he made some funny noises.

Four critical regions were analysed: the verb of the adjunct clause (playing), the post-verbal noun (the piano), the verb of the main clause (made), and the post-verbal region, that is, the rest of the sentence (some funny noises). As Roberts & Felser did, Hopp (2015) observed greater difficulty in both monolinguals and L2 learners when the garden-path sentences contained plausible direct objects relative to when the subjects of the embedded clause were implausible direct objects to the verb in the main clause as well as the other control conditions. These was evidenced by longer first- and second pass reading times of the main clause verb where there is disambiguation but also total reading time and regressions in the same region relative to the others. The increased difficulty was assumed to reflect greater effort required for reanalysis. In terms of gaze data measures, this was evidenced by longer second pass reading times and total reading times as well as quantitatively more regressions. This was further evidenced by lower accuracy for sentences with plausible direct objects.

Building on the Roberts & Felser study, Jacob & Felser (2016) further examined garden-path effects by subject-object ambiguities in bilinguals, and more specifically, explored the effects of semantics disentangled from syntax on ambiguity resolution. To do this, they used sentences such as (45) where there is disambiguation twice;

(45) While the gentleman was eating the burgers were still being reheated in the microwave.

The sentence has two points of disambiguation according to Jacob & Felser; the first is the auxiliary verb *were* which signals that the NP “the burgers” is the subject of the second clause and not the direct object of the first one. The second point where the erroneous interpretation becomes implausible is the VP "being reheated". In a reading study with eye-tracking, Jacob & Felser compared sentences with ambiguity such as (45) to unambiguous sentences such as (46) where a comma makes the subject interpretation impossible.

(46) While the gentleman was eating, the burgers were still being reheated in the microwave.

If reanalysis is complete and undertaken based on syntactic structure alone, the effects of ambiguity will be observed only at the auxiliary verb segment. Yet if the original misinterpretation persists, then effects of ambiguity will be present at both the auxiliary and the main verb. Jacob & Felser further manipulated the length of the sentence by creating similar sentences with an additional clause, as in (47) which, does not contribute towards a subject interpretation of the ambiguous segment.

(47) While the gentleman was eating the burgers that were really huge were still being reheated in the microwave.

In terms of accuracy of comprehension questions, the L2 learners had lower accuracy than the monolinguals but did so consistently across conditions. There was lower accuracy for long and ambiguous sentences relative to short and unambiguous ones respectively; this did not interact with group. Five measures of gaze data (first pass time, regression path time, total reading time, second pass time, proportion of regressions) were reported for three regions (auxiliary verb "were still", main verb phrase "being reheated" and wrap up phrase "in the microwave"). Effects of group and ambiguity were found for all measures for all three regions analysed, which was also the case with ambiguity for most measures, indicating slower reading for the L2 learners and for ambiguous sentences. The fact that the effects of ambiguity were observed in both regions supports the claim that the original interpretation is not completely abandoned but that reanalysis continues. The effects of ambiguity were stronger for the monolinguals and the longer sentences in the first region only. This was interpreted by Jacob & Felser as evidence that both the monolinguals and the L2 learners are qualitatively similar in that they are garden-pathed and sensitive to ambiguity but that the L2 learners are less able to repair their misinterpretation. There were no interactions of group by ambiguity at the point of semantic disambiguation.

Other studies examined bilinguals' ability to detect garden path ambiguities using agreement and case disambiguation cues as well as the impact of the L1. Gerth, Otto, Felser & Nam (2017) used the self-paced reading paradigm with comprehension while testing L2 learners of German with Italian, Russian, Korean as their native language and German native speakers to investigate the processing of subject-object ambiguous sentences while manipulating case or agreement. Unambiguous sentences had a first NP with a different number to the main verb (agreement condition) or had a first NP with masculine gender where the accusative differs overtly from the nominative. Accuracy questions required a yes or a no as a response and were related to the argument structure of the sentences heard. All groups showed higher accuracy with both case and agreement disambiguation, reflecting sensitivity to morphosyntactic cues during processing even in L2 learners with intermediate proficiency. However, only the native speakers showed an interaction; this reflects the fact that the benefit of agreement disambiguation was larger than the benefit of case. The Korean L1 speakers had greater difficulty disambiguating by agreement than by case.

Effects of ambiguity were also found for reading times at the critical region and the spill-over region for all groups; reading times were slower for ambiguous sentences than unambiguous ones. During incremental processing, all four groups made use of case and agreement to disambiguate which resulted in faster reading time. A group by ambiguity interaction was due to the fact the L1 Russians showed delayed sensitivity to ambiguity relative to the other L2 learners while the L1 Koreans showed this effect early on. Apart from this there was no evidence to support the idea than the native language influences participants' patterns for reading time. Proficiency in German (as measured by score of the web-based placement test from the Goethe Institute website[30]) influenced the participants' ability to comprehend ambiguous sentences across groups; lower proficiency lead to a higher likelihood of being able to revise misinterpretations. The proficiency scores did not predict the difference in benefit between disambiguation by case relative to disambiguation by agreement.

L2 learners have also been shown to use previous context to aid disambiguation. In a self-paced reading study, Pan & Felser (2011) tested L1 Chinese learners of English. Participants heard sentences with temporary ambiguity ("with") which was then resolved towards either an NP or VP attachment as in (48a) and (48b);

(48a) Bill glanced at the customer with ripped jeans and then walked away.

(48b) Bill glanced at the customer with strong suspicion and then walked away.

Before each critical sentence, participants heard a couple of related sentences that bias a listener towards an NP or a VP interpretation, as in (49a) and (49b) respectively, which were crossed across the two types of experimental sentence.

(49a) Bill walked into a shop that he knew the police were keeping an eye on. There was only one other customer in the shop. The customer was wearing old and filthy clothes, whereas the sales assistant was dressed very smartly.

(49b) Bill walked into a shop that he knew the police were keeping an eye on. There were two other customers in the shop. One customer was wearing old and filthy clothes, whereas the other one was dressed very smartly.

---

[30] https://www.goethe.de/en/spr/kup/tsd.html

For the native controls, sentences disambiguated in favour of VP attachment were read faster after disambiguation than sentences with NP attachment disambiguation. This was consistent irrespective of the preceding context. However, this was not the case with the L2 learners who showed an interaction of the previous context with the attachment; they read sentences with NP attachment when the preceding context biased the listeners in that direction as (49a). Conversely, they read sentences with VP attachment when the context was in such a way biasing as (49b). This suggests the native speakers applied a more constant structurally-driven interpretation of the ambiguous sentences whereas the L2 learners fluctuated to a greater extent and made greater use of preceding context.

In sum, the emerging consensus from the currently available literature shows that even late sequential bilinguals, in other words, L2 learners, process sentences incrementally and thus, experience garden-path effects. Moreover, L2 learners appear to be able to use cues to aid disambiguation in real time. The main difference appears to be the relative difficulty L2 learners have in overcoming the garden path effect and reconstructing a syntactic representation / correct interpretation of a sentence once disambiguated. However, there is currently no published work, to our knowledge, which addresses the issue of language processing in simultaneous or early bilinguals. Therefore, what is unclear is how early bilinguals with naturalistic exposure to English process garden-path sentences. This study examines how early bilinguals with naturalistic exposure process garden path sentences in the form of reduced relative clauses and compares them to monolingual native speakers and adult later L2 learners of English.

The present study

This present study builds on existing research which has focused on the distinction between native and non-native processing and has thus, tested late sequential bilinguals. More specifically, it compared garden-path sentence processing in simultaneous or early sequential bilinguals with naturalistic exposure to English from childhood and who may be classed as native speakers of English to monolingual English speakers and late learners with mostly classroom exposure to English. The motivation for this study lies in the fact that it is unclear whether the difficulty to revise misinterpretations in L2 learners will be ameliorated in the native bilinguals due to a potentially higher degree of proficiency in the language tested. Furthermore, it can be motivated by the fact that differences were observed between the monolingual and bilingual children in the equivalent study in Chapter 4 in that the bilingual children did not show the same increase in looks to the incorrect destination in the ambiguous region as the monolingual children did. Yet, in children even at the ages of 8-11 years, the linguistic system is still in development. It, therefore, remains an open question as to whether these differences persist once fully adultlike competence is in place. This study used the same visual-world eye-tracking study with picture selection as in the study in Chapter 4. We manipulated the ambiguity of the sentence (i.e. temporarily ambiguous with reduced relative clause vs. unambiguous non-reduced relative clause) and the number of referents in the visual stimuli (one referent vs. two referents) as with the respective study in children.

The research questions for this study are;

1) Do native bilingual adults with early naturalistic exposure to English process structurally ambiguous sentences similarly to monolingual adults or L2 learners?

2) Do native bilingual adults recover from garden-path effects similarly i.e. equally successfully and over the same time course as monolingual adults or L2 learners?

3) Do native bilingual adults with early naturalistic exposure to English utilise number of referents as a disambiguating cue in interpreting garden-path sentences? If they do so, are

there differences between the two groups as an endstate result or in the change of gaze data over time?

Predictions for the first research question are the most straight-forward and can be addressed with the total looks towards a given stimulus by condition in each time window examined. It is expected that the native bilinguals will opt for structurally simple and lexically biased interpretation of VP attachment and, as a result, will experience garden-path effects. This means that for the gaze data there will be a) more looks to the incorrect destination after hearing the ambiguous PP and b) fewer looks towards the correct destination upon hearing the second disambiguating PP in the ambiguous condition than in the unambiguous one. It is further expected that this effect of ambiguity will be modulated by the referential context; the effect of ambiguity will be either restricted to or more pronounced in the one-referent condition and either absent or less pronounced in the two-referent condition where the presence of a second referent in the visual stimuli is posited to aid disambiguation. Similar results are also anticipated for comprehension accuracy and reaction times although these are not just a sign of garden-path effects but more so evidence for the participants' inability to recover from them.

In line with Pozzan & Trueswell (2016), it is expected that the L2 learners will experience garden-path effects but that they will have greater difficulty revising misinterpretation, evidenced in an effect of ambiguity on the comprehension accuracy. Yet, it is expected, they will still use referential context to disambiguate in the ambiguous condition. What is less clear given previous work is how the native bilinguals will perform relative to the other two groups and whether the effects observed so far will be stronger or less so for the group in question. It can be reasonably assumed that the bilinguals with early naturalistic exposure will pattern with the monolinguals and not with the L2 learners considering the predictions of the Shallow Structure Hypothesis. This will mean stronger effects of ambiguity on the comprehension accuracy or looks to the incorrect and correct destination for the ambiguous and disambiguating region respectively for the L2 learners relative in the bilingual natives and the monolinguals. Moreover, as previous studies only report proportions of looks to a specific stimulus as an aggregate in a given time window, little can be assumed about how these looks will change over time depending on the condition and by group. This is potentially important, as looks may not differ overall but a steeper increase for one group or condition may reflect more efficient processing or greater ease of processing respectively.

Predictions for the second research question are like those for the first one in terms of direction. The end-result of recovery from garden-path sentences should be reflected in the comprehension accuracy and potentially the time needed to select a picture to answer the comprehension question. The same effects of ambiguity modulated by the number of referents in the visual stimuli are expected to be found if participants have difficulty recovering from initial erroneous interpretations. In Pozzan & Trueswell (2016), this effect and interaction was indeed observed for act out accuracy and was particularly strong for the L2 learners. We anticipate a group by ambiguity and referential context interaction when comparing the L2 learners to the other two groups reflecting the effects found in Pozzan & Trueswell. What remains an open question is the performance of the bilingual natives relative to either monolinguals or L2 learners. Considering the Shallow Structure Hypothesis, one would reasonably predict that the native bilinguals will pattern similarly to the monolinguals rather than the L2 learners.

A further way to examine recovery from garden-path sentences in addition to end-result accuracy is the change in gaze data over time in the critical regions in which a participant hears the ambiguous or the disambiguating PP, or in the segments thereafter. This has not been undertaken in previous studies. Faster recovery from garden-path effects would be reflected in a steeper increase in looks towards the correct destination (potentially also a steeper decline in looks towards the incorrect destination) once the participants hear the second PP which disambiguates between interpretations. This would be assessed by examining the significance of the predictor variables and their interactions on the intercept term. An effect of structure – a slower rise in looks to the correct destination – is expected for the ambiguous sentences as participants will reorient their gaze to the correct picture slower and later. If the presence of a second referent aids disambiguation, then there will be an interaction of referential content (although a main effect is also consistent with this) and ambiguity on the intercept term. If one group shows faster recovery from garden path effects, this should be evidenced by an effect of group (and an interaction with other variables) on the intercept term. The most crucial region to examine this is the time window after the onset of the disambiguating phrase, i.e. the second PP, although the same effects subsequent regions may also be interpreted in a similar fashion. As recovery from misinterpretation is argued to be the main difficulty in L2 processing, it is most likely that the L2 learners will recover more slowly from the garden-pathing than the monolinguals even in the sentences for which they accurately answer the comprehension question. The existing literature does not make explicit predictions about the performance of the bilingual natives, but the previous hypotheses made for the first research question are also relevant at this point.

The third research question can be addressed primarily by the interaction of referential context with ambiguity; the specific metric is more informative about the stage at which the participants can utilise referential context to disambiguate. More specifically, it is expected that the presence of the second referent in the visual stimuli will aid processing in ambiguous sentences and much less so in unambiguous ones, although a main effect of referential context may still be present. Given the literature on L2 language processing (e.g. Pozzan & Trueswell, 2016), it is expected that these effects will be observed for all three groups. What remains to be seen is the magnitude and the timing of the effects across groups.

5.2. Methods

5.2.1. Participants

The participants were the same as those in Study 1a. For more details, see Chapter 3.

5.2.2. Design

The experimental design is the same as that for Study 1A. For more details, see Chapter 4.

5.2.3. Materials

The materials for Study 1BB were the same as those for Study 1A. For more details, see Chapter 4

5.2.4. Procedure

The procedure for Study 1BB was the same as in Study 1B. For more details, see Chapter 4.

5.2.5. Analyses

The same measures were analysed and are reported as in Study 1B. The analysis is the same as for this study. For more details, see Chapter 4. The same two sets of contrasts between participants were applied as in Study 1AB (i.e. model 1: monolinguals vs. bilinguals and then native bilinguals vs. L2 learners; model 2: native vs. non-native speakers and then monolingual vs. bilingual speakers) for the accuracy data and the reaction times. For more detail, see also Chapter 3.

5.3. Results

5.3.1 Accuracy

5.3.1.1 ML vs. BL; then NS BL vs. NNS BL

Figures 1 is a summary of the accuracy scores by group and condition.

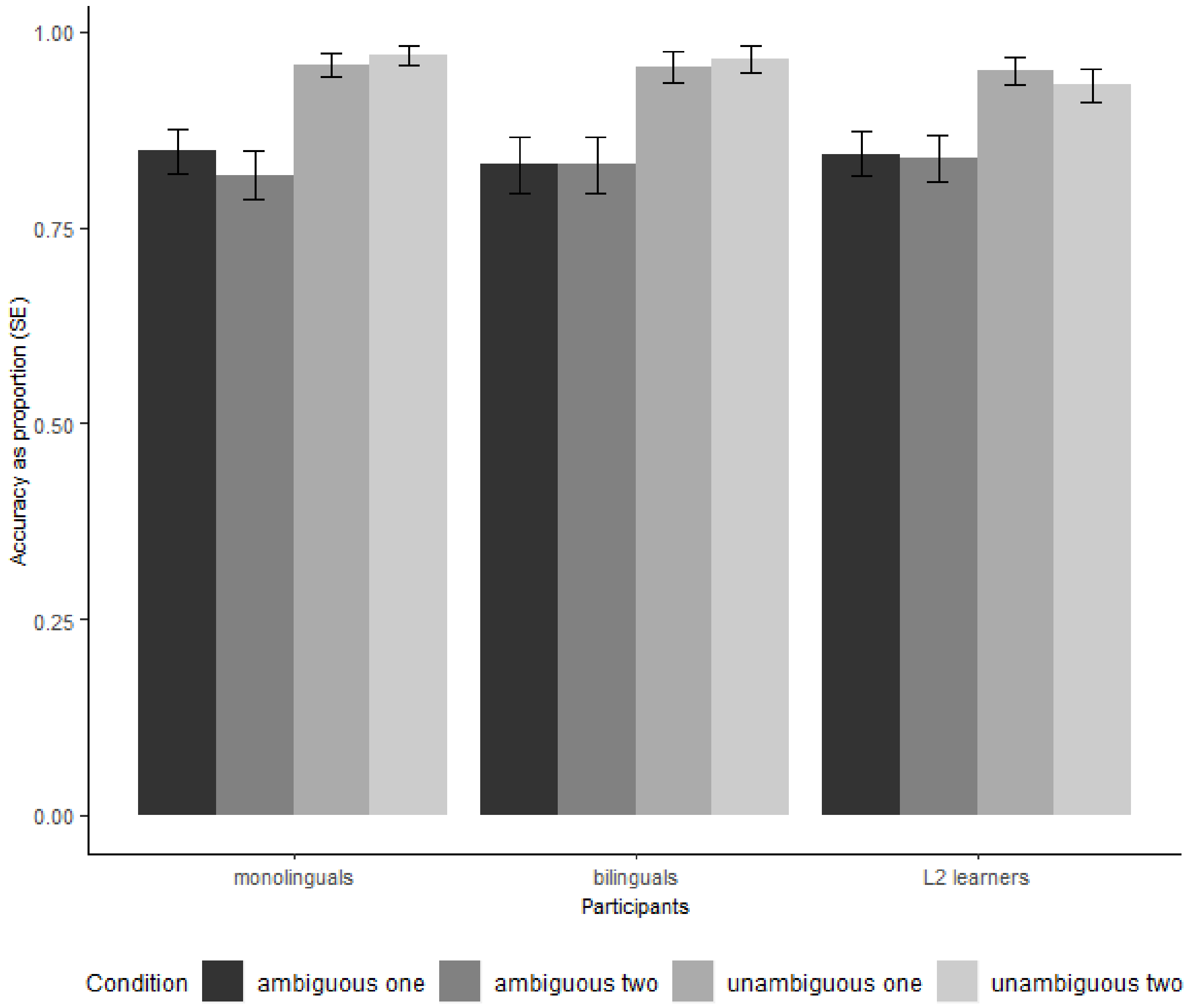

Figure 35 Accuracy as ratio by group and condition

There was a main effect of ambiguity (β = 0.8066, SE = 0.10, z = 7.833, p < 0.001) but no effect of group for either comparisons (β = -0.0907, SE = 0.20 z = -0.449, p = 0.653 for ML vs. BL; β = -0.0305, SE = 0.18, z = -0.167, p = 0.867 for NS BL vs. NNS BL) or context (β = -0.0099, SE = 0.10, z = -0.098, p = 0.922). No interactions were significant. Accuracy was higher for unambiguous sentences than for ambiguous ones across all groups and both referential contexts (see Figure 1).

5.3.1.2 NS vs. NNS; then ML vs. BL NS

As with the initial contrast, there was a significant effect of ambiguity only (β = 0.8065964, SE = 0.11, z = 7.833, p < 0.001) and no effect of group for either contrast (β = -0.0758, SE =

0.20, z = -0.384, p = 0.701 for NS vs. NNS; β = -0.0528, SE = 0.19, z = -0.284, p = 0.776 for ML vs. BL NS) or referential context (β = -0.0099, SE = 0.10, z = -0.098, p = 0.922). Again, no interactions were significant. Accuracy was higher for the unambiguous condition than for the ambiguous one (see Figure 1).

The most consistent finding is that accuracy is lower for the ambiguous sentences than for the unambiguous ones. The number of referents in the visual stimuli did not impact accuracy in either contrast. This may appear to be contradictory to the literature; if the referential context is expected to aid disambiguation, this is most plausibly going to happen if processing would otherwise struggle. In this case accuracy is overall higher than in the Pozzan & Trueswell study where for L2 learners' accuracy on the ambiguous one-referent condition was at chance performance. All three groups were equally as accurate.

5.3.2 Reaction Times

5.3.1.1 ML vs. BL; then NS BL vs. NNS BL

There was a trend towards an effect of group in the first contrast (β = 141.086, SE = 72.563, t = 1.944, p = 0.054; monolinguals vs. bilinguals) but not in the second contrast (β = -52.522, SE = 67.338, t = -0.780, p = 0.437; native speaker bilinguals vs. non-native bilingual speakers). There was also a main effect of ambiguity (β = -179.685, SE = 27.534, t = -6.526, p < 0.001) and context (β = 92.596, SE = 27.504, t = 3.367, p = 0.001). There was an interaction of ambiguity with group in the first contrast (β = -98.655, SE = 37.557, t = -2.627, p = 0.009), an interaction of context with group in the first contrast (β = 78.254, SE = 37.459, t = 2.089, p = 0.037), an overall ambiguity by referential context interaction (β = -57.520, SE = 27.499, t = -2.092, p = 0.037) and a trend for a three-way interaction of ambiguity by context by group in the first contrast (β = -65.197, SE = 37.450, t = -1.741, p = 0.082). Figure 2 provides an overview of the raw data by group and condition

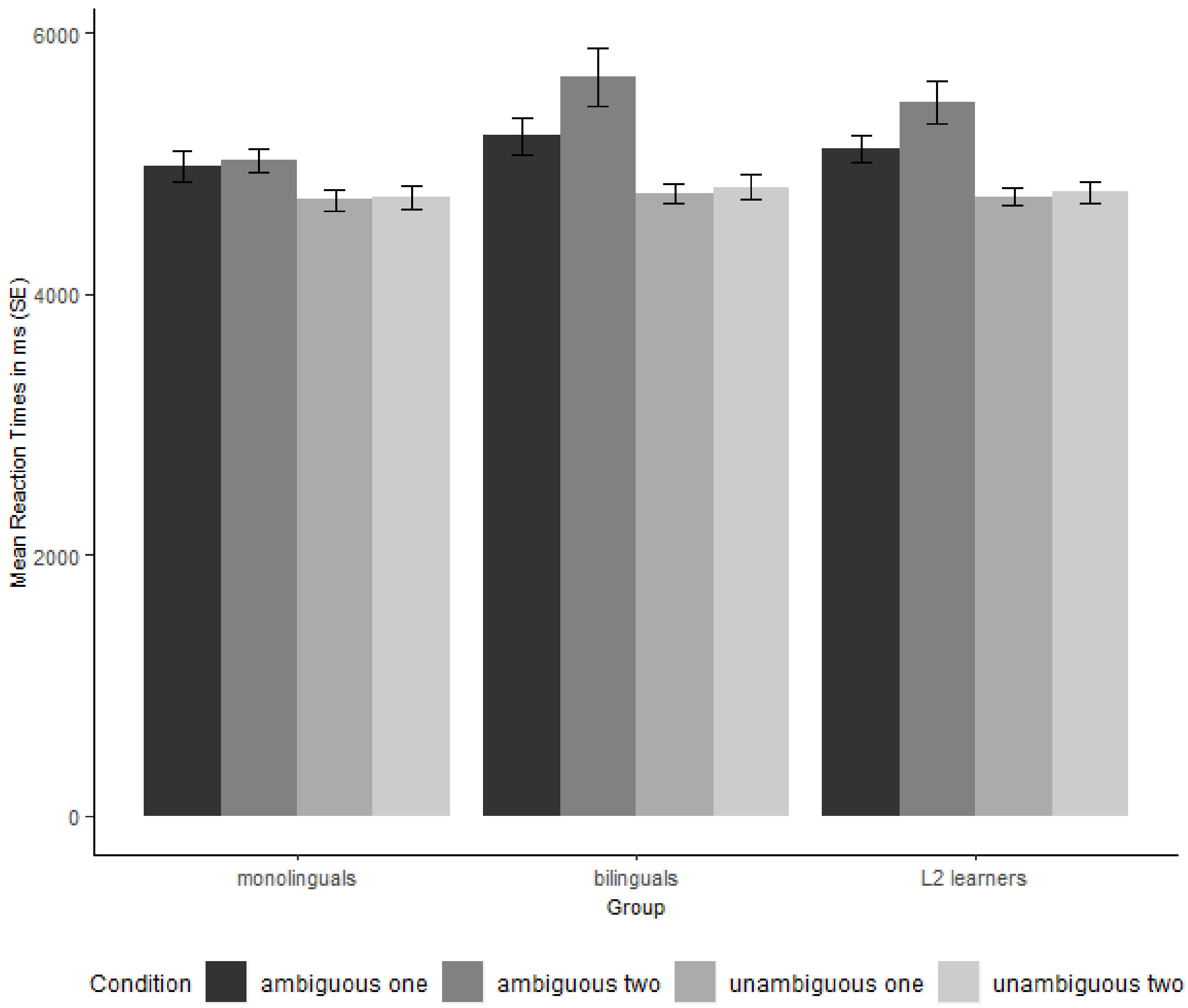

Figure 36 Average adjusted reaction time in ms (SE) by group and condition

The monolinguals were slightly faster than the bilinguals in both groups. Across groups, reaction times were longer for ambiguous trials than for unambiguous ones and for trials in the one-referent context than the two-referent context, although the latter is more pronounced in the two groups of bilinguals (hence the relevant interaction). Pairwise comparisons by ambiguity showed an effect of referential context in the ambiguous trials ($t(726) = -2.1593$, $p = 0.031$) but not the unambiguous ones ($t(830) = -1.1575$, $p = 0.247$).

The effect of ambiguity was significant in both one-referent ($t(562) = 3.4847$, $p = 0.001$) and two-referent contexts ($t(555) = 4.1776$, $p < 0.001$) but larger in the latter. This entails that, while ambiguity was a consistent source of difficulty for the participants from all groups, the effect of the second referent in the visual stimuli was limited to the ambiguous sentences. Surprisingly however, it was in the opposite direction; reaction times were slower for comprehension questions in 2-referent conditions than for those trials in the 1-referent conditions.

To examine the ambiguity by group interaction in the first contrast, the data were split into data from monolinguals and the two groups of bilinguals. The effect of ambiguity was significant in the bilingual groups ($W = 133170$, $p < 0.001$) but only trending for the monolinguals ($W = 46735$, $p = 0.080$)[31]. This suggests, and is confirmed by visual inspection of the data, that both groups of bilingual speakers slowed down more on the ambiguous sentences relative to the unambiguous ones than the monolingual participants. The marginal three-way interaction and a visual inspection of Figure 2 suggests that the unexpected effect of context found in the ambiguous but not unambiguous sentences, as discussed above, is driven by the bilinguals rather than the monolinguals.

[31] For pairwise comparisons to break down significant interactions, a Wilcoxon signed-rank test was used was the data was not normally distributed instead of a paired-sampled t-test.

5.3.2.2 NS vs. NNS; then ML vs. BL NS

There was an effect of group in the second contrast (monolingual vs. bilingual native speakers; (β = 132.23, SE = 66.80, t = 1.979, p = 0.048) but not for the first contrast (native vs. non-native speakers; β = 17.72, SE = 73.48, t = 0.241, p = 0.810). There was also a main effect of ambiguity (β =-179.75, SE = 29.39, t = -6.116, $p < 0.001$) and referential context (β = 92.86, SE = 27.37, t = 3.393, $p < 0.001$). There was an ambiguity by context interaction (β = -57.63, SE = 27.37, t = -2.106, p = 0.035) and an ambiguity by group interaction in the second contrast (β = -78.19, SE = 36.98, t = -2.114, p = 0.0370); the context by group in the second contrast interaction also trended towards significance (β = 61.03, SE = 34.44, t = 1.772, p = 0.077). As with the accuracy data, the results from the second contrast corroborate the findings from the first contrast which point to an overall difficulty caused by ambiguity, longer reaction times for the 2-referent condition, which is exclusive to the ambiguous, sentences and that both groups of bilingual speakers appear to be clustering with the other rather than with the monolinguals. The latter was not observed in the accuracy data where analyses yielded no effects between any of the groups.

5.3.3 Gaze data

5.3.3.1. ID region ("on the plate")

Looks to Incorrect Destination

Contrast 1: ML vs. BL; then NS BL vs. NNS BL

Figure 3 provides an overview of the looks to the incorrect destination in the ID region where there is ambiguity and potential for garden-path effects. There was a main effect of group in the first contrast on the intercept term (monolinguals vs. bilinguals, β = 0.045, SE = 0.02, t = 2.016, p = 0.047) but not on the second contrast (native bilingual speakers vs. L2 learners; β

= -0.013, SE = 0.02, t = -0.627, p = 0.532). There was also a main effect of ambiguity (β = -0.048, SE = 0.01, t = -3.968, p < 0.001) and context (β = -0.0241, SE = 0.011, t = -2.113, p = 0.035) on the intercept term while the linear term for time was also significant (β = 3.323, SE = 0.49, t = 6.857, p < 0.001). There were two trends on the intercept term; group by ambiguity in the first contrast (β = -0.029, SE = 0.02, t = -1.784, p = 0.078) and context by ambiguity by group in the second contrast (β = -0.025, SE = 0.01, t = -1.705, p = 0.089).

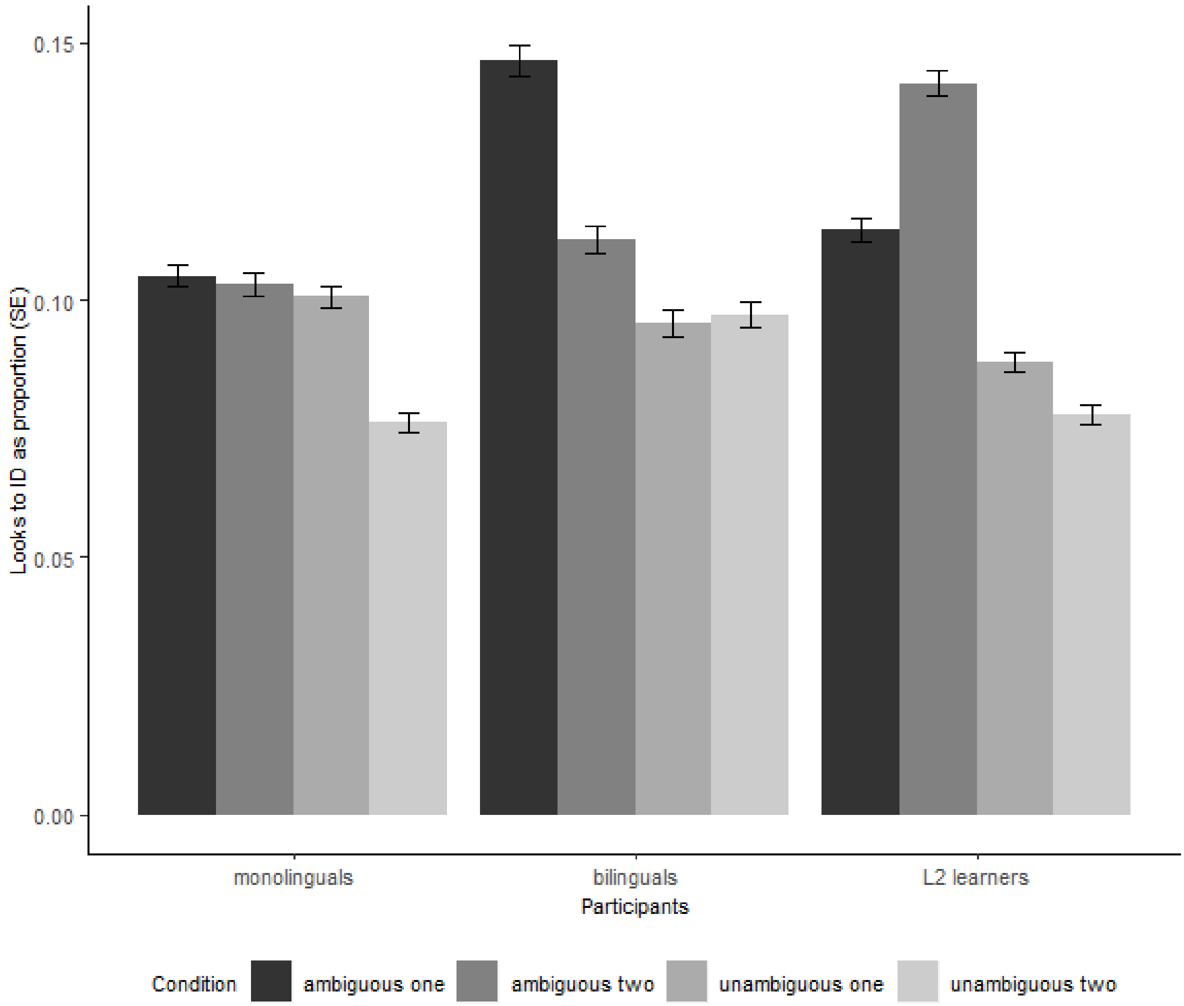

Figure 37 Looks to incorrect destination (the plate) in the ID region (i.e. "on the plate") by group and condition overall

There were fewer overall looks to the incorrect destination in the ID region for monolinguals relative to the two bilingual groups, but the two bilingual groups did not differ to one another in this respect. Overall, there were also more looks towards the incorrect destination image for the ambiguous sentences than for the unambiguous ones. Taken together, this suggests that the increased looks towards the incorrect destination picture reflects garden-path effects, as the participants interpreted the ambiguous PP as modifying the verb. There were also more looks to the incorrect destination in the one-referent condition overall suggesting the participants used the presence of the second referent in the visual stimuli to disambiguate. The interaction of group by context and the three-way interaction confirm a visual overview of the data that this is not consistent across groups.

Pairwise comparisons were carried out between the monolinguals and both groups of bilinguals to address the group by ambiguity interaction. For the bilinguals, there was a main effect of ambiguity: looks to the incorrect destination were more when the sentence was ambiguous ($\beta = -0.062$, $SE = 0.02$, $t = -3.92$, $p < 0.001$). This was completely absent in the monolinguals ($\beta = -0.014$, $SE = 0.02$, $t = -0.79$, $p = 0.434$). Despite a significant effect of referential context, this effect was not observed in the pairwise comparisons. No group by ambiguity and/or referential context interaction was observed in the model for the data for the two bilingual groups. In other words, as confirmed by visual inspection of the data (Figure 3), the group by ambiguity interaction for the first contrast can be attributed to the fact that there are effects of ambiguity in both bilingual groups across referential contexts but in the opposite direction. These are not observed not for the monolinguals.

The three-way interaction between group in the second contrast, ambiguity, and referential content shows that the bilinguals looked at the incorrect destination in the ambiguous condition when there was one referent more so than when there were two. The reverse was true for the L2 learners. The interaction was, therefore, examined by refitting the model without between participant comparisons to the data from the bilinguals and the L2 learners separately. The bilinguals showed a strong effect of ambiguity ($\beta = -0.08$, $SE = 0.02$, t = -

3.72, $p < 0.001$) but no effect of referential context ($\beta = -0.002$, SE = 0.02, t = -0.08, p = 0.937) or an interaction between ambiguity and context ($\beta = -0.03$, SE = 0.02, t = -1.45, p = 0.147).The L2 learners showed only a trend for an effect of ambiguity and referential context ($\beta = -0.05$, SE = 0.03, t = -1.94, p = 0.064 and $\beta = -0.05$, SE = 0.03, t = -1.835, p = 0.067). The ambiguity by referential context interaction was not significant.

On the linear term, there was an effect of group in the second contrast ($\beta = -1.768$, SE = 0.61, t = -2.889, p = 0.004) but not first contrast ($\beta = 1.077$, SE = 0.66, t = 1.633, p = 0.103). Other main effects on the linear term were ambiguity ($\beta = -1.583$, SE = 0.48, t = -3.271, p = 0.001) and context. The effect of ambiguity on the linear term suggests that looks to the incorrect destination increase and remain more for the ambiguous conditions relative to the unambiguous ones while the respective effect on referential context is due to a brief increase in looks to the incorrect destination for the 1-referent condition. There was a group by ambiguity interaction for both contrasts ($\beta = 1.374$, SE = 0.66, t = 2.082, p = 0.037 and $\beta = -1.523$, SE = 0.61, t = -2.487, p = 0.013 respectively) and a group by context interaction in the second contrast on the linear term ($\beta = 1.535$, SE = 0.61, t = 2.507, p = 0.012). Figure 4 provides a visual overview of the changes in looks to the incorrect destination over time in the ID region by group and condition.

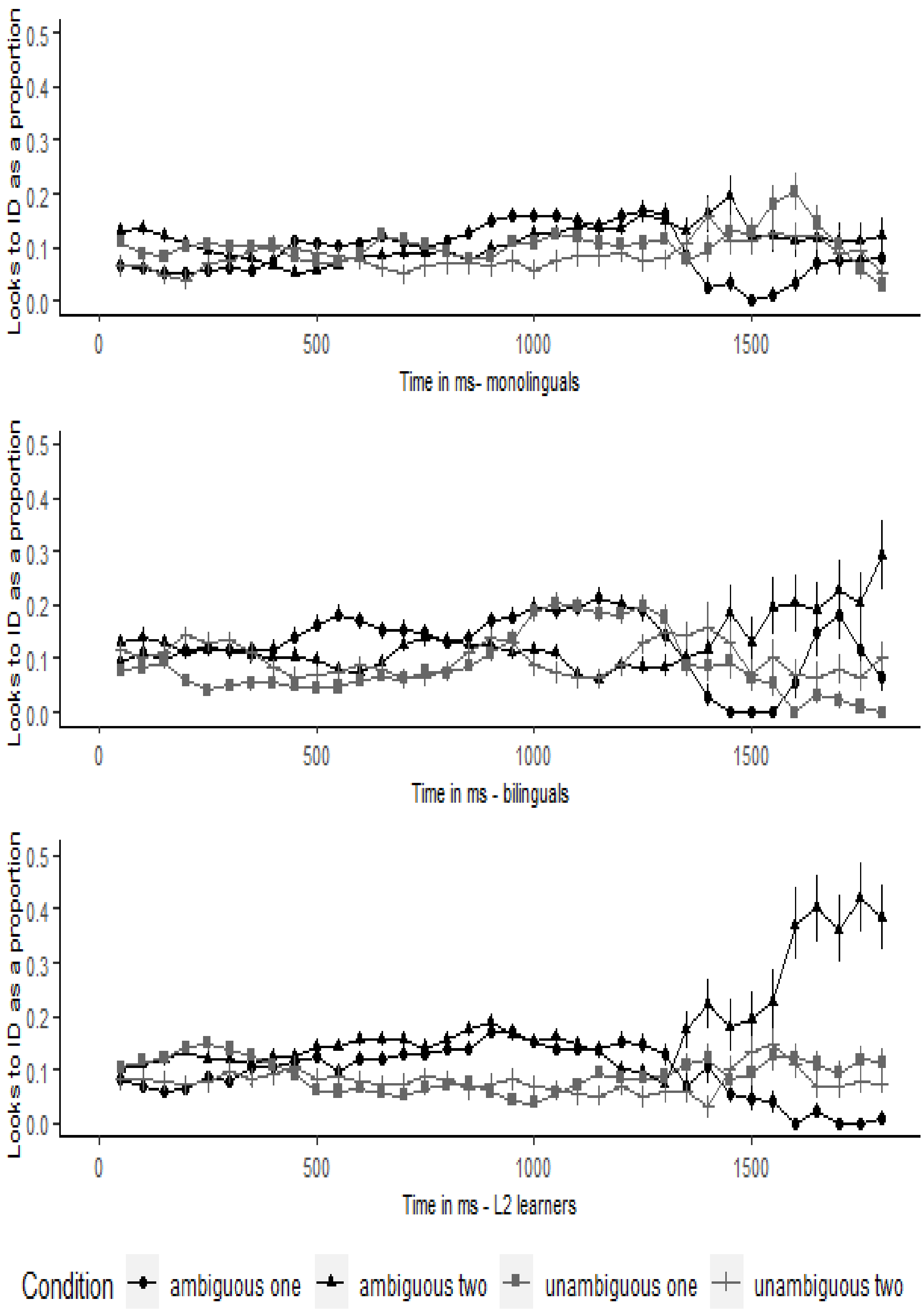

Figure 38 Looks to the incorrect destination (the plate) in the ID region (while the participants heard "on the plate") by group and condition over time

The results show that the three groups differed from one another in how the looks to the incorrect destination changed over time based on the ambiguity of the sentence but also that the native bilinguals and the L2 learners are differentially impacted by the presence of the second referent. The monolinguals showed a gradual increase in looks towards the incorrect destination in the segment of the sentence where garden pathing was expected. The bilinguals showed an earlier increase in looks towards the incorrect destination for the ambiguous sentences with one referent relative to the other three conditions 500-1,000ms post-onset. Looks to the incorrect destination were subsequently driven in the bilinguals by the referential context and were consistently lower for the 2-referent context. For the L2 learners looks to the incorrect destination increased faster and remained more throughout the segment for the ambiguous sentences relative to the unambiguous ones with the referential context seemingly having no impact. This suggests that the L2 learners showed the strongest garden-path effects. The main effect of ambiguity on the linear term suggests a steeper increase in looks to the incorrect picture which was persistent across the segment.

Contrast 2: NS vs. NNS; then NS ML vs. NS BL

For the second set of contrasts (native vs. non-native speakers, then monolinguals vs. bilinguals with nativelike competence), there was no effect of group in the first contrast (β = 0.0968, SE = 0.02, t = 0.426, p = 0.671) and a trend in the second contrast on the intercept term (β = 0.040, SE = 0.02, t = 1.978, p = 0.0507. There was also a main effect of ambiguity (β = -0.048, SE = 0.01, t = -3.968, $p < 0.001$) and referential context (β = -0.0241, SE = 0.01, t = -2.113, p = 0.035) on the intercept term while the linear term was also significant (β = 3.323, SE = 0.49, t = 6.857, $p < 0.001$), reflecting the findings from the contrast 1. There was also a marginal group by ambiguity interaction in the first contrast (native speakers vs. non-native, $\beta = -0.028$, $SE = 0.02$, $t = -1.679$, $p = 0.097$) on the intercept term. On the linear term, there was a main effect of ambiguity ($\beta = -1.583$, $SE = 0.48$, $t = -3.271$, $p = 0.001$) and referential context ($\beta = -1.963$, $SE = 0.48$, $t = -4.061$, $p < 0.001$), an effect of group in the second contrast ($\beta = 1.692$, $SE = 0.61$, $t = 2.782$, $p = 0.005$), a group by ambiguity interaction in the second contrast ($\beta = 1.792$, $SE = 0.61$, $t = 2.945$, $p = 0.003$) and a group by referential context interaction in the first contrast ($\beta = 1.368$, $SE = 0.67$, $t = 2.058$, $p = 0.0395$). The second set of contrasts reflects the findings in the first comparison.

Looks to Correct Destination

Contrast 1: ML vs. BL; then NS BL vs. NNS BL

There were no main effects or interactions on the intercept term. The linear term was significant ($\beta = 7.903$, SE = 0.58, t = 13.525, $p < 0.001$). There were two main effects and interactions on the linear term; a main effect of group in the monolinguals vs. bilinguals contrast ($\beta = 2.293$, SE = 0.79, t = 2.880, p = 0.004), a main effect of ambiguity ($\beta = 3.889$, SE = 0.58, t = 6.661, $p < 0.001$), a group by context interaction in the first contrast ($\beta = 1.626$, SE = 0.79, t = 2.043, p = 0.041) and an ambiguity by context interaction ($\beta = -1.336$, SE = 0.584, t = -2.288, p = 0.022). Looks to the correct image increased over time. They increased faster and began to do so earlier on for the unambiguous sentences. An increase in looks to the correct destination is not anticipated at this segment as the sentence has not yet been disambiguated. This could be explained on the basis of prediction. The participants know that the verb “put” needs a destination given its argument structure and, knowing that the second referent (i.e. the apple already one the plate) cannot be that look to the only remaining plausible picture (i.e. the plate). Figure 5 depicts the looks towards the correct destination in the ID region aggregated by group and condition while Figure 6 shows the respective change over time.

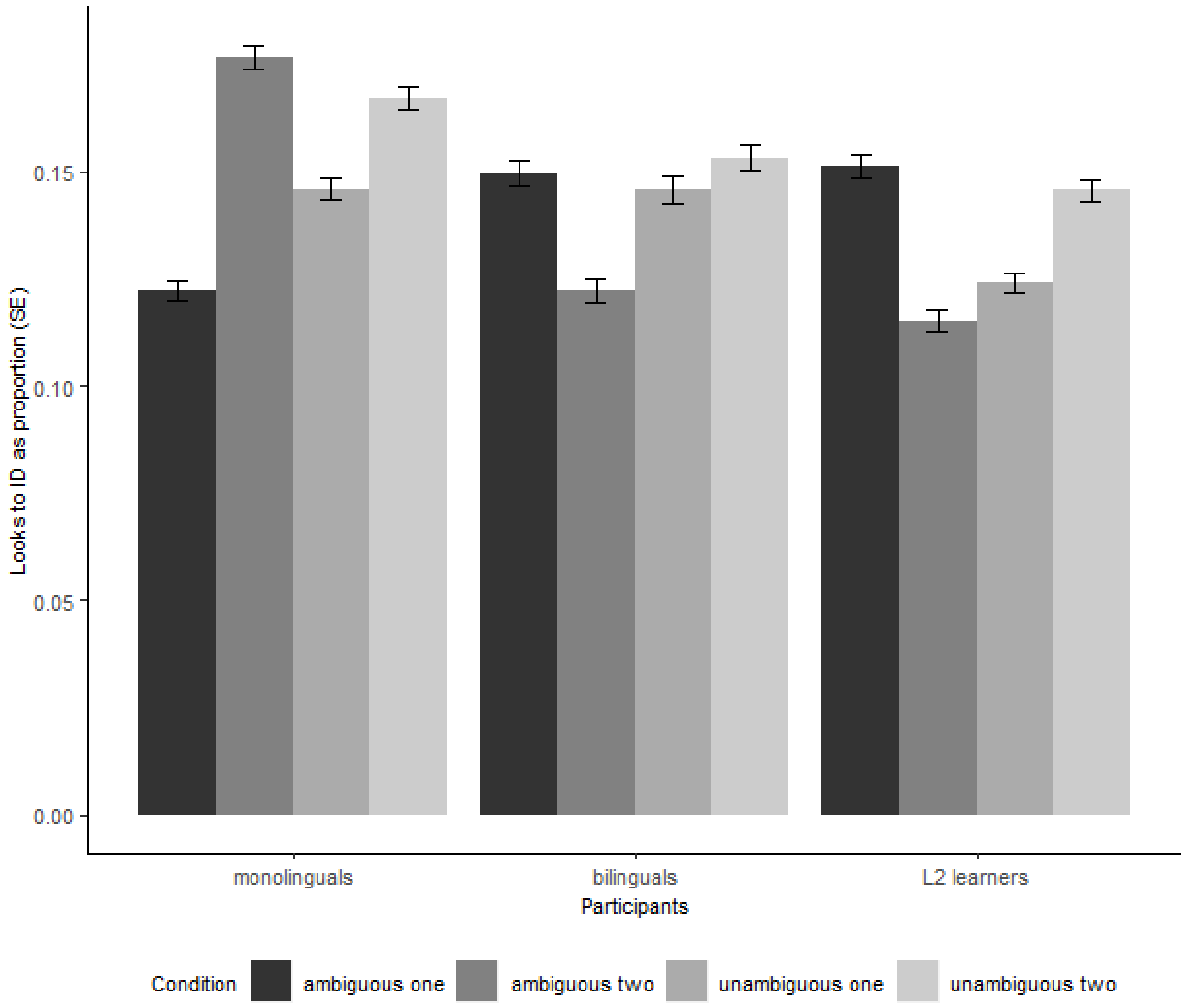

Figure 39 Looks to correct destination (the bag) in the ID region ("on the plate") by group and condition overall

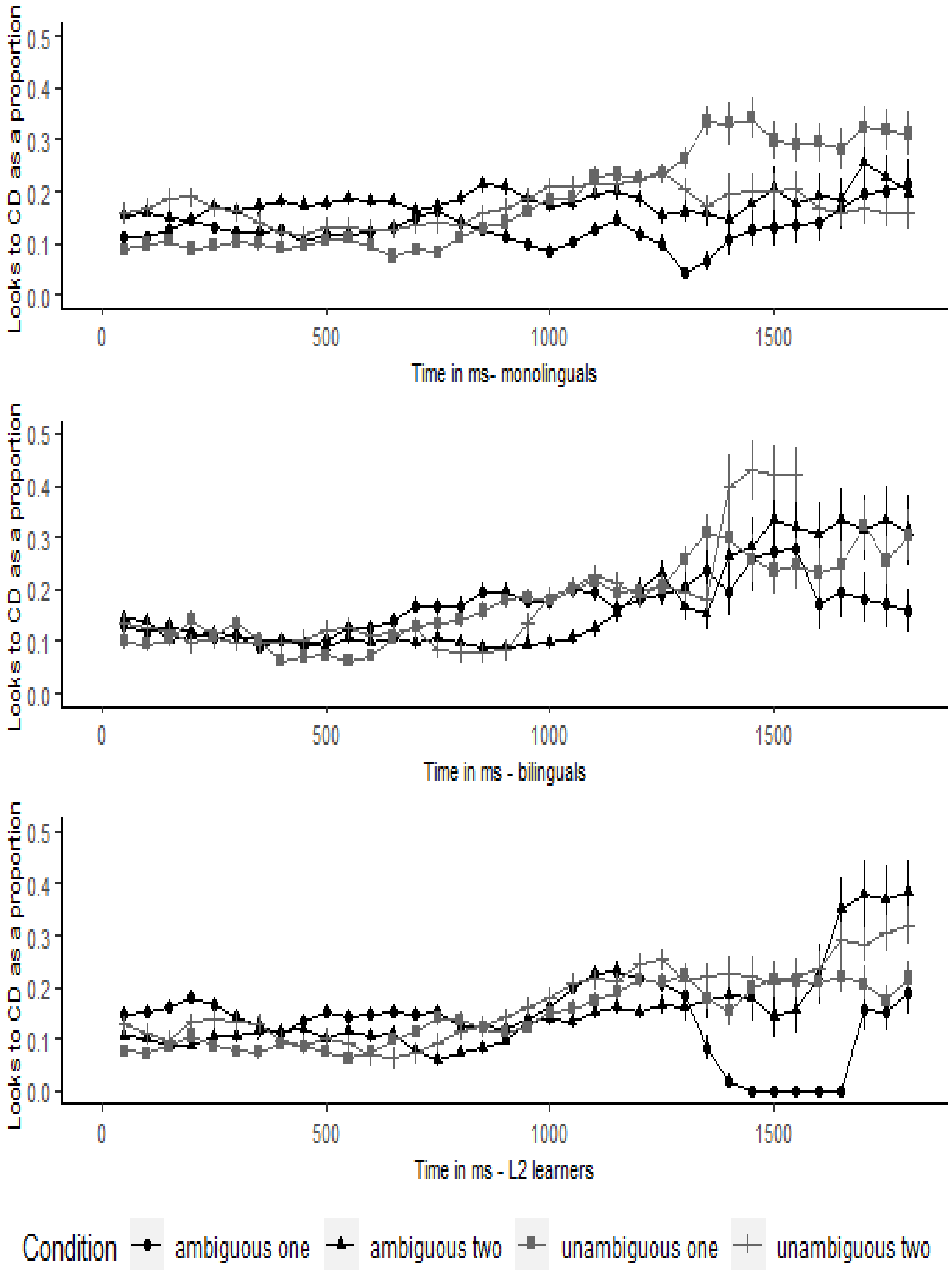

Figure 40 Looks to the correct destination (the bag) in the ID region (while the participants heard "on the plate") by group and condition over time

Contrast 2: NS vs. NNS; then NS ML vs. NS BL

Similarly to the first set of contrasts, there were no main effects or interactions on the intercept term. The linear term was significant ($\beta = 7.903$, SE =0.58, $t = 13.525$, $p < 0.001$). The same main effects and interactions were significant as with the previous contrast (main effect of group in the monolinguals vs. bilinguals contrast; $\beta = 1.743$, SE = 0.80, $t = 2.178$, $p = 0.029$; main effect of ambiguity; $\beta = 3.889$, SE = 0.58, $t = 6.661$, $p < 0.001$, group by context interaction in the first contrast; $\beta = -1.687$, SE = 0.80, $t = -2.107$, $p = 0.0351$).

5.3.3.2. CD region ("on the table")

Looks to Incorrect Destination

Contrast 1: ML vs. BL; then NS BL vs. NNS BL

There was no main effect of group for either contrast ($\beta = 0.009$, SE = 0.02, $t = 0.433$, $p = 0.669$ and $\beta = -0.01$, SE = 0.02, $t = -0.706$, $p = 0.487$) on the intercept term or an effect of ambiguity ($\beta = -0.016$, SE= 0.01, $t = -1.581$, $p = 0.114$). The only main effect on the intercept term was that of referential context ($\beta = 0.031$, SE = 0.01, $t = 2.809$, $p = 0.006$). Looks to the incorrect destination were higher for the two-referent context than for the one-referent context and are shown in Figure 7. There were no interactions, although the context effect was numerically greater in the ambiguous sentences.

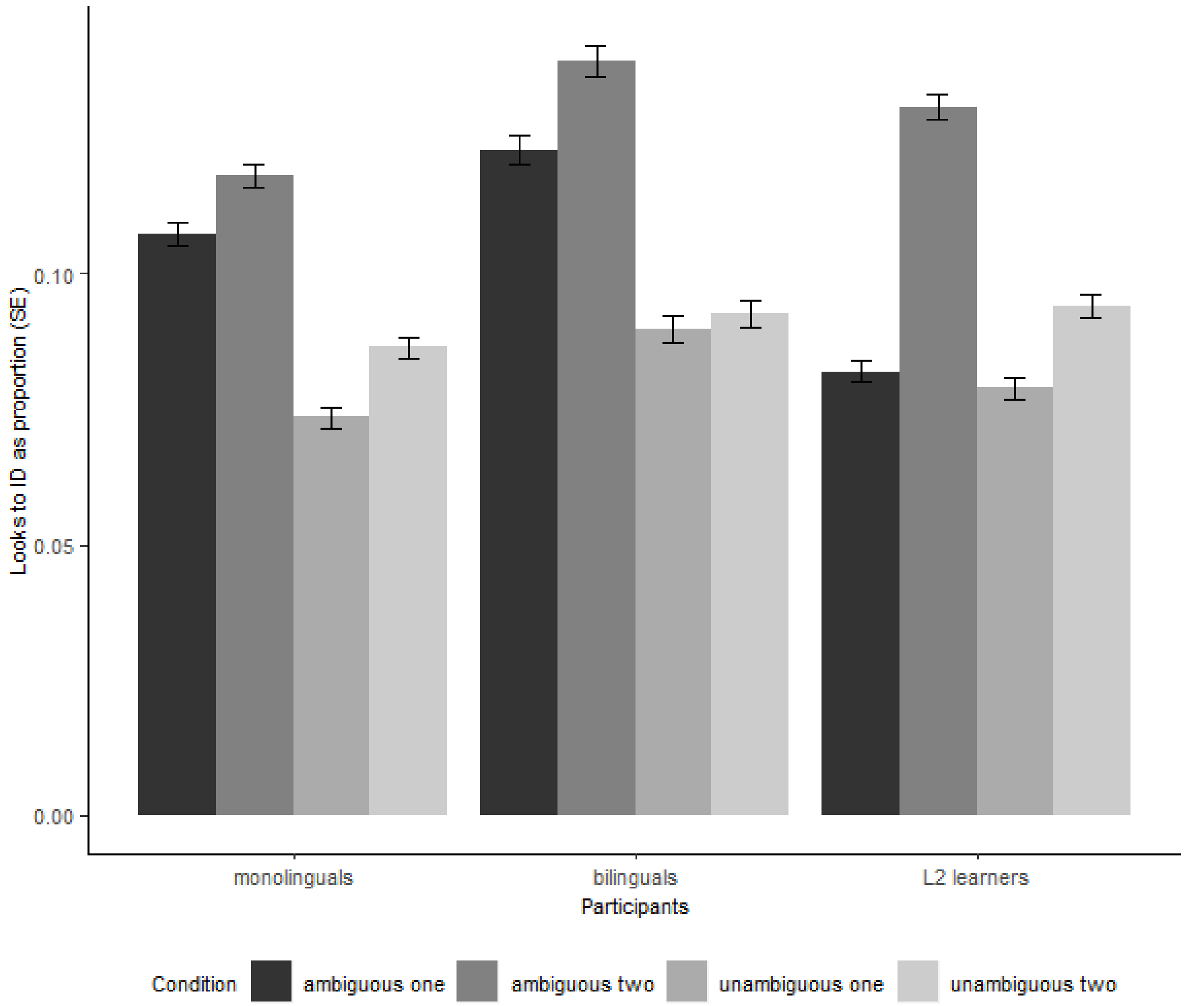

Figure 41 Looks to incorrect destination (the plate) in the CD region ("in the bag") by group and condition overall

The linear and quadratic terms were significant predictors (β = -1.581, SE = 0. 46, t = -3.424, p = 0.0001 and β = 1.497, SE = 0.46, t = 3.27, p = 0.001). On the linear term, there was a main effect of context (β = 1.748, SE = 0.46, t = 3.804, p < 0. 001), a group by ambiguity in the second contrast (β = -1.467, SE = 0.59, t = -2.502, p = 0.012), and a group by ambiguity by referential context in the second contrast (β = 1.226, SE = 0.59, t = 2.091, p = 0.037). These effects and interactions are depicted in Figure 8.

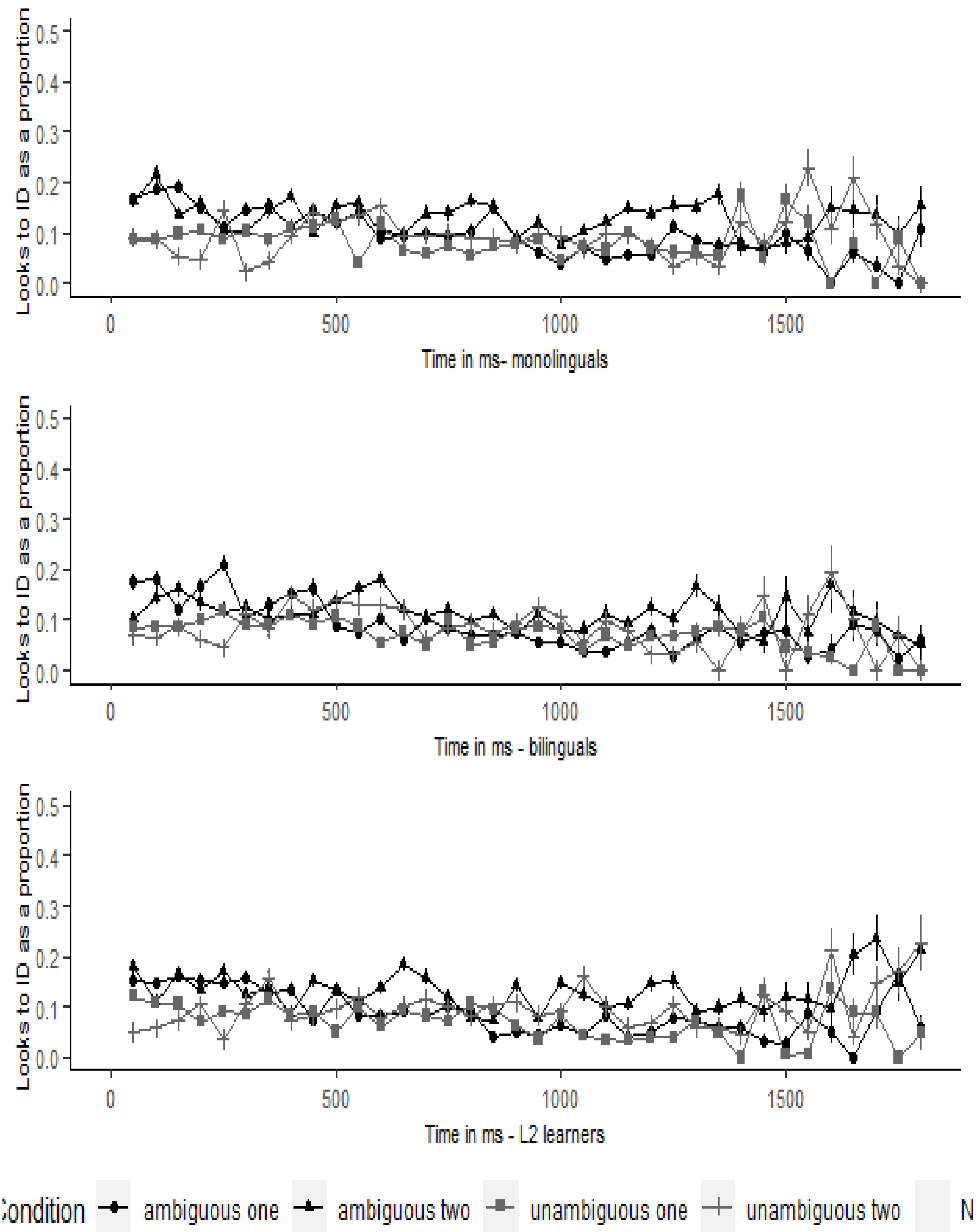

Figure 42 Looks to the incorrect destination (the plate) in the CD region (while the participants heard "in the bag") by group and condition over time

Looks to incorrect destination declined more rapidly in the one referent condition as evidenced in Figure 8. We do not interpret this interaction as it is not meaningful. The group by ambiguity interaction for the bilingual natives vs. L2 learners contrasts is due to the fact that looks to the incorrect destination remain mostly more for the ambiguous sentences than the unambiguous ones for the native bilinguals but for the L2 learners where there is a rapid decline in the respective looks in the first 500 ms of the critical region. This may indicate that the garden-path effects are lingering for longer in the bilinguals than in the L2 learners.

Contrast 2: NS vs. NNS; then NS ML vs. NS BL

The contrast between native speakers and non-natives first followed by monolingual and bilingual natives mirrored the results from the first contrast. There was only a main effect for referential context on the intercept term ($\beta = 0.031$, SE = 0.01, $t = 2.809$, $p = 0.006$) and the linear ($\beta = -1.581$, SE = 0.46, $t = -3.424$, $p < 0.001$) and quadratic term were both significant ($\beta = 1.497$, SE = 0.46, $t = 3.271$, $p = 0.001$). There was also a main effect of context on the linear term ($\beta = 1.748$, SE = 0.46, $t = 3.804$, $p < 0.001$) as well as a group by context interaction ($\beta = -1.482$, S = 0.64, $t = -2.331$, $p = 0.02$) and a group by ambiguity by context interaction ($\beta = 1.578$, SE = 0.64, $t = 2.480$, $p = 0.0132$), both in the first contrast (native vs. non-native speakers) on the same polynomial term. Group differences appearing in the native vs. non-native contrast mirror the different pattern of behaviour observed in the first set of contrasts which was different for the L2 learners relative to the other two groups.

Looks to Correct Destination

Contrast 1: ML vs. BL; then NS BL vs. NNS BL

There were no main effects on the intercept term, only a group by context interaction in the second contrast (native bilinguals vs. L2 learners; ($\beta = 0.046$, SE = 0.02, $t = 2.057$, p =

0.040). The linear and quadratic terms were significant (β = 8.794, SE = 0.8, t = 11.141, p < 0.001; β = -9.199, SE = 0.8, t = -11.761, p < 0.001 respectively). There were no main effects on the linear term but three interactions: 1) a group by ambiguity interaction (β = 2.068, SE = 1.0, t = 2.062, p = 0.039), 2) a group by context interaction (β = -2.599, SE = 1.0, t = -2.590, p = 0.009), both in the second contrast and 3) a context by ambiguity interaction overall (β = 2.714, SE = 0.79, t = 3.444, p = 0.001). The group by context interaction on the intercept term reflects the fact that the presence of the second referent resulted in fewer looks to the correct destination in the bilinguals but had no impact on the L2 learners.

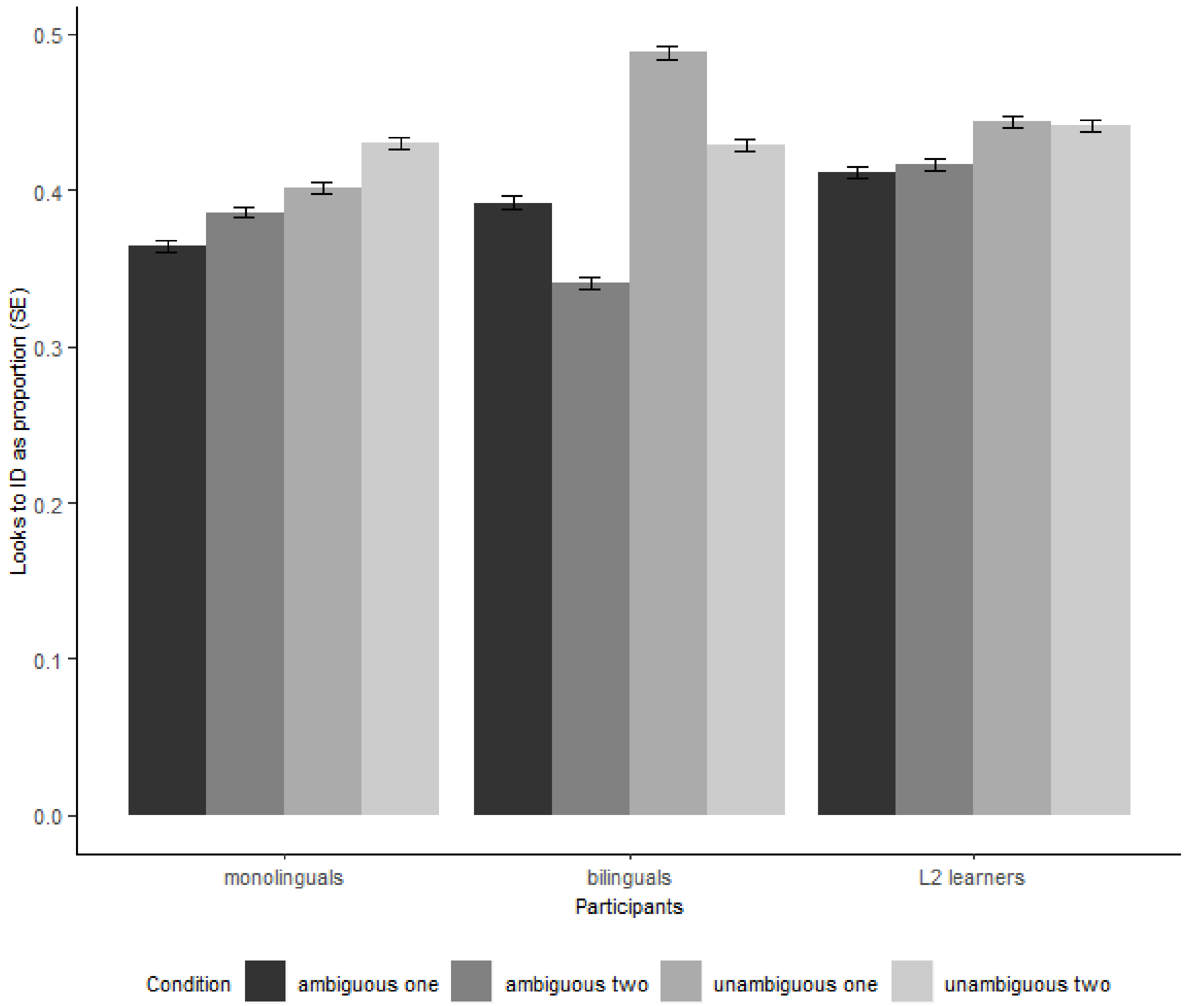

Figure 43 Looks to the correct destination (the bag) in the CD region (while the participants heard “in the bag”) by group and condition overall

The group by ambiguity and group by referential context interactions in the native bilinguals – L2 learners contrast on the intercept reflect different changes over time for the two groups. For the native bilinguals, looks to the correct destination remain, counterintuitively, consistently higher for ambiguous sentences and the 1-referent context conditions for the first 1,000ms. In the L2 learners, looks to the correct destination increase faster for the 2-referent ambiguous condition and the slowest for the 1-referent ambiguous condition, although after about 700ms the patterns become more changeable. Figure 12 provides an overview of the change in looks to the correct destination over time by condition for each group separately.

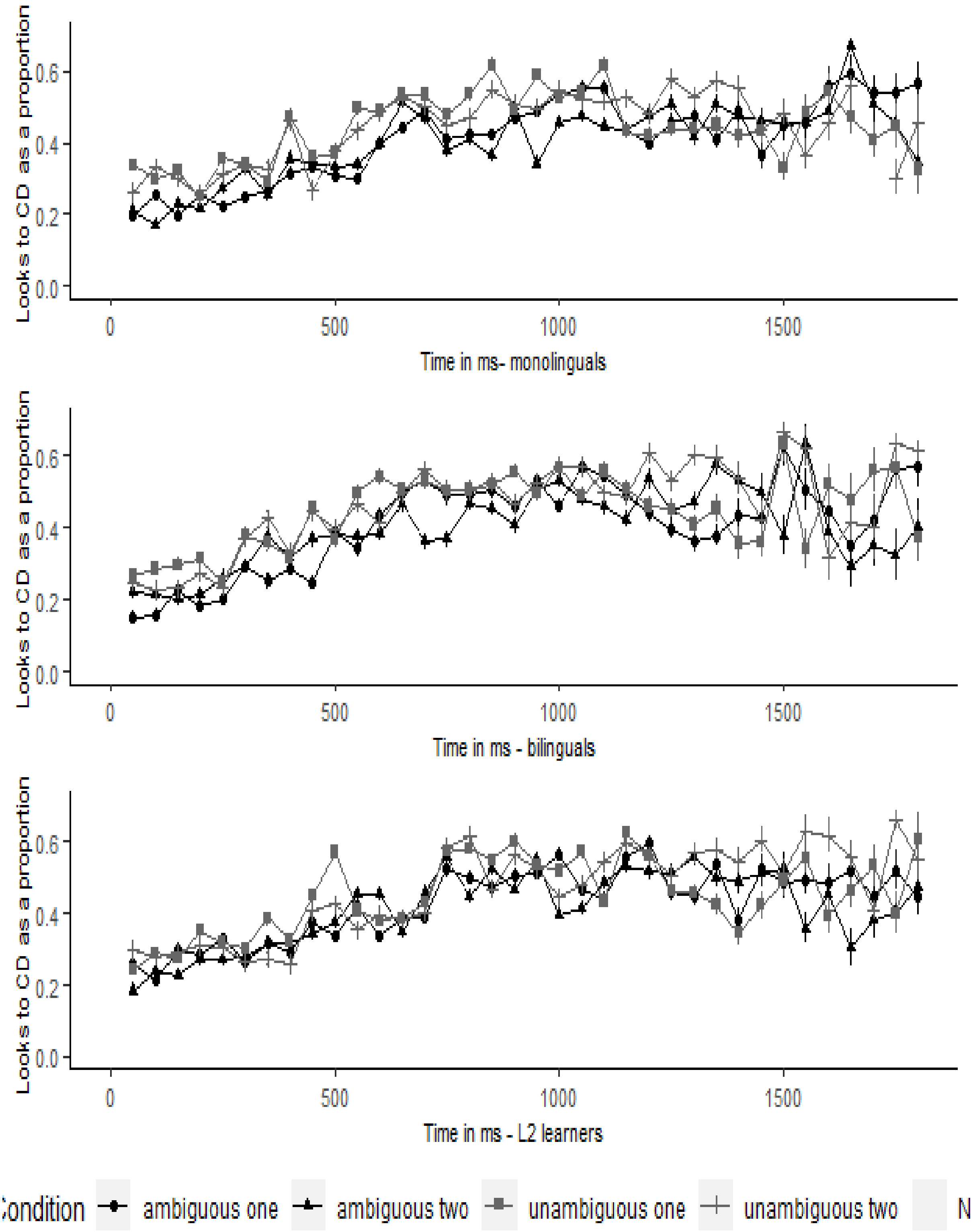

Figure 44 Looks to the correct destination (the bag) in the CD region (while the participants heard "in the bag") by group and condition over time

Contrast 2: NS vs. NNS; then NS ML vs. NS BL

In the second set of contrasts, there were no main effects or interactions on the intercept term. Only an effect of group in the first contrast (natives vs. non-natives) was marginally significant (β = 0.077, SE = 0.04, t = 1.864, p = 0.065). As with the first set of contrasts, the linear and quadratic terms were both significant (β = 8.794, SE = 0.79, t = 11.141, $p < 0.001$ and β = -9.199, SE = 0.78, t = -11.761, $p < 0.001$). On the linear term, there was a significant ambiguity by context interaction (β = 2.714, SE = 0.79, t = 3.444, $p < 0.001$) which was also found for the first set of contrasts, a group by context in the first contrast (native vs. non-native, previously second contrast, (β = -2.407, SE = 1.1, t = -2.211, p = 0.027) while the group by ambiguity interaction in the first contrast was marginally significant (β = 1.976, SE = 1.1, t = 1.817, p = 0.069). These effects point to the different patterns shown for the L2 learners relative to the other groups in the previous set of contrasts. The ambiguity by context interaction on the linear term across all groups is due to the initially fewer looks for the 1-referent ambiguous condition and the initially slower increase relative to the 2-referent ambiguous condition, but the reverse is true for the unambiguous sentences. In other words, looks to the correct destination are lowest and increase the slowest for the hardest condition for the participants. The presence of the second referent also leads to more looks to the correct destination in the later stages of the segment examined (post 1,000ms) for both ambiguous and unambiguous conditions, although the reason for this is unclear.

5.3.3.3. FINAL region ("before going to school")

Looks to Incorrect Destination

Contrast 1: ML vs. BL; then NS BL vs. NNS BL

There was a main effect of group in the second contrast (bilinguals vs. L2 learners; $\beta = -0.033$, SE = 0.02, $t = -2.170$, $p = 0.033$) but not for the first contrast on the intercept term (monolinguals vs. both groups of bilinguals; $\beta = 0.007$, SE = 0.02, $t = 0.384$, $p = 0.702$). There was also a main effect of referential context ($\beta = 0.031$, SE = 0.011, $t = 2.727$, $p = 0.007$) on the intercept term but no main effect of ambiguity ($\beta = -0.017$, SE = 0.01, $t = -0.154$, $p = 0.877$). The only interaction on the intercept term was group by context in the first contrast (monolinguals vs. bilinguals; $\beta = 0.035$, SE = 0.012, $t = 2.270$, $p = 0.023$). The linear and quadratic terms were significant ($\beta = 1.191$, SE = 0.47, $t = 2.525$, $p = 0.012$ and $\beta = -1.357$, SE = 0.47, $t = -2.894$, $p = 0.004$ respectively). On the linear term, there were no main effects and only a three-way group by context by ambiguity interaction in the first contrast ($\beta = 1.339$, SE = 0.64, $t = 2.092$, $p = 0.037$).

There were more looks to the incorrect destination for the bilinguals relative to the L2 learners. The main effect of referential context is due to the fact there were more looks to the incorrect destination in the 2-referent context. To further explore the group by context interaction, the data were analysed separately for monolinguals and bilinguals/L2 learners. There was an effect of referential context for the two groups of bilinguals with more looks towards the incorrect destination in the 2-referent context ($W = 4224500$, $p < 0.001$) but this was completely absent in the monolingual data ($W = 1741200$, $p = 0.499$). The gaze data are presented in Figures 45 and 46.

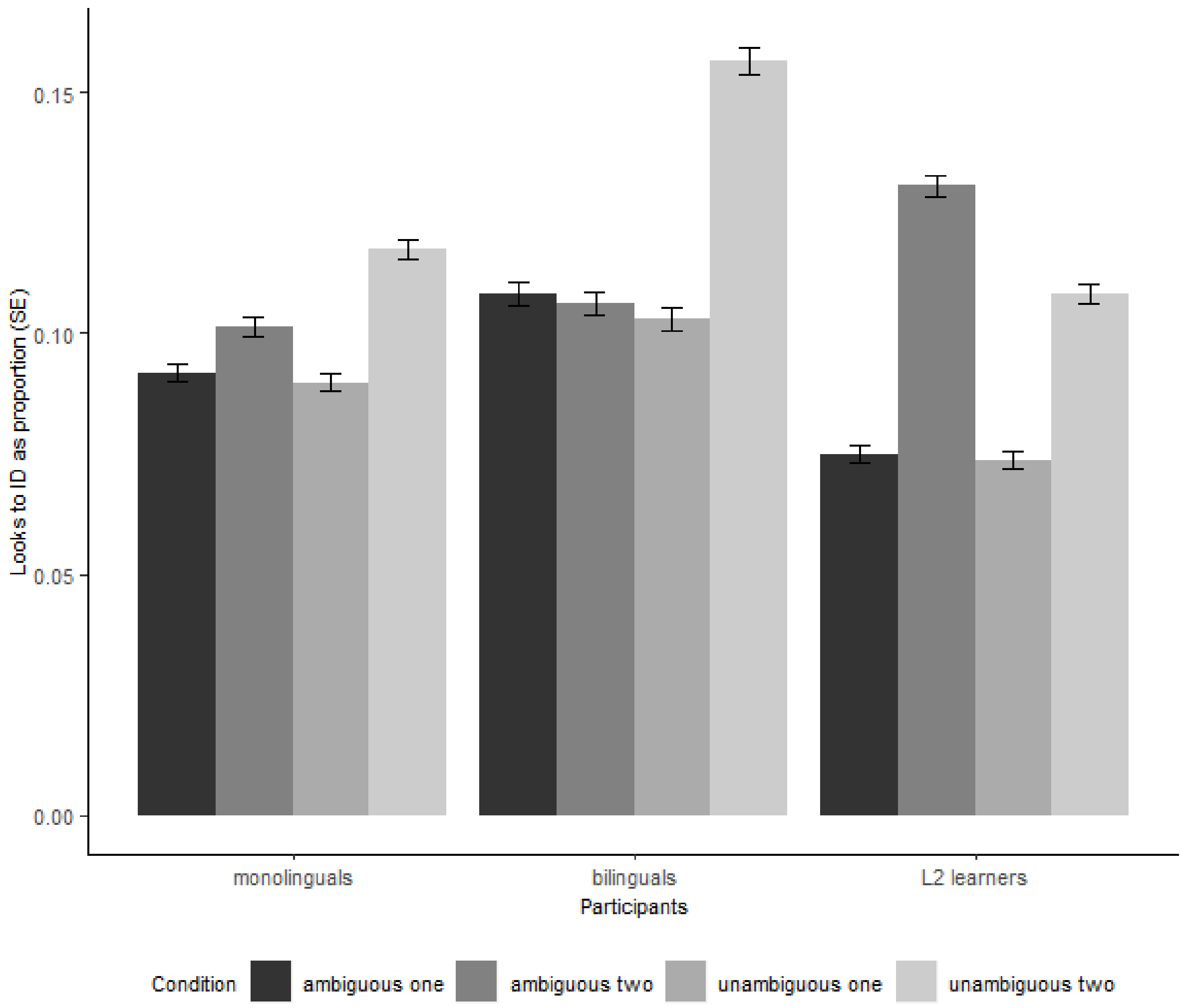

Figure 45 Looks to the incorrect destination (the plate) in the FINAL region (while the participants heard before going to school") by group and condition overall

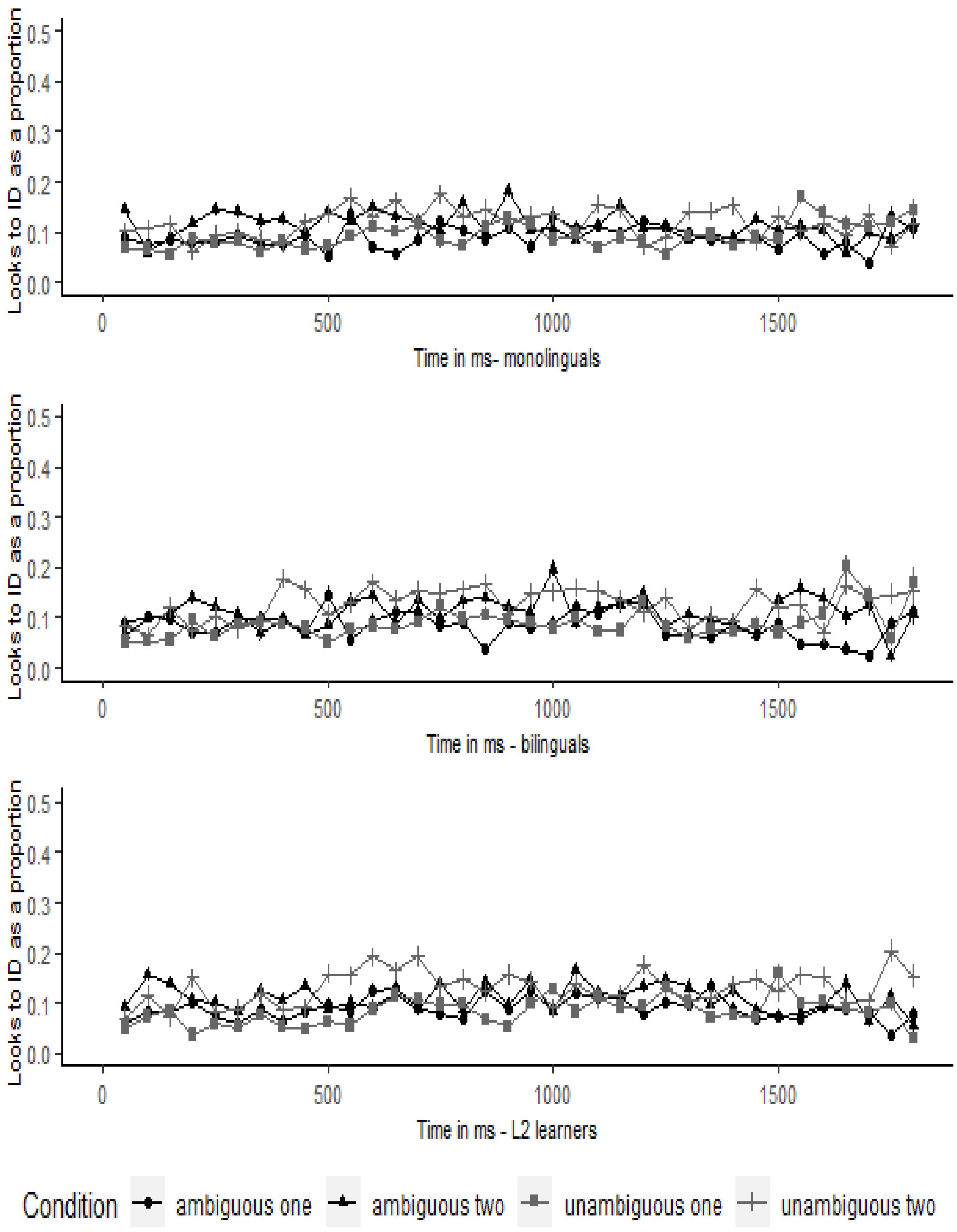

Figure 46 Looks to the incorrect destination (the plate) in the FINAL region (while the participants heard "before going to school") by group and condition over time

Contrast 2: NS vs. NNS; then NS ML vs. NS BL

The main effect of context on the intercept term was found in the second set of contrasts as in the first (β = 0.031, SE = 1.13, t = 2.753, p = 0.006). The linear and quadratic terms were again significant (β = 1.203, SE = 0. 47, t = 2.553, p = 0.011 and β = -1.345, SE = 0.47, t = -2.870, p = 0.004). The only significant interaction on the intercept term was one of group by context in the contrast between native speakers and L2 learners (β = 0.061, SE = 0.03, t = 1.966, p = 0.04949). There were no main effects on the linear term, as was the case with the first comparisons, and only the three-way interaction of group by context by ambiguity in the monolinguals vs. bilinguals contrast reached significance (β = 2.662, SE = 1.28, t = 2.081, p = 0.038). To address the interaction on the intercept term, the data were analysed individually for the native speakers – monolingual and bilingual - and the L2 learners. The effect of context was found for both native speakers (W = 4788000, p = 0.029) and also for the L2 learners (W = 1408200, p < 0.001). The effect was larger for the L2 learners than for the two native speaker groups. Taken together with the results from the first set of comparisons, the data suggest that the patterns of the bilinguals differ from those of both the monolinguals and the L2 learners. The monolinguals show a reduced effect of referential context, the L2 learners show a consistently enhanced effect and the bilinguals show a large effect in the unambiguous condition.

Looks to Correct Destination

Contrast 1: ML vs. BL; then NS BL vs. NNS BL

There was only a main effect of ambiguity on the intercept (β = 0.053, SE = 0.02, t = 3.009, p = 0.003) but not on the linear term (β = -0.76, SE = 0.70, t = -1.087, p = 0.277031). There were more looks to the correct destination when the sentence was unambiguous than when it was ambiguous regardless of group or referential context (see Figure 13). The effect of context was significant on the linear term (β = -2.312, SE =0.70, t = 3.296, $p < 0.001$) but not on the intercept term (β = 0.022, SE = 0.02, t = 1.050, p = 0.296). No other effects were significant on either term. The linear and quadratic terms were significant (β = -2.231, SE = 0. 95, t = -2.355, p = 0.019). There were no interactions on the intercept term and only a group by context interaction on the linear term (β = -2.312, SE = 0.70, t = -3.296, $p < 0.001$) in the monolingual vs. bilingual contrast and a group by context by ambiguity interaction in the natives vs. non-natives contrast (β = -1.940, SE = 0.89, t = -2.176 0.029). The effect of context on the linear term is difficult to interpret. It appears that there are initially more looks to the correct destination in the 2-referent conditions and this is more visible in the L2 learners. However, the absence of any group by context interaction does support claims about differences between groups.

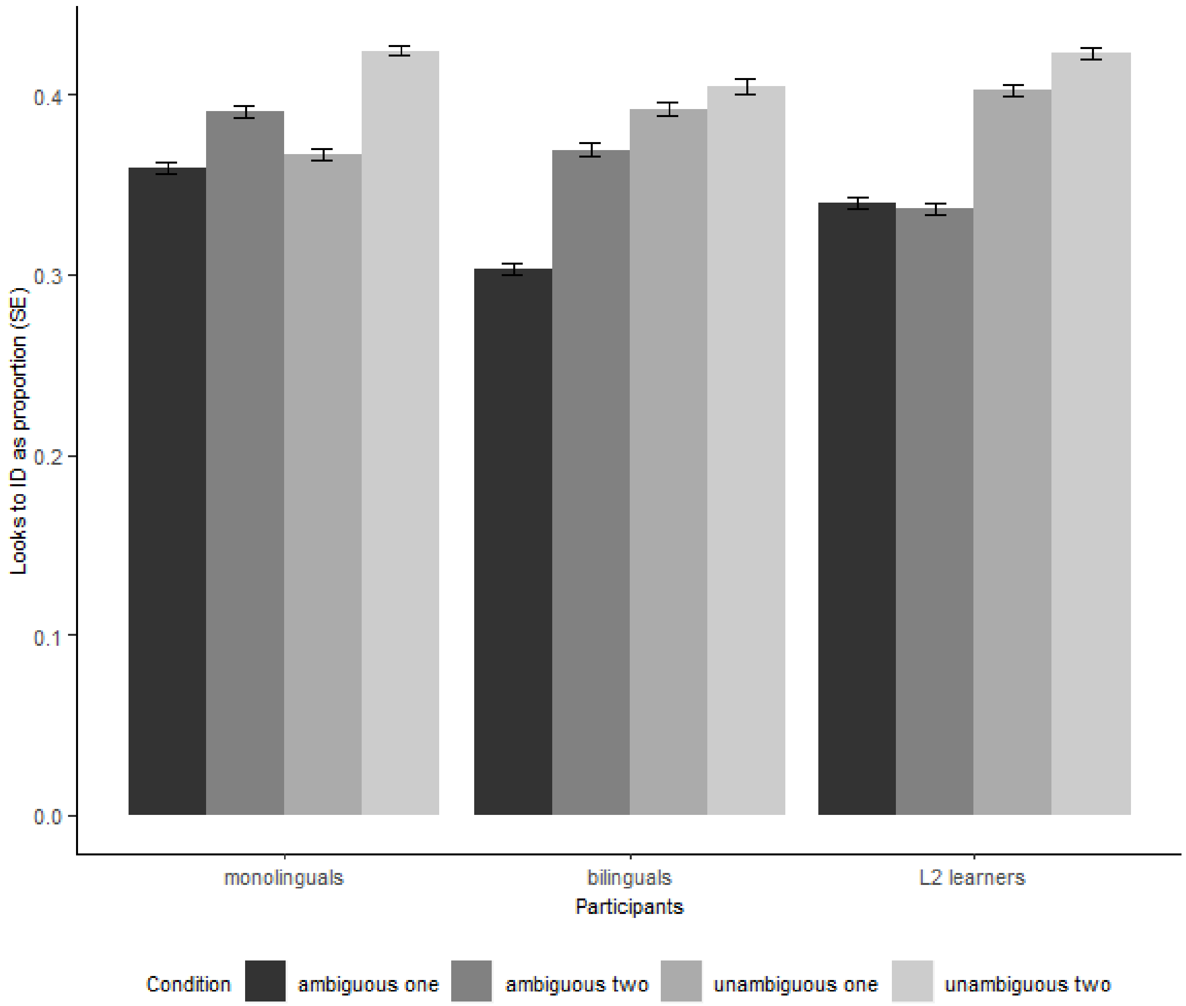

Figure 47 Looks to the correct destination (the bag) in the FINAL region (while the participants heard before going to school") by group and condition overall

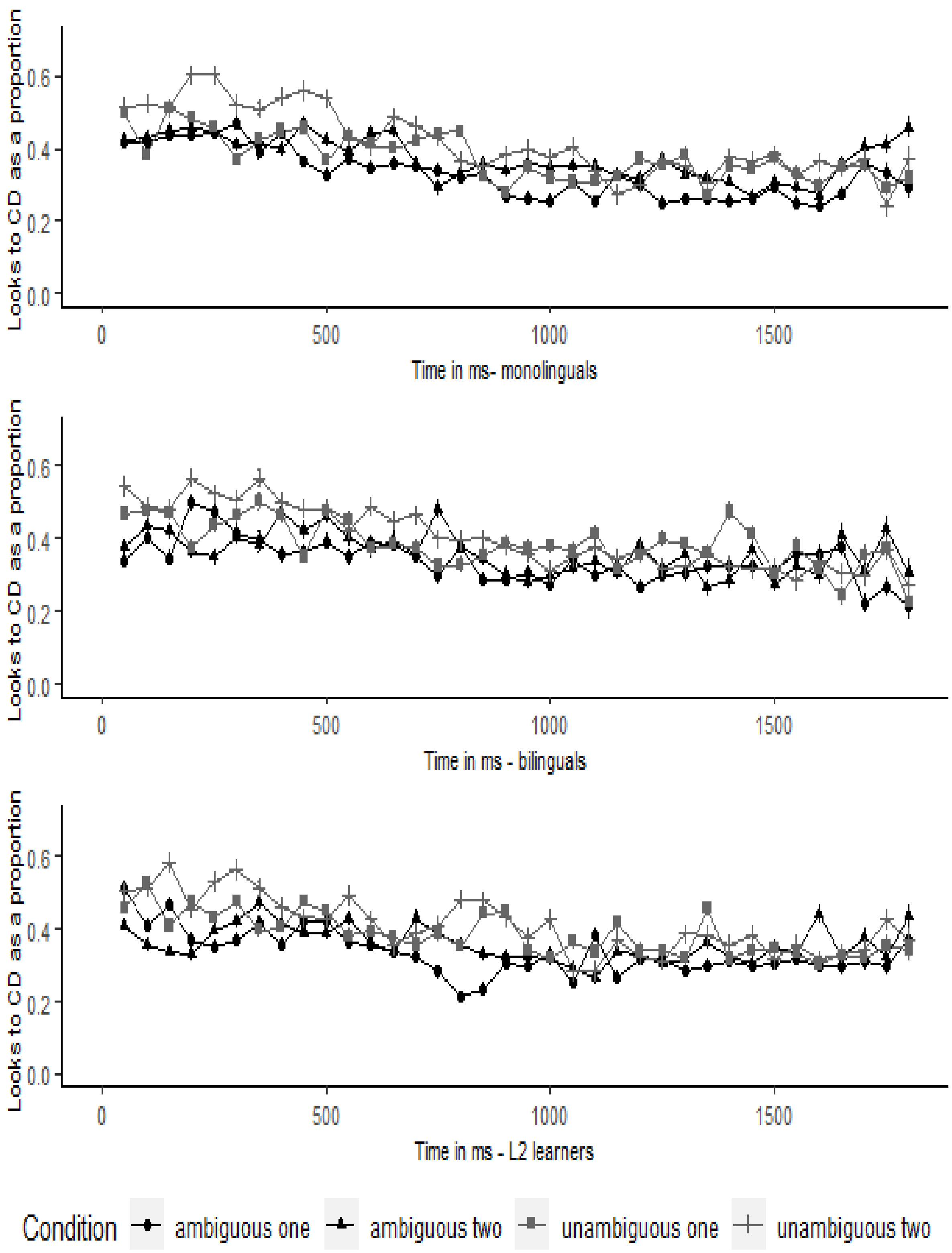

Figure 48 Looks to the correct destination (the bag) in the FINAL region (while the participants heard "before going to school") by group and condition over time

Contrast 2: NS vs. NNS; then NS ML vs. NS BL

The analyses for the second set of contrasts (native vs. non-native speakers first; then monolingual vs. bilingual native speakers) mirrored the effects found in the first contrasts. There was a main effect of ambiguity on the intercept ($\beta = 0.055$, SE = 0.02, $t = 2.604$, $p = 0.022$) but not on the linear term ($\beta = -0.712$, SE = 0.70, $t = -1.013$, $p = 0.311$). The effect of context was significant on the linear term ($\beta = -2.297$, SE = 0.70, $t = -3.268$, $p = 0.001$) but not on the intercept term ($\beta = 0.022$, SE = 0.02, $t = 1.085$, $p = 0.281$). The linear term was significant suggesting an increase in looks to the correct destination over time in a linear manner ($\beta = -9.219$, SE = 0.70, $t = -13.107$, $p < 0.001$). There were no significant interactions on the intercept term. Two interactions on the linear term were significant; a context by group interaction in both contrasts ($\beta = 5.002$, SE = 1.90, $t = 2.620$, $p = 0.009$ and $\beta = -3.811$, SE = 1.766, $t = -2.158$, $p = 0.0309$).

5.4. Discussion

To summarise, there were no differences between groups or referential context but only main effects of ambiguity (lower for ambiguous sentences) for accuracy; these differences were, nonetheless, consistent across groups. For reaction times, participants were overall slower for ambiguous sentences than for unambiguous sentences, which is intuitive, but unexpectedly also slower overall in the 2-referent condition than in the 1-referent condition; this, however, was driven predominantly by the bilinguals. The bilinguals were generally slower than monolinguals, in particular for ambiguous sentences but showed similar processing patterns in a number of segments. In the ambiguous region, looks to the incorrect destination were more for the ambiguous sentences than the unambiguous ones. In the subsequent regions (disambiguation and end of sentence); looks to the correct destination are lower for the ambiguous sentences. These differences become significant mostly in the way the looks to the correct destination over time change. The effect of referential context was inconsistent. When observed, it was either in the opposite direction from what would be predicted (more looks to the incorrect destination in the 2-referent than in the 1-referent condition) or was not limited to ambiguous sentences. This, however, was found across all groups and group by context interactions were mostly non-significant. This section addresses each question individually and discusses the findings in the light of previous works in relation to these.

RQ1: Do native bilingual adults with early naturalistic exposure to English process structurally ambiguous sentences similarly to monolingual adults? Do they differ to L2 learners?

Off-line measures are not directly informative about processing but allow for inferences to be drawn. For the study in question, participants heard an ambiguous prepositional phrase (e.g. “on the plate”) which may be interpreted as a modifier to the verb phrase (“put the apple”) or as a modifier to the second noun phrase (“the apple”); while both options are plausible, participants are more likely to build the first interpretation. It is only after the second prepositional phrase (“in the bag”) that the second interpretation is the only grammatical interpretation, albeit the originally less preferred one. The fact that responses to comprehension questions about ambiguous sentences have lower accuracy scores and longer reaction times when answered correctly is consistent with attributing the source of the difficulty in garden-path sentences to the need to reanalyse upon discarding the initial misinterpretation. The absence of effects of group or interactions of group with ambiguity in either set of contrasts entails that all groups process sentences in a qualitatively similar, i.e. incremental, manner. Trends towards group differences or the group by ambiguity interaction signal a more pronounced slowdown for both bilingual groups relative to the monolinguals in the reaction times. This may be suggestive of greater difficulty recovering from garden-path effects but does not suggest qualitative differences between groups in terms of processing. An overall different picture emerges from Pozzan & Trueswell (2016) for both the monolinguals and the L2 learners[32]. For both groups, accuracy was lower for ambiguous sentences than for unambiguous. However, in Pozzan & Trueswell, the monolinguals did not make any errors in the unambiguous sentences. This was not the case with this study where the monolinguals did make errors in the comprehension questions. Furthermore, this study failed to find differences in accuracy between monolingual and both groups of bilinguals. In contrast, Pozzan & Trueswell (2016) showed lower act out accuracy for the L2 learners relative to the monolinguals. This is probably due to the lower proficiency of the L2 learners in the Trueswell et al. study relative to the ones in this study. One should note that there are differences between the two studies in terms of the comprehension measures; this study used comprehension question accuracy whereas Trueswell et al. used act out. It is unclear how this impacts the differences in results, but it would be reasonable to expect that, in the context of the visual world paradigm, the act out would be easier than the comprehension questions. The reason for this is that the act out occurs immediately after the instruction whereas the response needs to be to a comprehension question following the ambiguous sentence. Moreover questions may be harder than declaratives and imperatives as a structure as they involve syntactic movement (of the wh-phrase).

More direct evidence for incremental processing comes from the gaze data and, in particular, the looks to the incorrect destination when the participants hear the first/ambiguous prepositional phrase. When participants heard the phrase “on the plate”, there were more looks towards the picture of the plate – upon which there was no apple - in the ambiguous sentences than in the unambiguous ones. This indicates that the parser interpreted the

[32] Native bilinguals were not included in Pozzan & Trueswell

ambiguous PP as the destination of the verb “put”. Moreover, as the linear term was significant, looks to the incorrect destination increased over time indicating that the preference towards the wrong interpretation strengthened. This may be due to the fact that the initial interpretation was not formed immediately after hearing the ambiguous PP. The main effect of ambiguity on the linear term shows that looks towards the incorrect destination were persistently more for the ambiguous condition and increased significantly more rapidly.

However, the effect of garden-pathing was not consistent across groups as evidenced by the main effect of group and the interaction of group with ambiguity. In particular, the L2 learners and the bilingual natives behaved similarly and looked at the incorrect destination more in the ambiguous sentences than in the unambiguous ones. This was not the case for the monolinguals. These differences were not just confined to the total amount of looks towards the incorrect destination in the relevant region but also differed in how their looks changed over time post onset as evidence by the group by ambiguity interaction on the linear term for all contrasts. Both groups of bilinguals showed an increase in looks to the incorrect destination 500-1,000 ms post onset of the ambiguous PP. For the bilingual natives, however, the increase was limited to the ambiguous sentences in the 1-referent condition only, whereas the same increase was found for all ambiguous sentences for the L2 learners. This increase was more limited and consistent across conditions in the monolinguals. The fact that L2 learners did not make use of referential context in contrast to the native bilinguals is inconsistent with the Shallow Structure Hypothesis which predicts a greater reliance on non-structural cues, and as such, larger effects of referential context in this case, for the L2 learners. Instead this discrepancy could be more readily viewed as an inability to utilise various types of information in real time, and as possible, relative to those bilinguals who acquired English at a younger age and with more naturalistic exposure.

The results from this segment are partly in line with Trueswell et al where there is a similar increase in looks towards the incorrect destination which resulted in a main effect of ambiguity equivalent to the effect of ambiguity on the intercept term in this study. The findings from this study differ to Trueswell et al. in two ways; there is no ambiguity by context interaction and it is not observed for the monolinguals. In Trueswell et al., the subsequent pairwise analyses revealed the effect of ambiguity was present only in the 1-referent condition. This study does not undertake pairwise comparisons as there was no interaction of ambiguity by context. Moreover, a group by ambiguity by context interaction on the linear term suggests that the increase in looks to the incorrect destination for 1-referent ambiguous sentences only is limited to the bilingual natives. Two factors may have contributed to these discrepancies in the results; for this study, the data were analysed using mixed effects linear regression, which is more conservative, as standard practice in current psycholinguistic research. This would be different to Trueswell et al. (1999) but Pozzan & Trueswell (2016) also analysed the data this way. However, the latter study tested participants with lower proficiency and exclusively classroom exposure. In addition to this, the use of orthogonalised polynomial regression means that the contribution of each independent variable is examined on each term separately. This however, would not convincingly account for the differences between groups. Another difference between two studies is the speed of presentation; in Trueswell et al., all sentences were reportedly read out in under 3,300ms and gaze data beyond that time point were not analysed. For this study, the sentences were read out substantially slower and segmentally (minimum time was 7,800). This was done so that the task was identical with the one the children undertook and allows for longer windows of gaze data to be analysed.

The results for the gaze data are similarly consistent with Pozzan & Trueswell in that there are more looks to the incorrect destination in ambiguous sentences. The difference between

Pozzan & Trueswell and this study is that in the former there is no interaction between group and ambiguity suggesting similar garden-path effects between natives and non-natives. In this study, differences are found. However, Pozzan & Trueswell analysed gaze data from the onset of the ambiguous segment as a whole whereas this study examines it in segments and also looks at changes over time. The results are also consistent with other studies on garden-path sentences such as Hopp (2015), who showed greater difficulty with ambiguous sentences in a number of metrics in an eye-tracking study on subject-object locally ambiguous sentences, such as first and second pass-reading times and regression likelihood.

RQ2: Do native bilingual adults recover from garden-path effects similarly i.e. equally successfully and over the same time course as monolingual adults?

As with Trueswell et al. (1999), accuracy was lower for ambiguous sentences than unambiguous sentences. The lower accuracy scores for the ambiguous sentences in comparison to the unambiguous ones suggests that bilinguals have some difficulty with garden-path sentences that persists even after they have heard the entire sentences. This should be interpreted as a potential difficulty in revising initial misinterpretations rather than an inability to comprehend these types of sentences as only about 1 in 6 comprehension questions were answered incorrectly for the ambiguous sentences. This was consistent across-groups as evidenced by the absence of an effect of group or any group interaction on the accuracy in any between-group comparison as well as a visual inspection of the raw data which shows very small numeric differences between all three groups. This suggests that the bilinguals and the L2 learners were equally as (un)successful as the monolinguals in recovering from the garden-path effects. This is in contrast to Pozzan & Trueswell (2016) who found that the L2 learners were less able to recover from misinterpretation. One possible reason for this discrepancy is that the L2 learners in the aforementioned study had lower proficiency than the L2 learners in this study.

The persistent difficulty is also reflected in the differences in reaction times; participants were slower to select the correct picture to answer the comprehension question after the end of the sentence. However, the reaction times also point to group differences not captured by accuracy. The bilingual natives had the longest reaction times and there was a trend for the two groups of bilinguals and the bilingual natives to be slower than the monolinguals. Moreover, group by ambiguity interactions observed suggest a more pronounced slowdown for ambiguous sentences relative to the unambiguous ones in both groups of bilinguals relative to the monolinguals. In sum, all three groups were equally successful in recovering from garden-path effects in ambiguous sentences after hearing the complete sentence but, when doing this successfully, the bilingual natives are actually slower than the monolinguals in doing so and speed-wise similar to the L2 learners.

Evidence for recovery from garden-path effects can also be obtained from the online measures. In the disambiguating region, looks to the incorrect destination decline over time significantly as evidenced by the negative co-efficient of the linear term. Conversely, looks to the correct destination in the correct destination region where there is disambiguation increase rapidly from about chance (ca 25%) to over half; the linear term is statistically significant with a positive co-efficient. These two results suggest that the participants are moving away from the erroneous original interpretation towards the correct one. In both cases, there is a group by ambiguity interaction in the native bilinguals vs. L2 learners contrast on the linear term and marginally on the natives vs. non-natives contrasts. Visual inspection of the data suggested a more prolonged decrease in looks to the incorrect destination and increase in looks to the correct destination for the L2 learners relative to the native bilinguals and, to a lesser extent, of the latter relative to monolinguals for the looks to the incorrect destination.

The fact that, in the region of disambiguation, i.e. the correct destination region ("on the plate"), looks to the incorrect destination are still significantly higher for ambiguous than unambiguous sentences for all groups reflects the fact that the misinterpretation may persist (see Slattery et al., 2013 for residual misinterpretations across sentences). In the segment in question, the difference in looks to the correct destination between ambiguous and unambiguous sentences does not yet result in significant effect overall. This is due to the fact that the differences emerge with time. Even so, in terms of raw numbers, all groups' looks to the correct picture more in the unambiguous sentences than in the ambiguous ones.

An asymmetry persists for the ambiguous sentences even in the final region of the sentence ("before going to school"). There is an effect of ambiguity with fewer looks to the correct destination. One explanation for this is, again, the potential difficulty in revising initial misinterpretations. The absence of any group by ambiguity interaction means that this difficulty is consistent across all groups. Trueswell at el. analysed data in 200ms bins after the onset of the ambiguous PP but do not report on gaze data following disambiguation. For this reason, it is difficult to infer whether or not there are difficulties in syntactic reanalysis of local ambiguity on the basis of the gaze data. The results from Trueswell et al.'s act out which show significantly lower accuracy[33] for ambiguous sentences, in particular for 1-referent contexts, even in adults are consistent with the results from comprehension accuracy in this study suggesting some lingering difficulty with garden-path sentences.

RQ3: Do native bilingual adults with early naturalistic exposure to English utilise number of referents as a disambiguating cue in interpreting garden-path sentences? If they do so,

[33] Albeit, the number of errors was very small overall for adults.

are there differences between the two groups as an end –state result or in the change of gaze data over time?

The effects of context were unclear in this study given the interaction between context and ambiguity and the fact that the effect of context was not in the expected direction. The referential context did not have any impact on the processing of garden-path sentences in the off-line measures for any of the groups regardless of the ambiguity. This was also attested with the adult data from Trueswell et al. (1999) for the act out accuracy[34]. Two considerations could account for this. Firstly, effects may not be easily observable at levels of accuracy at or near ceiling as there is very limited scope for higher accuracy in one condition relative to the other. The L2 learners in this study are closer to ceiling than those in Pozzan & Trueswell. In terms of reaction times, referential context had the opposite effect of what would be expected; participants were slower to respond by selecting a picture when there were two referents in the visual stimuli than when there was only one. This was found to be driven by the ambiguous sentences and for the two groups of bilinguals. In this sense, the native bilinguals performed similarly to the L2 learners and differently to the monolinguals. This is the first study to report reaction times; previous studies reported accuracy and gaze data. One plausible explanation for the effect in the reaction times is that the second referent acted as a distractor hence slowing the participants down.

In terms of gaze data, the region of predominant interest is the incorrect destination region, when participants heard the ambiguous PP and the region immediately thereafter when there is disambiguation. At the first point, looks to the incorrect destination would signify that the participants have interpreted the ambiguous PP to modify the verb per se rather than its object, in other words they are a measure of garden-pathing. It was expected that looks to the

[34] In the Trueswell study participants needed to move an object to one location; in this study, they answered a comprehension question. This may account for the discrepancy in the accuracy rates.

incorrect destination would be reduced in the 2-referent condition as the presence of the second referent in the visual stimuli would act as a cue for an NP- rather than a VP-modifying reading. In the incorrect destination region, i.e. once the participants heard the ambiguous PP "on the plate", there was an effect of referential context as there were overall fewer looks to the incorrect destination in the 2-referent condition than in the 1-referent condition suggesting that the participants utilised the referential context to aid processing. In other words, they saw a second apple which was on a plate and were guided to interpret the ambiguous prepositional phrase to modify the NP rather than the VP. But there are two reasons this was not consistent with Trueswell et al. (1999) and Pozzan & Trueswell (2016). Firstly, this occurred in both ambiguous and unambiguous conditions. Moreover, there were differences across groups as evidenced by the interactions of context with group in the native vs. non-native bilingual contrast. More specifically, the native bilinguals are the only group who showed the effects in the expected direction; they looked less at the incorrect destination in the 2-referent than the 1-referent condition when the sentence was ambiguous but there was no such effect in the unambiguous conditions. In contrast, the L2 learners showed the effect in the opposite direction while no differences were found for the monolinguals. This may reflect difficulty utilising the referential cue in real time which comes to function as a distractor. Looks to the incorrect destination are generally lower for the monolinguals than for the two groups of bilinguals. The effect in this direction persisted in the correct destination region, i.e. when the participants heard the second prepositional phrase "in the bag" which disambiguates the sentence in favour of the less likely interpretation but also in the final region of the sentence ("before going to school").

Weak evidence for the effect of context can also be found for the looks to the correct destination in the region where there is still ambiguity. While there were no overall main effects or interactions, looks to the correct picture began to increase in the incorrect destination region about 800ms post onset for all conditions with the exception of the ambiguous 1-referent condition. It is unexpected that this would occur at this particular region as the sentence is still ambiguous in terms of PP attachment. One plausible explanation could be prediction; the participants know, given the argument structure of the verb "put" that a location of movement will be needed. Therefore, they look at the only picture which could depict the destination. In the unambiguous sentences, the incorrect destination is never a plausible destination whereas in the ambiguous ones, it may plausibly be a destination but the presence of the second referent guides the parser to interpret the ambiguous PP as modifying the object of the verb instead of the verb.

Looks to the correct destination in the disambiguating, i.e. correct destination, region (when participants heard "on the plate") also show effects of referential context which varies by group. There were no overall differences and only the monolinguals showed more looks to unambiguous than ambiguous and more looks to correct destination for 2-referent context than 1-referent context but across both ambiguous and unambiguous sentences. The interaction of context by ambiguity on the linear term across groups suggests that the increase in looks over time varies by condition in conjunction to other variables. This reflects more looks overall across groups to the correct destination for the 2-referent condition in the ambiguous sentences but not in the unambiguous ones and a more protracted increase in the first 800ms of the segment. This means the presence of the second referent in the ambiguous sentences provided an early cue for disambiguation when needed but did not have this effect when it was not. The effect of the second referent is also present in the later part of the segment, i.e. after 1,000ms but for both ambiguous and unambiguous sentences. The reasons for this are less clear.

There were, however, differences in change over time in looks for the two groups of bilinguals as evidenced by the group by context interaction on the linear term for the respective contrast. For the L2 learners, there was a steeper increase in looks to the correct destination in the 2-referent ambiguous condition in comparison to the respective 1-referent condition indicating the L2 learners built the correct interpretation faster in the presence of the disambiguating cue. This was not found for the native bilinguals. Given the absence of a group by context interaction on the linear term for the relevant contrast, we assume that the changes in time for the native bilinguals and monolinguals are statistically the same. This means only the L2 learners seem to use the visual cues to disambiguate in the condition needed at the point of disambiguation. The latter observation may be predicted under the Shallow Structure Hypothesis; the non-native speakers are relying on non-structural information more so than the native speakers when a new interpretation is needed. This is inconsistent with Trueswell at al. who found that monolingual adults used referential context to disambiguate. This could be attributed to differences in experimental design; the sentences in this study were read out substantially slower than in Trueswell, this may have enabled native speakers to avoid garden-path effects. This being the case, processing in the non-native speakers would be assumed to be slower than that of the native speakers.

The results from the gaze data differ from those in other studies to some degree in terms of the effect of referential context. In Trueswell et al. (1999), there was an increase in looks to the incorrect destination for the ambiguous condition only in the 1-referent condition but not in the 2-referent condition. This means garden-pathing was avoided due to the presence of the second referent in the visual stimuli. In contrast to Trueswell et al. (1999), the study in this thesis finds a more inconsistent pattern. One reason for this may be the difference in the measure used to capture gaze data. Trueswell et al. calculated proportion of trials where there was a fixation to the incorrect destination as well as the total time fixating on the incorrect destination 200-1,000ms after the onset of the ambiguous segment. These were also analysed in 100ms time windows. This study relies on the proportion of looks towards the incorrect destination relative to the all pictures for 200ms segments in each critical time window in each trial individually, similarly to the measure adopted for the gaze data in the studies on filler-gap dependencies. This measure allows for a larger number of observations and was used with time as a variable in order to capture changes over time. That way, it avoids a limitation of using fixations which may be the by-product of accumulation of looks to a given picture, which will happen inevitably, over time and was acknowledge by Trueswell et al (1999:111). This study did not show such effects overall based on effects of the intercept term which is the closest measure to what was analysed in Trueswell et al. (1999). A differential impact of context depending on ambiguity on the linear term was at found at various segments and but does not support the notion that the second referent provided an effective disambiguating cue in the ambiguous condition more so than in the unambiguous condition (e.g. looks to correct destination in incorrect and correct destination regions). This is because the effect is sporadically found and was often not in the predicted direction.

The results are also somewhat different to Pozzan & Trueswell in terms of impact of referential content; looks to the incorrect destination were fewer for the 2-referent conditions predominantly in the ambiguous sentences but also in the unambiguous sentences. Crucially, Pozzan & Trueswell found no three-way interaction between group, ambiguity and referential context indicating that the benefit of the second referent for disambiguating was not particular to the only one of the two groups. However, the results in the present study are more inconsistent with regards to how context interacts with ambiguity making comparisons challenging.

5.5. Conclusion

This study tested how monolinguals, bilinguals with early naturalistic exposure and L2 learners process sentences with local syntactic ambiguity where the wrong interpretation is likely to be initially preferred, i.e. garden-path sentences. More specifically, the study examined whether all three groups could utilise visual cues to guide the parser towards the unlikely but ultimately correct interpretation. Results were reported for accuracy, reaction times as well as gaze data for two metrics across four segments of the sentences. The results showed overall greater difficulty with ambiguous sentences relative to unambiguous ones, i.e. ones which had an overt relative clause. This was found across the majority of metrics and was consistent across groups. All groups appear to initially misinterpret garden-path sentences and show some degree of difficulty in reanalysing the ambiguous sentences relative to unambiguous ones. The native bilinguals performed similarly to the L2 learners in terms of reaction times rather than the monolinguals despite equal accuracy scores. Referential context did not seem to have the expected impact given previous works; the presence of the second referent lead to more looks towards the correct destination in the ambiguous sentences but also for unambiguous sentences. Paradoxically, the same asymmetry was also found for looks to the incorrect destination. The latter does not suggest that referential context has a disambiguating effect. The precise impact of referential context appears unclear and inconsistent. Differences between groups are inconsistent and difficult to establish given the data.

Chapter 6 Discussion

6.1. Summary of findings

This thesis investigated processing of morphosyntax in bilingual children and two different types of bilingual adults based on age of onset and quality of input and exposure. To this end, four visual-world paradigm studies with eye-tracking were conducted; two with bilingual children and two with bilingual adults (early or simultaneous with naturalistic exposure and late sequential with classroom exposure mainly) and their monolingual controls. Two structures were tested; wh-questions and garden-path sentences.

Study 1A tested processing of wh-questions in children in a follow up to Contemori et al. (2018). The results showed that the bilingual children were equally as accurate in responding to comprehension which-questions as monolingual children and equally as fast at doing so. Both groups had greater difficulty with object questions than subject questions. The mismatch in number of the two NPs increased accuracy in object questions but not subject questions and decreased reaction times. The on-line processing data showed qualitative similarities between the two groups of children but differences in timing and cue utilisation. Both groups of children looked at the correct picture as much and did so more for the subject questions than the object questions confirming the difficulty in processing the latter. For the object questions, looks to the correct picture were initially below 50% suggesting the participants had misinterpreted them as subject questions. This never happened for the subject questions. Moreover, the increase in looks to the target picture was slower than for the subject questions which were consistently higher. The difference between the two groups was that the monolingual children increased their looks to the correct destination more quickly than the bilingual children indicating faster processing. Moreover, only the monolinguals showed an effect of number mismatch in the gaze data; exclusively for the object questions, there were more looks to the correct picture in the number mismatch condition relative to the number match condition and these also increased faster. This was not found for the bilingual children suggesting they could not use the cue in real time to disambiguate in favour of the ultimately correct reading.

Study 1B tested the same structure in monolingual and bilingual adults and yielded broadly similar results but differences between the groups. Subject questions were easier to process than object questions; accuracy was higher, reaction times were lower, looks to the correct picture were overall, but also initially, higher and increased at a faster rate. Object questions showed the same pattern of looks to the correct picture being below 50% after hearing the first NP, consistent with a potential misinterpretation of them as subject questions. The effect of number mismatch was found in the accuracy data and resulted in higher accuracy in the object questions only. Pairwise comparisons on the basis of a group by number match interaction showed that number mismatch was used by the monolinguals only. The effect of number match was also found in the gaze data where it resulted in a more rapid increase in looks to the correct picture. Nevertheless, this was not confined to object questions alone or only to the monolingual adults.

The three groups of adults showed both similar patterns of processing but also some differences and the native bilinguals varied in their patterning from both the monolinguals and the L2 learners. In terms of accuracy, both monolinguals and native bilinguals outperformed the L2 learners but did not differ from one another. In terms of reaction times, the native bilinguals were slower than the monolinguals but did not differ from the L2 learners. The results from the gaze data were similar; although all groups looked at the correct picture overall equally, the monolinguals increased their looks over time faster than the two groups of bilinguals who did not differ from one another. Taken together, this suggested slower processing in the bilinguals relative to the monolinguals but qualitatively similar in that processing is incremental. Even though the native bilinguals appeared to pattern with the L2 learners in term of processing, despite differences in accuracy, differences between the two groups were still present even in the processing data. More specifically, the impact of the number of the first NP was different in the two groups of bilinguals; processing a sentence where the first NP was in the plural was harder than when it was in the singular for the L2 learners based on the gaze data. This asymmetry was not observed for the native bilinguals.

Study 2A examined how monolingual and bilingual children process garden-path sentences with reduced relative clauses where the number of visual referents was manipulated (1 vs. 2) similarly to Trueswell et al. (1999). Both groups had difficulty with the ambiguous sentences but potentially differed in the timecourse of the garden-path effects. In the ambiguous region, there were differences in the looks towards the incorrect destination, which is consistent with a garden-path misinterpretation of the sentence which was modulated by the ambiguity. The monolingual children showed an increase in looks towards the incorrect destination. This was not found for the bilingual children. The difficulty in processing ambiguous sentences is further evidenced by the gaze data; there were more looks towards the correct destination picture, suggesting a correct interpretation of the locally ambiguous sentence, at the point of disambiguation. Moreover, the absence of a group by ambiguity interaction for the looks towards the correct destination at this point indicates that both monolingual and bilingual children equally had greater difficulty with the ambiguous sentences. Furthermore, in the final segment of the sentence immediately after disambiguation, looks to the correct destination declined more slowly for the ambiguous sentences than the unambiguous ones indicating a more protracted processing of garden-path sentences due to their ambiguity which was consistent for both groups of children. Accuracy in the comprehension questions was lower for the ambiguous sentences than the unambiguous ones which suggests that reanalysis and re-interpretation remains challenging even at the end of the sentence. The fact that, even in the ambiguous sentences, accuracy was over 70% suggests that the lower accuracy is not due to an inability to understand the ambiguous sentences but is more readily associated with a difficulty recovering from the garden-path effect, consistent with the finding in Trueswell et al. (1999). The absence of a group by ambiguity interaction for accuracy suggests the bilingual children did not have greater difficulty reanalysing and reinterpreting the ambiguous sentences. These effects were not observed in the reaction time data indicating that there was no speed-accuracy trade off. Neither the bilinguals nor the monolinguals showed substantial use of referential context in order to disambiguate neither in the accuracy, reaction times nor the gaze data. This is consistent with Trueswell et al. (1999) who found that the difference between children, albeit younger than the ones in the present study, is not in the presence of garden path effects but in the ability to use information from sources other than syntax to guide the parser towards a disambiguation with a particular interpretation. In fact, the only evidence of this effect was in the gaze data in the region immediately after disambiguation. Participants looked to the correct destination more for trials with two referents in the visual stimuli than one referent, but this was not confined to the ambiguous sentences.

Study 2B tested garden-path sentences in monolingual and bilingual adults. The experimental task was the same as in in Study 2A and the participants were the same as in Study 1B. As with the children, all three groups of adults had greater difficulty with the garden path sentences than the ones with the overt relative clause. Evidence was found both for garden-path effects but also for the role of referential context as a disambiguating cue albeit weaker than in previous studies. The findings for similarities or differences between groups are mixed.

Evidence for garden-path effects include overall a) more looks to the incorrect destination for all groups, consistent with a misinterpretation of the ambiguous sentence at the ambiguous segment of the sentence – these increased over time, suggesting the misinterpretation becomes stronger; b) fewer looks to the correct picture, reflecting a correct interpretation of the sentence, for the ambiguous sentences at the point of disambiguation relative to the respective point for the unambiguous for monolinguals and native bilinguals. The bilinguals experienced garden-path effects suggesting they processed the sentences incrementally. In fact, there were more looks to the incorrect destination in the ambiguous segments of the sentences for the two groups of bilinguals relative to the monolinguals but no difference between the two groups of bilinguals.

Accuracy for the comprehension questions was lower for the ambiguous sentences than the ambiguous ones and time needed to respond successfully to the comprehension question was respectively longer. Taken in conjunction, the latter two findings do not indirectly point to garden-path effects without the looks to the incorrect destination, but suggest a difficulty recovering from misinterpretation even after hearing the entire sentence.

This study also found that referential context impacted processing but to a weaker extent and differently than in other studies. There were fewer looks to the incorrect destination in the ambiguous region, but this was not limited to the ambiguous condition as in previous studies.

At the disambiguating segment, there was a slower increase in looks to the correct picture for the ambiguous 1-referent condition relative to the ambiguous 2-referent condition. This discrepancy was not found in the unambiguous condition suggesting that the presence of the second referent facilitated interpretation of the ambiguous sentence. The presence of the second referent also resulted in a slower decline in looks towards the target destination in the final segment of the sentence relative to the 1-referent condition.

The differences between the three groups were unclear. There were no significant differences in comprehension accuracy between the three groups; which was high for both ambiguous and unambiguous sentences. Moreover, there was no interaction of ambiguity and group in any contrast indicating the difficulty with ambiguous sentences is consistent across the three groups. However, the findings are different for the reaction times needed for participants to respond accurately. There was a trend for both groups of bilinguals to be slower than the monolinguals but not for the L2 learners to be slower than the native bilinguals. Crucially, the ambiguity by group interactions in the monolingual vs. bilingual contrasts suggests the bilinguals slow down more for the ambiguous sentences relative to the monolinguals, indicating a potentially greater difficulty reanalysing the sentence.

In addition to this, the monolinguals and bilinguals appear to differ in the timing of the effects founds when the data are analysed for each group individually. For example, pairwise comparisons driven by a group by ambiguity interaction showed that the effect of ambiguity on the looks to the incorrect destination in the ambiguous segment failed to reach significance in the monolinguals but did so for the two groups of bilinguals. On the other hand, the native bilinguals and the L2 learners differed in their looks to the incorrect destination which was modulated by ambiguity and referential context. Looks to the incorrect destination increased more rapidly for the L2 learners in the ambiguous than in the unambiguous conditions regardless of the referential context, whereas the native bilinguals showed an increase in looks to the incorrect destination which was more specific to the ambiguous 1-referent condition, albeit only briefly. This suggests that the bilingual adults utilised referential context to guide disambiguation when needed whereas the L2 learners did not. A further difference is that L2 learners looked at the incorrect destination more in the final segment in the sentence relative to the native bilinguals and the monolinguals, suggesting a persistence of garden-path effects. In the same region, referential context did not impact the monolinguals but lead to fewer looks to the incorrect destination for the two bilingual groups.

The remaining subsections of this chapter address research questions 1-6 of this thesis individually. The research questions on child processing (i.e. 1-3) are addressed in section 6.2 while the research questions pertinent to adult processing (i.e. RQs 4-6) are included in section 6.3. Subsequently, the findings from all studies are discussed in light of theoretical accounts which they may inform. The chapter then proceeds to a discussion of the limitations and possibilities for future research. Lastly, the chapter outlines its concluding remarks.

6.2. Morphosyntactic processing in bilingual children

6.2.1. Do bilingual children differ in morphosyntactic processing from monolinguals in terms of overall accuracy and efficiency?

The results from the studies in this thesis indicate that bilingual children were overall as accurate and efficient as the monolingual counterparts in processing morphosyntax. For the study on which-questions, there were no differences between the two groups in terms of comprehension accuracy, where the figures for each group were essentially identical, nor were there any differences for the reaction times for the reaction times. The bilingual children had greater difficulty with object questions than subject questions, consistent with the broader literature for this morphosyntactic feature (e.g. Avrutin, 2000; Deevy & Leonard, 2004; De Vincenzi, Arduino, Ciccarelli & Job, 1999; Friedmann, Belletti & Rizzi, 2009; Goodluck, 2005; Jakubowicz & Gutierez, 2007; Stavrakaki, 2006).

Interestingly the asymmetry in the two types of which-question does not vary between groups meaning it is not a source of additional difficulty for the bilingual children as suggested by the absence of group by all structure interactions. The latter finding is consistent with previous work on language processing in children which have shown similar size effects in monolingual and bilingual children in a range of studies with bilinguals with numerous language combinations; bilingual children slow down equally to monolinguals in self-paced listening tasks when they successfully detect grammatical errors (Blom, 2011; Blom & Vasić, 2011; Chondrogianni & Marinis, 2012, 2016; Chondrogianni et al., 2015a; Kaltsa et al. 2016; Vasić et al. 2012) or when they process structurally complex linguistic phenomena such as passives (Marinis, 2007) or interpretation errors in picture-matching tasks (Marinis, 2008). In these studies, the bilingual children are equally as accurate and show the same grammaticality effects, in the same direction and of the same size, as the monolinguals. In other words, exactly like the monolinguals, they slow down when there is something erroneous in the sentence or when the sentence does not match the picture.

One study which shows differential difficulty in processing for bilinguals relative to monolinguals is Marinis & Saddy (2013) which examined the processing of different types of passive sentences using a self-paced listening with picture-matching task. Marinis & Saddy showed that bilinguals had greater difficulty with passive sentences relative to active ones than monolinguals did based on both accuracy and reaction times. This was based on the observation that the bilingual children showed increased reaction times for passive sentences and for sentences where the picture did not match the sentences even after the critical segments but with the monolingual children the effects were found only on the critical segments per se. This would, however, indicate slower processing but not insensitivity to the morphological information such as passive inflections or the preposition by.

The results from the study on garden-path sentences mirrors the results from the first study. Both groups had greater difficulty with the garden-path sentences than with sentences with overt embedded relative clauses. This is evidenced by the lower accuracy for the ambiguous sentences. Consistent with the study on which-questions and with the majority of other studies on morphosyntactic processing in bilingual children, there was no interaction of group and ambiguity or an effect of group on accuracy. The bilingual children were equally as accurate as the monolinguals – in fact, the accuracy score was identical for the ambiguous sentences at 72.8%, suggesting the bilingual children did not have disproportionate difficulty with the ambiguous sentences. The reaction time data are consistent with the first study in this thesis and differs from the previous work which use self-paced listening in that there are no differences between the two groups in terms of reaction time. This may be due to the fact that the study measured response times to click a picture as an answer to a comprehension question; therefore, this was upon hearing the entire sentence and would not reflect real-time processing in the same way that a self-paced listening task would. Reaction times also failed to show an effect of ambiguity. This may be due to the fact that the reaction times were computed by subtracting the duration of the sentence from the total time. This was done as the unambiguous sentences were about 500 ms longer than the ambiguous ones as they contained the overt complementiser and verb "that is". However, using the raw data did not yield any results either. In this respect, the findings from this study differ from those in other studies on processing which usually show longer reaction times for the linguistic feature assumed to be harder (e.g. Marinis, 2007; Marinis & Saddy, 2013) or where there is ungrammaticality (Blom, 2011; Blom & Vasić, 2011; Chondrogianni & Marinis, 2012, 2016; Chondrogianni et al., 2015a; Kaltsa et al. 2016; Vasić et al. 2012). Two points need to be taken into account, however, the reaction time metrics are different in this study to the previous studies and the gaze data are perhaps a better measure of relative differences of speed of processing between groups.

The results from the second study on children are in line with all previous studies on garden-path sentences with children which have measured accuracy in act out tasks (Choi & Trueswell, 2010; Hurewitz et al., 2000; Kidd et al., 2011; Meroni & Crain, 2003; Snedecker & Trueswell, 2004; Trueswell et al. 1999). Children find sentences with local syntactic ambiguity hard to process and may misinterpret them. The majority of the studies tested children aged around five years. This study tested children in the age range of 8-11 years. A study by Weighall (2008) tested children aged 5, 8 and 11 years but did not include an unambiguous control condition. Given the minimal lexical differences between ambiguous and unambiguous sentences, the additional complexity is to be attributed to the structural properties of the two types of sentences.

One difference between previous studies and the current one is that the bilingual children are normally slower than the monolinguals; this was not found in the current one. There are several plausible explanations related to the experimental design, linguistic feature tested and the nature of the bilinguals. The reaction times in this segment do not reflect the on-line difficulty the children encountered at the point of difficulty as was the case in the previous self-paces listening studies. Perhaps the most likely explanation is the high degree of proficiency in English of the bilingual participants recruited for this thesis relative to those in other studies. All participants completed a range of linguistic assessments and the bilingual children scored within age-appropriate norms for monolinguals even though they were competent in the home language based on parental questionnaire. Evidence that bilingual children are slower than monolinguals can be found in the study on which-questions as the looks to the correct picture increase more rapidly for the monolinguals over time. However, this is consistent across conditions suggesting a more general discrepancy in processing speed which mirrors what has been found in studies using self-paced listening (Blom, 2011; Blom & Vasić, 2011; Chondrogianni & Marinis, 2012, 2016; Chondrogianni et al., 2015a; Kaltsa et al. 2016; Marinis, 2007; Marinis & Saddy, 2013; Vasić et al. 2012). This is partially established for the study with garden-path sentences where the monolinguals reorient their looks to the incorrect destination in the ambiguous region faster than the bilinguals but there are no effects of group or interactions for the looks to the correct destination in the disambiguation segment.

In sum monolingual and bilingual children process complex morphosyntactic structures equally as accurately and qualitatively similarly albeit slower.

6.2.2. Do bilingual children successfully build syntactic representations in real time as monolingual children do? Do they revise them, if needed, as monolingual children?

The evidence from the two studies in this thesis suggests that both bilingual and monolingual children process language similarly in that there is incremental processing and recovery from misinterpretation but differ predominantly in terms of timing.

In the study on which-questions, both groups of children showed signs of incremental processing in that looks towards the correct picture matching the sentence heard dropped initially for object questions suggesting it was initially interpreted as a subject question. This is a direct replication of the trajectory of looks to the correct picture as demonstrated in the study by Contemori and Marinis (2018) with younger monolingual children. In this sense, this study is consistent with both Omaki et al. (2014) and Atkinson et al. (2018) who also show incremental processing and consequent misinterpretation of filler-gap dependences in younger children. This means both bilingual and monolingual children do not wait to hear the entire sentence before building a representation but do so on the basis of incoming material. As a result, initial interpretations may be wrong.

The bilingual children showed a more protracted increase in their looks to the correct picture for both subject and object questions. The results from the study on which-questions mirror previous studies which show slower processing using the visual world paradigm (Blom, 2011; Blom & Vasić, 2011; Chondrogianni & Marinis, 2012, 2016; Chondrogianni et al., 2015a; Kaltsa et al. 2016; Marinis, 2007, 2008; Vasić et al. 2012).

More importantly, the differences in increase in looks to correct picture between the two groups are not modulated by the type of structure being processed. This means, the two different trajectories for subject and object do not differ across groups, thus supporting the notion that the bilingual children show the same patterns of incremental processing as monolinguals. Consistent with the findings of this thesis on incrementality, previous studies on morphosyntactic processing in bilingual children have shown that the slowdown in both groups in not only equal but is also observed in the same segments. In Marinis & Saddy (2013), perhaps the only study with sentences which involve reanalysis, the bilingual children show a slowdown in those segments where the monolingual children do and under the same conditions suggesting that they are sensitive to the need to reanalyse misinterpreted sentences. The observation of these effects in subsequent segments for the bilingual children was not interpreted by Marinis & Saddy (2013) as suggestive of qualitative differences between the two groups but that the same patters are observed for longer in the bilingual children. This could be the by-product of slower processing and, in this sense, would be in line with the findings in this thesis. However, Marinis & Saddy also found a larger increase in reading times for the passive sentences in the bilinguals relative to the monolingual children, suggesting greater difficulty despite similar processing mechanisms. Something similar was not found in the studies in this thesis.

The results from the study on garden-path sentences show weaker effects of incrementality and misinterpretation. In the ambiguous region, there was no overall effect of ambiguity on the looks to the incorrect destination but an effect on the linear term in addition to an interaction of group by ambiguity. This means that, although looks were overall not different between ambiguous and unambiguous conditions, here there was a faster increase for the ambiguous sentences for the monolinguals not found in the bilinguals. This would be consistent with the assumption that the monolinguals are committing to a wrong interpretation over time prior to the disambiguating material. Main effects were observed for gaze data in other studies for the equivalent looks to the incorrect destination in the ambiguous region (e.g. Snedecker & Trueswell, 2004; Trueswell et al. 1999). The difference could be attributed to the use of multilevel analyses in this study, consistent with current practise in psycholinguistic research, which are generally more conservative, the difference in the age of the participants (they were older in this study; but then other studies have also found the aforementioned effects for adult controls similarly to the adult participants in Study 1BB), or the nature of the dependent variable (fixations calculated by manual coding as opposed to ratio of looks recorded electronically for this study[35]) among others.

The two studies should also be considered in light of perhaps the only study on incremental processing and revision in bilingual children (Papangeli & Marinis, 2010). In this study, Papangeli & Marinis tested garden-path sentence processing in Russian-Greek bilingual children using a self-paced listening task. The bilingual children did not show garden-path effects as the monolingual children did; they did not slow down at those segments which disambiguated in favour of the unlikely interpretation. The bilinguals and the monolinguals in the study on garden-path sentences in this thesis did not differ on a range of measures of gaze data, in particular looks to the correct destination; for both groups there was repeatedly an effect of ambiguity with fewer looks to the correct destination for the ambiguous sentences. This indicates difficulty in processing but do not definitively demonstrate garden-path effects. Looks to the incorrect destination are the most direct measure to demonstrate the latter. The study in this thesis on garden-path sentences shows similar findings as Papangeli & Marinis; although the garden-path effects were weak, there was an increase in looks to the incorrect destination at the point of ambiguity suggesting a misinterpretation of the sentence, but this was steeper in the monolinguals. This would suggest that, consistently with Papangeli &

[35] In this respect, the studies in this thesis are similar to Lemmerth & Hopp (2019).

Marinis, only the monolingual children showed some degree of misinterpretation. This is perhaps not surprising given that the children in Papangeli were in the age range of 9-15 years whereas the children in this study were aged 8-11. Moreover, the fact that the failure to observe garden-path effects online in the bilingual group with various L1s in this thesis suggests that the findings from Papangeli & Marinis are not due to cross-linguistic influence from the L1, in this case Russian. One criticism of the Papangeli & Marinis (2010) study is that the region of disambiguation, where one would expect to find a slowdown, is the last segment in the sentence for which reaction times were recorded. It may be plausible that a slowdown could have been observed in the subsequent segment as a spill over effect. The results from the study in this thesis do not support this. There were two segments analysed following the ambiguous segment. None of them showed an effect of ambiguity for the looks to the incorrect destination or an interaction of group by ambiguity.

On the other hand, the results from the study on which-questions are in contrast to the findings from Papangeli & Marinis; both groups showed a bias to misinterpret object which-questions as subject questions. In fact, the differences in the rates of increase between the two groups might suggest that recovery is slower in the bilingual children. However, this should be seen not as a susceptibility to ambiguous or syntactically complex sentences but as a manifestation of generally slower processing. In other words, the bilingual children behave similarly to the monolinguals and misinterpret ambiguous which-questions this was not found in the study on garden-path sentences in the bilinguals to the same degree that monolinguals do.

There are a number of reasons as to why this may be the case with the garden-path sentences. Firstly, wh-questions are more frequent in oral language. This would entail more quantitatively adequate exposure for probabilistically-driven biases to be formed for both groups. Not only are garden-path sentences less frequent in the input, they are typically heard (or read) alongside some type of disambiguating information. In oral language, intonation usually disambiguates either by marking clause boundaries or by providing an emphasis. Clause boundaries are usually also marked in written language through punctuation. Therefore, genuinely ambiguous garden path sentences are highly infrequent and largely non-natural for speakers of a language.

It is possible that the difference in findings may be attributable to the nature of the linguistic features themselves. In the case of the which-questions, the wh-phrase is in a non-canonical position and its presence overtly signals that another position further downstream in the sentence will be empty but unavailable. Under the Active Filler hypothesis (Frazier & Clifton, 1989), this is known as a filler-gap dependency and it is assumed the parser anticipates the upcoming gap and will attempt to close this dependency as early on as possible in order to minimise cognitive load. Subject wh-questions are thus, the simpler option. Under alternative structural accounts, the wh-phrase is associated with the verb's subcategorization frame (Pickering & Barry, 1991; Pollard & Sag, 1994; Sag & Fodor, 1995); the unavailability of the empty position would be due to the speaker's knowledge that the verb's subcategorization frame has already been satisfied. Following this account, the parser would anticipate the need to integrate the dislocated constituent into the verb's frame. Yet even in this case, the wh-phrase signals an upcoming verb and the need for structure to accommodate the subcategorization the frame of the verb. However, by definition with garden-path sentences, the parser has no overt signals towards anticipating local ambiguity or the need for reinterpretation. It may be that filler gap-dependencies are easier to process and reanalyse than garden-path sentences as the non-canonical word order is a cue to the parser that upcoming structure will be needed for interpretation of the sentence. In garden-path sentences, context may guide the parser towards a particular interpretation, however, it is unclear whether this functions in the same way as morphosyntactic cues.

To conclude, both monolingual and bilingual children build syntactic representations incrementally for which-questions and reanalyse upon disambiguation. Bilingual children are slower to recover from misinterpretation, but this is not disproportionately difficult relative to the monolinguals. The bilinguals did not appear to build erroneous misinterpretations to the same degree as monolinguals in the garden-path sentences.

6.2.3. Do bilingual children make use of various cues during language processing to facilitate processing in the same way as monolingual children?

For the study with which-questions, the disambiguating cue was number mismatch between auxiliary verb and the first NP. The rationale was that, in sentences such as "Which donkey are ….", the parser will infer that the first noun phrase is not the subject of the verb, as that would be a violation of subject verb agreement. Therefore, it needs to have a different syntactic function. The effect of number mismatch was selective in the accuracy data for structure but not group. Accuracy was higher for number mismatch than match for object questions but not subject questions. Crucially, this was the case for both monolingual and bilingual children. This is a direct replication of Contemori and Marinis (2018) for the younger monolinguals. The two groups differ, however, in the reaction times and critically the gaze data. For the reaction times, the bilingual children benefitted from number mismatch across subject and object questions whereas there were no effects for the monolinguals. In contrast, there was a faster increase in looks to the correct picture for the monolingual children in the object questions when the number of the auxiliary and of the first NP did not match but no such effect for the subject questions or the bilingual children in either questions type. This indicates that both groups of children could utilise the cue in question to disambiguate but did so over a different timecourse. The bilinguals were able to use it by the end of the sentence as were the monolinguals but only the monolinguals were able to use it in real time as measured by the gaze data.

In Contemori et al., the effect of number mismatch was larger in the children, who also showed larger effects of misinterpreting the object questions, than in the adult controls. It is, therefore, reasonable to conceptualise the use of number mismatch as a compensatory mechanism which was largely redundant in the adult controls. The results from this thesis are a replication of the findings for the monolingual children albeit at an older age. Yet the bilingual children did not appear to use this mechanism as they heard the sentences. The study by Contemori did not report reaction times and as such, cannot be compared with the results from this thesis on this measure.

The results from this study are partially in line with those from Roesch & Chondrogianni (2016) who used an off-line task to investigate how nominal case in German can function as a disambiguating cue in processing which-question in French-German children. The results from Roesch & Chondrogianni for accuracy reflect what was found in this study for the gaze data; it was namely reduced and more nuanced for the bilinguals. Whereas monolinguals used case in all conditions (cue at the first NP, cue at the last NP and cue on both), the simultaneous bilinguals did so only for double cue and the early sequential bilinguals did not benefit from it at all. However, the children in this study were mostly simultaneous bilinguals and were not able to use the cue to disambiguate in real time even at an older age. On the other hand, in this study there were no differences in use of number in the off-line data for accuracy in the comprehension questions. In both studies, the cue is purely linguistic in nature. However, the two studies differ in several ways; the language they are testing the children in (English vs. German), the age of the participants (older in this study) and length of exposure (about 2-3 years in Roesch & Chondrogianni, substantially longer for this study) as well as the experimental design. It is perfectly plausible that the children in the Roesch & Chondrogianni study cannot use the cue in real time either and that the ability to do so increases with age and/or exposure for the bilinguals and becomes more automated. A further reason is the nature of the linguistic features tested in both studies; case is highly idiosyncratic and lexicalised in German and may therefore, require a larger quantity of input certainly for acquisition and potentially for its use in real time processing. This would explain why the simultaneous bilinguals still used it less than the monolinguals in Roesch & Chondrogianni. Subject –verb agreement, however, is more systematic and may be readily abstracted with less input. It may be possible that this is also the case for children to be able to use it.

The results from the study of which-questions could also be seen in conjunction with other studies on how bilingual children utilise information encoded in inflectional morphology in real time in a comparably small time window which, conveniently, have also used the visual-world paradigm. These studies have investigated how bilingual children utilise grammatical gender on determiners in order to predict the upcoming noun – on the assumption that the upcoming noun will match the determiner for grammatical gender. Lemmerth & Hopp (2019) tested German-Russian children on the younger end of the age range to this thesis who were either simultaneous or early sequential bilinguals. They found that the early bilinguals consistently utilised gender as evidenced by an increased amount of looks to a picture with the corresponding grammatical gender as the determiner. However, the sequential bilinguals showed this effect only for those nouns where the grammatical gender was consistent in both Russian and German. Lew-Williams (2017) did the same for Spanish with simultaneous English-Spanish bilinguals and English monolinguals with exposure to Spanish through immersion education aged 6 or 10 years, similarly to the participants in this thesis. Lew-Williams found that the simultaneous bilinguals utilised grammatical gender to predict the upcoming noun, but the late sequential bilinguals did not, although they behaved similarly in other eye-tracking studies. The results from the study in this thesis are broadly in line with these previous studies in that they point to a reduced use of cues to predict upcoming material and therefore, aid disambiguation. Taken together the results from the two earlier studies suggest that use of gender cues in bilinguals is contingent on the context of acquisition of the additional language; bilingual children who acquired the additional language more similarly to monolingual children will exhibit a pattern of processing akin to those observed in monolinguals. This mirrors findings in the literature for processing in adult L2 learners (Dussias, 2003; for a review on native vs. non-native processing, see Cunnings, 2017). However, the findings of this study show even more reduced use of morphological cues in real time relative to the previous literature, as the bilingual participants are mostly simultaneous, some being early sequential bilinguals and only about 1 in 5 being classed as late sequential bilinguals based on parental reports. There are at least two reasons which may account for this difference. Previous studies have investigated grammatical gender on the determiner. The first reason relates to the nature of the morphological features tested. The use of grammatical gender on the determiner to predict the upcoming noun involves a computation at the phrase level between the head and the specifier. Grammatical gender of nouns is highly idiosyncratic to individual lexical items and therefore, would be susceptible to differences in the input bilingual children may receive relative to their monolingual peers. However, the number of determiners in any language is likely to be limited and their frequency is high. Therefore, using the determiner's gender to predict the upcoming noun may be less of a process of structural computation than an indirect process of lexical retrieval. In contrast, the use of number mismatch between noun phrase and verb is not just one of number but also subject verb agreement. It involves computation at the sentential level. Although there is little involvement of lexical knowledge, it can be conceptualised as a highly structural process. The second reason why the results differ somewhat may be methodological. Both Lew-Williams and Lemmerth & Hopp used time of first fixation as the dependent variable. This study calculated proportion of looks towards the correct picture in

200 ms bins for 2,000 ms starting at 200 ms after the onset of the auxiliary verb where there is the cue which could act as a disambiguation. Essentially, the former captures speed while the latter captures change over time. Both methods of analyses have merit and it is unclear how the different methodologies may have impacted the results. Capturing change over time may have been more appropriate for the purposes of this study, as analysing the data by individual bins would have reduced statistical power, something potentially problematic given the more complex experimental design in terms of independent variables relative to the previous studies.

The results from the study on garden-path sentences showed that neither monolingual nor bilingual children utilised referential context in order to disambiguate faster. This is consistent with Trueswell et al. (1999); in the ambiguous region, they found an effect of ambiguity, looks consistent with the erroneous interpretation were more for the ambiguous sentences but there was no difference between referential contexts, nor did the latter interact with ambiguity. The results for the respective gaze data from this study support this. Results from the accuracy of the act out replicated these findings both in Trueswell et al. and in this study. This is in contrast to the adult controls in Trueswell et al. (1999) but consistent with the adult participants in Study 1BB of this thesis. The children in Trueswell et al. were younger than in this study (4-5 years as opposed to 8-11); however, the developmental change required for children to utilise this non-structural information appears not yet to have taken place. For some of the analyses, the presence of the second referent appears to have had the reverse effect than anticipated; this is the case for the reaction times; participants were marginally slower to respond to the comprehension question by selecting a picture when there were two referents in the same picture. The presence of the second referent also lead to an increase in looks to the incorrect destination in a number of segments. Counterintuitive as this may be, it is also attested in the Trueswell et al. study where looks to the incorrect destination were more for the 2-referent than the 1-referent context. Even though, this is not the main point of the Trueswell et al. (1999) study, it fits into a broader picture in the relevant literature (e.g. Snedeker & Trueswell, 2004; Weighall, 2008) that looks in the ambiguous condition are not compatible with what would be expected if children were using context to aid disambiguation. There are three possible exceptions to this in this study. There was an interaction of referential context by ambiguity on the looks to the correct destination in the region of disambiguation (i.e. when participants heard the second PP); looks to the correct destination began to increase later for the ambiguous 1-referent condition than for the other three conditions (2-referent ambiguous and the two unambiguous conditions). Moreover, looks to the correct destination declined more slowly in the final segment of the sentence.

Trueswell et al. did not analyse the gaze data in windows timed to the onset of the different segments but conducted analyses for 200ms time windows following the onset of the ambiguous phrase. This corresponds loosely to the segments post ambiguity analysed in this study. They found no benefit for the 2-referent context in disambiguating the sentence; visual inspection of the gaze data confirms this. In this sense, the results from this study show somewhat a similar absence of effects, or at best only tentative effects of children utilising referential context to disambiguate. Nonetheless, the children in this study are older and the effects occur after the ambiguous region and are related more so with looks to the correct destination, indicating that the value of the second referent is more readily postulated as an aid to comprehension or recovery from misinterpretation at the end of the sentence rather than in real time disambiguation or avoiding garden-path effects.

In all cases, these effects of referential context did not interact with ambiguity in the way they did for the adults in Trueswell et al. (1999). This indicates that the referential context aided comprehension, but this did not happen at the point of ambiguity and was not used to disambiguate between two possible interpretations. Moreover, there was no interaction between group and context in any analyses where there was a significant effect of context. This entails that both bilingual and monolingual children were impacted by referential context in the same manner, but this suggest they context aided disambiguation in ambiguous sentences.

The question now arises about what differentiates the two linguistic structures examined in this thesis. The absence of effects of referential context was found in both groups and can be justified on the basis of development. This, however, cannot account for the difference in use of number mismatch between the monolingual and bilingual children in the study on wh-questions. There is little reason to expect bilingual children to underperform their monolingual peers in terms of using visual cues to guide their interpretation of a sentence as long as they can comprehend the syntactic structure as this is not a language mediated source of information. What might make number mismatch between the two NPs harder for the bilingual children to use in real time relative to the monolinguals, as evidenced by the gaze data, is that they are grammatical in nature. The mismatch between the number of the first NP and the auxiliary verb can only be utilised as a disambiguating cue through subject-verb agreement which in English includes both person and number. Bilingual children may not be equally as able to use purely structural information in real time as monolingual children with the equivalent proficiency in the language tested. This should not be interpreted as evidence that processing in bilingual children is less structural in nature ("shallow", as it has been argued for adult L2 learners, see Clahsen & Felser, 2006a; 2006b; Cunnings, 2017). Instead, it is potentially a reflection of the fact that processing is slower and thus, may not be able to undertake the required computation to integrate the information from the number mismatch in time. Therefore, the effect may be observed in the reaction times, a fine grained but still end state measure. In sum, it could be argued that the evidence is stronger for children to some degree utilising number mismatch between NPs to disambiguate (albeit not in real time for the bilingual) than it is for using referential context. The fact that number mismatch is a core part of morphosyntactic agreement and is present within the sentence may account for this. In contrast, referential context is extra-sentential and is part of the discourse or visual context. This would make it harder for both bilingual and monolingual children and, thus explain the absence of effects even at the stage of the comprehension question.

To summarise, the bilingual children used number mismatch between the two noun phrases to facilitate processing of object which-questions, as did the monolinguals, but did so only in the accuracy but not the gaze data. Similarly to the monolingual children, the bilinguals did not utilise referential context in the visual stimuli to disambiguate garden-path sentences.

6.3. Morphosyntactic processing in bilingual adults

6.3.1. Do bilingual adults process morphosyntax similarly to monolinguals or L2 learners in terms of overall accuracy and efficiency?

Bilingual adults had greater difficulty with the object questions than subject questions as evidenced by the lower accuracy scores and slower reaction times. This is consistent with the broader literature for both children and adult, monolingual and bilingual, typical and atypical groups across a range of languages (e.g. Avrutin, 2000; Crain & Fodor, 1985; Deevy & Leonard, 2004; De Vincenzi, Arduino, Ciccarelli & Job, 1999; Friedmann, Belletti & Rizzi, 2009; Goodluck, 2005; Jakubowicz & Gutierez, 2007; Juffs & Harrington; 1995; Stavrakaki, 2006; Stowe, 1986 among others). The native bilinguals differed to L2 learners; in accuracy, scores were overall marginally higher for the bilinguals than for the L2 learners but not in comparison to the monolinguals. This is similar with the results from the child data, where the bilingual and monolingual children were equally as accurate. However, unlike with the children, the bilingual groups also had greater difficulty with the object questions relative to the subject questions than the monolinguals. This interaction was not found in the native vs. non-native bilinguals contrast. In this respect, the bilinguals performed more similarly to the L2 learners than the monolinguals. In Contemori, Carlson & Marinis (2018), there was no difference between the two structures as accuracy scores reached ceiling effects for the adults, although the few errors made were almost all in object questions.

A similar picture emerges for the reaction time data. The participants were overall slower responding to object which-questions than for subject which-questions, which is consistent with the wider literature. However, the bilinguals were overall slower than the monolinguals and not faster than the L2 learners. Again, the bilingual groups showed greater susceptibility to the additional difficulty of object-questions relative to the monolinguals but did not differ from the L2 learners. In fact, the monolinguals, as an individual group, did not show a slowdown for object-questions. This is in contrast to the child data where there were no differences found between the monolingual children in terms of reaction times and where speed for each group was not modulated by the type of structure but does mirror other studies in terms of speed of processing in children (Blom, 2011; Blom & Vasić, 2011; Chondrogianni &Marinis, 2012, 2016; Chondrogianni et al., 2015a; Kaltsa et al. 2016; Marinis, 2007, 2008; Vasić et al. 2012). In terms of gaze data, there were overall more looks to the correct picture for the subject which-questions than for the object questions. This was also found in Contemori et al. and is in line with the findings for the previously discussed metrics. Differences in time course of gaze data are discussed in the subsequent section.

For the study on garden-path sentences, the results for the adults were similar to those of the children, at least with regards to overall accuracy and efficiency. Accuracy in responding to the comprehension question was lower for the ambiguous garden-path sentences than for the unambiguous ones with the unreduced relative clauses. In both studies with adults, accuracy was lower for the condition that would be considered the hardest. For the garden-path sentences, this is consistent with previous studies with adult control groups (e.g. Trueswell et al., 1999) even when the errors are minimal. While off-line results are not optimally informative regarding assumptions about processing, they point to a consistent difficulty with structurally complex sentences with an increased likelihood for a need to reanalyse them. This may be due to initial misinterpretation and the fact that the effect is found for offline measures indicates a persistent difficulty in abandoning the earlier interpretation. Similarly to the child study on garden-path sentences, there were no differences between the groups. In this respect, the results from this study differ with the ones from the adult study on which-questions. Two issues need to be considered at this point. Firstly, the effects in the second study were numerically very small and translated mostly into trends in terms of statistics. Second, the nature of the linguistic features tested is different. The first study tested wh-questions which while involve a need for reanalysis in the case of object questions, also contain morphosyntactic information (subject-verb agreement) as well as non-canonical word order which triggers the need to anticipate an empty position for the dislocated constituent. For garden-path sentences, the difficulty lies in the need for reanalysis, but there is no trigger to anticipate the need for gap-filling nor is the difficulty related to morphosyntax. It may be the case that the challenge in processing garden-path sentences is more cognitive than linguistic.

The findings are also partially in contrast to Pozzan & Trueswell (2016) where the L2 learners had lower accuracy than the monolinguals (although both studies showed an effect of ambiguity). It should be noted that Pozzan & Trueswell measured act outs and, in particular, towards the incorrect destination consistent with the initial erroneous reading. This study, on the other hand, reports comprehension accuracy on comprehension. In this study, the accuracy of the L2 learners is higher than the L2 learners in Pozzan & Trueswell although lower for the monolinguals. In the Pozzan & Trueswell study, the main effect of group is driven by the fact the monolingual participants essentially make very few mistakes. Moreover, the L2 learners in Pozzan & Trueswell were recruited in Italy and, having received only classroom exposure, may have had lower proficiency in English than those in this thesis. This is consistent with works in the literature which suggest that more naturalistic exposure leads to more nativelike processing (e.g. Dussias, 2003; Hopp, 2015, Pliatsikas & Marinis, 2013c; Roberts & Felser, 2011). The findings, in this sense, do not replicate other studies which show L2 learners to have difficulty with reanalysis (Jacob & Felser, 2016; Pozzan & Trueswell, 2016).

In terms of reaction times, the results for the adult study on garden-path sentences are partially consistent with the one on which-questions. The bilinguals from both groups are slower than the monolinguals but no differences were found between the two groups of bilinguals. This is consistent across the two studies with adults but differs to the respective study with children. The fact that reaction times are longer for ambiguous sentences further supports the notion that garden-path sentence processing entails some difficulty in reanalysing them which takes time even when done successfully. This is similar to both the adult study on which-questions but also the respective study with children.

In sum, the two groups of bilinguals showed similar patterns of processing to monolinguals in both the study on which-questions and the study on garden-path sentences. Object questions and locally ambiguous sentences had lower accuracy across groups relative to subject questions and unambiguous sentences respectively. This is indicative of incremental processing and consequent misinterpretation. The gaze data reflect this for both studies. For the study on which-questions, the two groups of bilinguals were slower than the monolinguals in terms of reaction times and changes to the gaze data.

6.3.2. Do bilingual adults successfully build syntactic representations in real time as monolingual adults? Do they revise them, if needed, similarly to monolinguals or have difficulty similarly to adult L2 learners?

It should be noted that the high rates of accuracy for the object questions even for the L2 learners (>90%) suggests that even the L2 learners were largely in a position to revise syntactic misinterpretations but failed more frequently than the other two groups. This is further supported from the reaction times, where longer time needed to respond to object questions when doing so accurately relative to subject questions could be a spill over effect of the need to reanalyse. The fact that this discrepancy is present in the two offline measures indicates that difficulties in reanalysis may persist. In contrast to the accuracy data, the impact of syntactic structure was different between bilinguals and monolinguals – but not between native bilinguals and L2 learners; the slowdown in reaction times for object questions was driven by the two groups of bilinguals and not the monolinguals. This would indicate that the bilinguals had greater difficulty recovering from garden-path effects even after they have heard the entire sentence suggesting persistently lingering misinterpretations.

This is further supported by the change in the gaze data over time. Subject and object questions have a completely different timecourse. This difference in patterns is consistent with what was found in children although the children showed a more protracted plateau below 50%. This may be accounted for on the assumption that the adults had more efficient processing skills than the children and began to reanalyse faster. Qualitatively, it is the same change over time and can be explained by garden-path effects which have occurred in all three groups. Where the results from the adult study differ from the children's study is in the effect of group and the interaction of group with structure on how the looks to the correct picture change over time. These differences appear to be associated with the monolinguals vs. bilinguals contrast and not with the natives vs. non-natives or native bilingual vs. L2 learners.

This would suggest that the two groups of bilinguals have a shallower increase or an increase with a later onset in looks to the target reflecting somewhat slower processing relative to the monolinguals. Interestingly, the differences are more pronounced in the subject questions which are assumed to be easier. This would indicate that monolinguals and bilinguals build syntactic interpretations incrementally although the monolinguals do so faster but recover from misinterpretation equally as efficiently. In this sense, the results from the adults mirror those from the children.

The native bilinguals also differed from the L2 learners but not the monolinguals in terms of how the number of the first NP impacts the increase in looks to the correct destination over time. For the bilinguals, there was no difference between the two but for the L2 learners, there was an initial lag for sentences with a plural first NP relative to those with a singular one but this increased more steeply. This may be explained under linguistic markedness with the plural being considered the marked form; therefore, the first NP plural trials may be harder than the singular first NP ones. Marked features have been proposed to be broadly harder for L2 learners (Eckman, 1977; White, 1987).

For the study on garden-path sentences, there is stronger evidence that participants are forming interpretations of the ambiguous sentences early on than in the respective study with children. In the adult study, there is an effect of ambiguity on the amount of looks to the incorrect destination whereas in the children's study, ambiguity has an effect on the increase of looks to the incorrect destination but not overall. This is consistent with the potential that adults process language faster than children, as the effects are observed at the earliest time window possible. This mirrors the findings of the study with which-questions, where adult participants were shown to initially misinterpret them in real time as subject questions. There were overall more looks to the incorrect destination at the point of ambiguity consistent with a garden-path effect and looks to the incorrect destination showed a more rapid increase for the ambiguous sentences than the unambiguous ones. This indicates that the participants are committing to an interpretation which is erroneous. This tendency continues numerically in the disambiguating segment but does not reach statistical significance and therefore, points to some lingering misrepresentations but not to the participants' inability to recover from them. Further evidence for this comes from the slower increase in looks to the correct destination in the disambiguation segment for ambiguous sentences than for unambiguous ones. This was also the case later on as there was the same effect on looks to the correct destination overall for the final segment of the sentence.

In all, accuracy rates, reaction times and the gaze data suggest that all three adult groups are able to recover from misinterpretation in the vast majority of trials, although some difficulty may occasionally persist. For the which-questions, reaction times and gaze data show a slower recovery for both groups of bilinguals relative to monolinguals but not ultimately less successful. For the garden-path sentences, all three groups are equally successful in recovering from misinterpretation.

6.3.3. Do bilingual adults make use of various cues during language processing to facilitate processing in the same way as monolingual adults? Do they differ in this respect to L2 learners?

For the study on wh-questions, evidence for this is derived from the interaction of structure and number match in the accuracy data; the effect of number mismatch is present only in the object questions and not in the subject questions. This is consistent with the accuracy data from the study with children. Number mismatch is expected to facilitate processing in the object questions under both Relativized Minimality (Rizzi, 1990, 2004; Friedmann et al. 2009) and similarity-based interference accounts (Gordon et al. 2001, 2002, 2004) so the accuracy data confirms predictions made under these accounts. The interaction of structure by number match interaction was found for the entire dataset. Nonetheless, pairwise analyses undertaken in light of a group by structure interaction indicated it was reached significance only in the monolinguals. It is unlikely that the bilinguals and L2 learners are unable to utilise the cue from number mismatch after hearing the entire sentence given that the bilingual children were able to do so. Absence of effects may be accounted for on the basis that the participants are at ceiling and thus the benefit from number mismatch is redundant or at least, challenging to test statistically due to small variance. But it is the monolinguals who are at ceiling performance. For the bilingual groups, on the other hand, in particular the L2 learners, the overall accuracy is slightly lower and therefore, the number mismatch could still contribute to disambiguation. Therefore, this explanation is implausible in this particular case. The effect was numerically very small as was the variance. Given that the sample size was smaller for each of the two bilingual groups relative to the monolinguals, it is possible that the subsequent loss of statistical power meant a smaller effect was not detectable.

For the reaction times, the picture is somewhat different. Overall, there was no effect of number match between the two NPs; the benefit of the number mismatch led to faster reaction times when the first NP was plural but not singular. This did not vary by group. The findings would not be consistent with either Relativized Minimality or similarity-based interference accounts as these do not predict the interaction found and the number of the first NP is conceptually not relevant to the claims they make. They may, however, be associated with markedness; the plural number is considered the marked form. As such, it may be harder to process and thus, the presence of the disambiguating cue may have been more valuable. For this metric, the effect appeared consistent across groups. This effect is also found in the child data but only for the bilingual children. The reasons for this discrepancy between groups across ages were unclear; it is difficult to conceptualise linguistic markedness and cues as facilitatory of a linguistic system under development as the bilingual children scored within age-appropriate norms in background language testing and had identical accuracy rates as monolingual children in this task. Viewing these findings in conjunction with the gaze data may offer an interesting explanation; the effects found for the reaction times are a reflection of the same processing mechanism that utilises linguistic information. However, in the bilingual children, it is not accessible in real time but can be deployed after the sentence is complete. The adult participants seem to be able to use it at both time points[36].

The value of number mismatch is further confirmed by analyses of the gaze data which provide an insight into real time processing. Looks to the correct picture were initially more for sentences with a number mismatch but increased more slowly in relation to the sentences with the number match. This indicates an initial advantage which disappeared after about 800-1,000ms. The same effect was found in the children but the timecourse was different; the emergence of the number mismatch was absent initially but then emerged and subsequently disappeared. This may be due to slower processing in children relative to adults which would indicate that effects observed in real time may occur later in the former. Furthermore, in the adult data, as with reaction times, the benefit from number mismatch was modulated by the number of the first noun phrase. Looks to the correct destination in the match condition were initially lower than in the mismatch condition relative to the respective sentences in the singular first NP conditions. The onset of the increase was slower than for the plural mismatch and the other three conditions, but steeper once it began. Conversely, there was little difference between match and mismatch in the singular first NP conditions. This suggests the benefit of number mismatch in real time processing is in line with what was found for the reaction times; it was larger in the otherwise harder condition with a plural first NP. This was completely absent in the children. It is unclear why this may be the case. To a

[36] It should be noted that analyses by groups separately were not undertaken for the adults as the interaction of group by number match by number of first NP was not significant.

certain extent, it is counterintuitive given that the effects related to difficulty would be more readily expected in populations with a developing system of language rather than those with a fully-fledged linguistic system. A tentative explanation may be that in a developing system any advantage conferred by frequency or un-markedness has not yet emerged.

There was no effect of referential context on comprehension accuracy for the garden-path sentences. This is in contrast to the study where there was a weak effect of number for the object questions only. It is unclear why this is the case; accuracy on the which-questions task is higher than on the garden-path sentences study which means it cannot be attributed to ceiling effects. However, the variability in the garden path study is much larger than for the wh-question study for the accuracy data even though it is similar across groups. In a statistical analysis using mixed effects model, the effect of a fixed variable is allowed to vary on the slope by subject. This entails that variability is reduced ("shrunk") towards the minimum shared across participants, the impact of outliers is reduced and as such, they are more conservative than a traditional ANOVA[37]. It is also contradictory to Trueswell et al. (1999) who found an ambiguity by context interaction; the second referent namely benefited act out in the ambiguous sentences. This, however, was driven by the fact that almost all erroneous act outs were in the ambiguous 1-referent condition. The crucial difference between the two studies may lie in the nature of the cues; in the which-questions, the cue is a grammatical one and occurs within the auditory stimuli whereas in the garden-path sentences, the cue is contextual and provided in the visual stimuli. This may explain the differences between the studies, but it is still unclear why context did not aid ambiguity resolution in the adult data in contrast to previous studies.

[37] ANOVA still takes into account this slope variance but in a different way. The main difference is mixed models take into by-subject and by-item variance all at once, while ANOVAs can only do one of these at a time.

The reaction times show different results; they were slower for the 2-referent context overall. This is in the opposite direction of the effects found in Trueswell et al. for accuracy and gaze data. While counterintuitive, this may be driven by the additional complexity added by the second referent in the visual stimuli. Therefore, participants needed more time to select the correct picture as they spent more time exploring which one is the appropriate answer. If the participants were using context to disambiguate, then they would not be predicted to be doing this. This effect was only marginal in the children suggesting the effect was substantially weaker. While with previous studies such as Trueswell et al. who show that the adults are impacted by the presence of cues in the referential context to a greater extent than the children, the fact that this occurs in an unexpected, i.e. the opposite, direction means that the results from this study are strongly inconsistent with previous ones. For both adult and children studies, there was no interaction of group with referential context suggesting that both monolinguals and bilinguals made similar (non)use of it by the end of the sentence. Overall, looks to the incorrect destination increased earlier on for the adults in the ambiguous segments although analyses per groups failed to show an effect. However, the adults showed some effects consistent with what would be expected given the previous literature, but only weakly; looks to the correct destination in the region of disambiguation increased faster for the 2-referent condition for the ambiguous sentences relative to the ambiguous trials with 1 referent in the visual stimuli but not for the unambiguous ones. In this sense, the adults showed some signs of sensitivity to referential context in real time at points where the children did not, even though they utilised number mismatch in the which-questions study. Given that these differences can be explained on the basis of developmental changes, differences between bilinguals and monolinguals are not expected. On the other hand, the use of number mismatch is using a grammatical cue; thus, differences between bilinguals and monolinguals are more probable irrespective of age. When the change to looks to the correct destination in the ambiguous region were broken down by groups, both groups of bilinguals showed a larger increase in looks to the correct destination than the monolinguals in the 2-referent relevant to the 1-referent context. These effects of context, in the same expected direction, continued into the final segment and lead to a slower rate of looking away from the correct destination. This was also found in the children's data suggesting continuity in terms of qualities of processing between the age groups. It indicates that cues can still aid disambiguation even after the critical segment.

In sum both monolingual and the two groups of bilingual adults used number mismatch to disambiguate object which-questions. However, there is only very weak evidence for the use of referential context being used to disambiguate garden-path sentences.

6.4. Implications for theories of language acquisition and processing in bilingual children and adults

This section discussed how the findings from the studies in this thesis inform numerous theoretical accounts of L2 acquisition and L2 processing. The accounts are divided into accounts regarding acquisition and accounts regarding processing. Each account originates either in bilingual child or bilingual adult language but many of them have been expanded.

6.4.1. Theories of L2 acquisition

It has been proposed in the literature that the syntactic representations of bilingual children are deficient or different to those of monolingual children (Haznedar, 2001, Vainikka & Young-Scholten, 1993, 1996a, 1996b, 1998a, 1998b, 2010). These accounts usually assume a reduced syntactic structure which becomes enriched with exposure. This is not possible to test as part of this thesis as the participants had substantial exposure to English. However, given the absence of differences observed between monolinguals and bilinguals, the results from the studies in this thesis do not support the prospect of differences in representation between bilingual and monolingual children. This is also likely to be the case for the adults given the overall high accuracy and the absence of between group differences on this measure. Alternative accounts have assumed there is no representational difference in adults but that the differences are due to mapping morphology onto words or morphemes (Goad & White, 2004; Lardiere, 2009; Prévost & White, 2000). While these accounts make predictions largely about inflectional morphology rather complex morphosyntax, the findings from the studies on which-questions which show few if any qualitative differences between bilingual and monolingual adults and children respectively, it is unlikely that any of the bilingual groups has difficulty mapping morphology in real time comprehension.

Numerous theories of acquisition postulate the source of difficulty not in the representation per se or the absence thereof but in the late exposure to these features in the second language (Hawkins & Chan, 1997; Hawkins & Hattori, 2006; Tsimpli, 2003; Tsimpli & Dimitrakopoulou, 2007; Tsimpli & Mastropavlou, 2008). The children in these studies were exposed to English as an additional language mostly from birth while some were first exposed to it after their native language but still young while going to nursery school. Very few children could be classed as late sequential bilinguals. Given this, it is not expected that these children would differ from monolinguals. Age of onset of acquisition and length of exposure did not correlate with overall accuracy suggesting that the bilinguals had received adequate exposure to acquire the morphosyntactic features when tested. However, the Interpretability hypothesis (Tsimpli, 2003; Tsimpli & Dimitrakopoulou, 2007; Tsimpli & Mastropavlou, 2008) makes some assumptions which may be relevant to the study on which-questions. Agreement features are assumed to be uninterpretable; they have no representation in the logical form, they have no conceptual or semantic information but are only represented at the linguistic system. Under the Interpretability hypothesis such features are assumed to be problematic in the L2. It should be noted that reduced use of number mismatch is not directly predicted under the Interpretability hypothesis but can be assumed by extrapolation. The latter makes crucial assumptions about timing of exposure (needed early on) and (un)availability of features in the L1 (difficulty if absent); in fact, it originates in adult L2 acquisition and has been expanded to child language. It offers, nonetheless, plausible and testable predictions about which features may be harder to utilise in real time. The bilingual children could not use number mismatch in real time even with early exposure to English and even though their native languages mostly had subject-verb agreement. Yet they did not differ from their monolingual peers in the garden-path sentences task. The adult data are more problematic for the Interpretability hypothesis as the differences between groups are limited and participants had lower accuracy on the garden-path sentences task, which does not involve uninterpretable features unlike the which-question study which does. The findings are also problematic for the Failed Functional Features Hypotheses (Hawkins & Chan, 1997, Hawkins & Hattori, 2006) for two reasons. Firstly, it assumes the need for exposure within an early time window; the bilingual children a delayed use of number mismatch despite their generally early exposure to English. Crucially, both bilingual adults and children seem to differ from their monolingual peers in terms of timing. Processing is slower but the evidence from the studies in this thesis does not suggest qualitatively different patterns of processing even when comparing monolinguals to L2 learners. The latter is not easily reconcilable with theories that assume an inaccessibility of features although stronger evidence for the use of number mismatch would be more unequivocally informative.

6.4.2. Theories of L2 processing

The results from this thesis are generally agnostic with regards to the role of the first language in processing in English. There are two reasons for this. The first is related to the first language of the participants. The majority of both adults and children spoke an additional language which had some morphological markers for the plural but also subject-verb agreement and wh-movement. This is related to both the fact that at least the former two appear in a majority of languages but also due to the fact that languages without such features are not spoken by large communities of speakers of English as an additional language in the UK. The second reason for this is related to length of exposure. Almost all participants in these studies had a substantial exposure to English within a naturalistic setting in excess of two years. The L2 learners have also more than two years exposure to English albeit non-naturalistically. This means that all groups have a high degree of proficiency and that their grammars are not initial state grammars. Therefore, to make claims about deficiencies in their grammar based on their L1 (e.g. Haznedar, 2001, Vainikka & Young-Scholten, 1993, 1996a, 1996b) or transfer in bulk of the L1 onto the L2 or second L1 as postulated under accounts such as the Full Access Full Transfer account (Schwartz & Sprouse, 1996) cannot be confidently substantiated for either adults or children. The data from the studies in this thesis are consistent with the notion that developmental milestones are similar in bilingual and monolingual children as has been assumed for simultaneous bilinguals (De Houwer, 1995). In other words, bilingual children processed language as their age-matched monolingual controls did. This thesis does not make this claim for language production but for processing and, by extension, comprehension. Processing is subject to developmental changes and both bilingual and monolingual children tested in this study appeared to be at the same or similar milestone. Moreover, it is assumed that qualitatively similar and equally as efficient mechanisms of processing, as established most clearly in the study on which-questions, require largely approximate linguistic systems or degree of linguistic proficiency between the two groups. The absence of a correlation between age of onset and accuracy in the study on which questions suggests that this is likewise the case for early sequential bilinguals - which has previously been debated (Meisel, 1999). The bilingual children in the two studies had similar accuracy as the monolinguals and similar processing speed. The predominant difference is the slower increase in looks towards the expected picture over time relative to the monolinguals. However, both groups of children experienced garden-path effects in both studies or had equally increased difficulty with sentences where garden-path effects are possible.

The results from these studies may also be discussed in terms of the Interface hypothesis (Sorace & Filliaci, 2006). The results from this thesis are not supportive of the Interface hypothesis. The Interface hypothesis makes predictions about ultimate attainment and has been phrased and tested in terms of off-line comprehension but more recently has been framed in terms of processing (Sorace, 2011). Previous work points to persistent difficulty with inflectional morphology (Lardiere, 1998, 2007; White 2003, 2009) The number (mis)match between subject and verb is a feature where syntax and morphology interface. However, it would be classified as an internal rather than an external interface Thus, it is possible under the Interface hypothesis that the L2 learners as well as the bilingual adults would not face difficulty with this linguistic feature. The bilinguals underutilised number mismatch in real time relative to monolinguals both in the case of adults and children. Garden-path sentences may be guided in their interpretation by the verb’s semantics (Hopp, 2015; Roberts & Felser, 2011) but the referential context, as it was manipulated in this thesis, is in itself an interface between syntax and an external interface, namely discourse. While the syntax-semantics interface is not typically an area of difficulty in advanced L2 learners (Dekydtspotter, Sprouse & Anderson, 1997; Dekydtspotter, Sprouse & Swanson, 2001; Dekydtspotter, Sprouse & Thyre, 2001; Slabakova, 2003, 2009), the syntax-pragmatics interface is predicted to be an area of difficulty under the Interpretability hypothesis. Yet contrary to the Interpretability hypothesis' predictions, the L2 learners do not differ in their (non)use of referential context to guide ambiguity resolution[38].

The results from this thesis are informative with regards to the Shallow Structure Hypothesis (Clahsen & Felser, 2006a, 2006b). This account fails to account for the findings in this thesis in two ways. First, its predictions are related to the age of onset of acquisition. In other words, it is an account that assumes a distinction between natives and non-native speakers of a language. This distinction, however, is not clear from the adult data in the studies carried out. The L2 learners showed broadly similar levels of accuracy as the monolinguals and did not underperform bilinguals with naturalistic exposure who would reasonably be classified as native speakers of the second language. On the other hand, on measures of on-line processing and reaction times, the native bilinguals patterned with the L2 learners on several measures and underperformed the monolinguals. In fact, in the majority of measures the significant effect of group was found in the monolinguals vs. bilinguals contrast and not in the native bilinguals vs. L2 learners or the monolinguals vs. non-native bilinguals contrast.

The second shortcoming of the Shallow Structure Hypothesis lies in the predictions about the nature of L2 processing which is assumed to be less reliant on structural cues but more so on cues from different sources. This is not substantiated in both studies with adults. In the study on which-questions, it may be reasonable to expect that the L2 learners will be more reliant on number match in order to process object which-questions. This was not found to be the case. Although the effects of number mismatch were small, there were no interaction of number match with group in any of the measures examined. This entails that the L2 learners were not more dependent on number (mis)match than the other two groups. The Shallow Structure Hypothesis does not predict differences between early bilinguals with naturalistic

[38] Although, the latter is harder to test as there are no strong effects of referential context in any of the measures.

exposure and monolinguals. The results from the study on garden-path sentences are also not particularly supportive of this proposal. It may be expected under the Shallow Structure Hypothesis that the L2 learners may rely more on the referential context than the other two groups. The data from the study on garden-path sentences for both adults and children do not show substantial evidence for facilitatory effect of referential context on ambiguity resolution on the any of the measures analysed. Most importantly, there is no evidence of the L2 learners using referential context in the expected manner, i.e. that the second referent aids processing in the ambiguous sentences only, to a greater or less degree than the other two groups.

The Shallow Structure Hypothesis does not predict differences between the monolingual and bilingual children as the latter have had early and naturalistic exposure to English and would not be classed as L2 learners. Differences in processing and reaction times are observed between the monolingual and bilingual children in the study on which-questions. Although these have not been interpreted as evidence for qualitatively different processing between monolingual and bilingual children, they are still not consistent with the Shallow Structure Hypothesis. The results from the study on garden-path sentences with children are not strongly informative about the theory in question as neither group appeared to utilise referential context which would have indicated a less purely structural parsing mechanism. It should be noted that the Shallow Structure Hypothesis does not imply inherently inferior processing in L2 learners relative to native speakers and has been reworded over time to more explicitly postulated differences in processing rather than a disadvantage for non-native speakers (see Clahsen & Felser, 2018). The findings from this thesis are not inconsistent with the Shallow Structure Hypothesis in that they fail to show a disadvantage in L2 processing relative to monolinguals. In fact, this thesis fairly consistently fails to show the predicted differences.

By extension, the findings from this thesis are inconsistent with theoretical accounts that assume a critical period and a need for early input for language learning (Birdsong, 1999; Johnson & Newport; 1989). Adults with late exposure to English (i.e. the L2 learners) did not show differences to bilingual adults with early exposure. On the other hand, bilingual children with early exposure to the second language do show differences when not expected to. The findings are also not in line with accounts which assume quality of exposure determines processing with more naturalistic exposure leading to nativelike patterns (e.g. Pliatsikas & Marinis, 2013c); adults with naturalistic exposure to the second language differ at points to monolinguals and perform similarly to L learners. The differences in the data for the children further support this as the bilinguals with early and naturalistic exposure[39] show differences in processing to the monolinguals.

The differences in processing between bilinguals and monolinguals, both adults and children, appear to be centred on processing speed. The bilinguals generally have slower reaction times to respond accurately and show a more protracted and less steep increase in looks towards an anticipated picture consistent with forming an interpretation. One way to tentatively account for this is on the basis of managing the two languages. It is assumed that during processing both languages will be active; therefore, one will need to be inhibited. As a result, processing may be slower. This would impact both groups of bilinguals irrespective of age and differentiate them from monolinguals. This has been shown so far predominantly for lexical access related to real word and non-word repetition or picture description (e.g. Gollan, Slattery, Goldenberg, Van Assche, Duyck, & Rayner, 2011; Martin, Costa, Dering, Hoshino, Wu, & Thierry, 2012; Roberts, Garcia, Desrochers, & Hernandez, 2002) but has also been

[39] We note that these children do not receive the same exposure as monolinguals; they differ in that they have quantitatively less exposure to English (as they are also exposed to the L1). However, the bilingual children in the studies in this thesis scored within age-appropriate norms on baseline tasks of language proficiency.

shown for sentence production (MacKay & Flege, 2004). It is not inconceivable that the same applies for online language processing. The need to balance activation from the two languages would be consistent with the findings in this thesis as it does not predict differences in accuracy which are not found between monolinguals and native bilinguals, or qualitative differences as established by interactions of group with other variables but would predict differences in reaction times and slower increases in looks to any given picture than for monolinguals at critical segments albeit in the same direction as monolinguals.

An alternative claim is that bilingual language processing is qualitatively similar to that of monolinguals but is more resource intensive (McDonald, 2006; Hopp, 2006, 2010). This entails that bilingual language processing begins to diverge from that of monolinguals under conditions where processing involves more cognitive demands (e.g. in terms of memory). This strand of accounts is explicitly related to sentence processing. The linguistic phenomena studied in this thesis potentially fall under the category of those structures which are cognitively demanding to process in that they involve non-canonical word order and/or ambiguity and subsequent need for reanalysis. The differences in speed or the reduced ability to utilise information in real time could readily be accounted for under less efficient processing in needing more effort. On the other hand, these accounts postulate that the difference lies between native and non-native speakers; they do not explicitly predict differences between monolinguals and bilinguals. In this sense, the findings from this thesis would require an amendment to these accounts in order to be more fully compatible with them. Crucially, however, this type of account makes explicit predictions about individual differences impacting processing, i.e. individuals with cognitive capacities, such as working memory, will perform similarly to monolinguals. The number of participants in this thesis renders modelling individual differences challenging and, as such, support for this type of account cannot be definitive.

6.5. Limitations and Future directions

The studies in this thesis provide new insights into morphosyntactic processing in bilinguals and, in particular children. There are two ways in which future work can build on this thesis to contribute further; expanding what is studied and who is studied – the latter being more of a limitation for this thesis. The first is to expand on the range of linguistic features and the nature of cues tested. This may be related to existing work available but also other linguistic features. Work on online processing of complex morphosyntactic syntactic structures where reanalysis is also involved, such as relative clauses or cleft sentences, in bilinguals would be welcome. Furthermore, there is a dearth of studies in bilingual processing in languages other than English which could examine grammatical cues not just based on subject-verb agreement, but where features such as nominal case could function as facilitators of disambiguation or processing in general. New studies have attempted to address this issue (e.g. Roesch & Chondrogianni, 2016) but have investigated the role of cues at the end stage and not in real time. Several studies have built on the experimental design of Trueswell et al. (1999) by manipulating the conditions under which non-linguistic cues may be used by children (Meroni & Crain, 2003; Weighall, 2008) but these have never been undertaken with bilingual children. Moreover, the impact of semantics in guiding sentence processing has been explored in monolingual (e.g. Kidd & Bavin; Snedecker & Trueswell, 2004) but not bilingual children. In addition to this, little is known about how (monolingual or bilingual) children with language impairment process language in real time and whether grammatical and/or non-grammatical cues can be utilised as a compensatory mechanism or whether they remain inaccessible.

The main limitation of this thesis is that it does not differentiate between subgroups of bilingual children. Research has shown that both bilingual children and adults differ in terms of their proficiency in either or both languages as well as in the quality and quantity of exposure/input resulting in different linguistic and overall cognitive abilities (e.g. Luk & Bialystok, 2013; De Cat, Gusnanto & Serratrice, 2018; Incera & Lennan, 2018; de Luca, Rothman, Bialystok & Pliatsikas, 2019). The studies in this thesis were not originally conceptualised as studies on individual differences; neither were the original studies upon which the work in this thesis was based upon. Moreover, large sample sizes are needed to model individual differences robustly as background measures need to be included, thus enlarging the number of predictor variables; the aforementioned studies broadly included between 60-120 participants. Recruiting more children for this thesis was not possible due to logistic and time-related constraints. Furthermore, as some of the manipulations in the studies, the expansion of the visual world paradigm to bilingual children and the use of time as a dependent variable is comparatively novel, it may be prudent to test the feasibility of the studies before endeavouring to explore individual differences. Examining differences in bilingual sentence processing in light of individual differences remains, nonetheless, a potentially fruitful and conceptually appropriate way forward. The question arising from this limitation concerns the potential impact on the results. The statistical analyses which allow for the random intercepts and slopes to vary by participant (among others) shrink the variance and reduce the impact of outliers. This entails that the differences between the two groups are most likely generic and observed across individual participants. It remains an open question as to the behaviour of specific subgroups of bilingual children for which there were too few participants to compare in these studies statistically. Given that the majority of children were exposed to English from birth and had similar age-appropriate proficiency scores with the monolingual controls and still showed slower processing, it is plausible that this would have been observed for bilingual children with less native-like levels of English proficiency. In fact, it is plausible that the differences in timecourse may even be more protracted. It is also possible that with bilingual children who substantially less proficient than their monolingual peers that one finds qualitative differences or differences observed at the end-state result (i.e. accuracy). Nevertheless, this would be confounded with the bilingual children's limited proficiency.

This step further in this direction for future research would be be valuable for an additional reason. It would namely be informative regarding accounts which postulate L2 processing to be similar to that of monolinguals but more taxing on cognitive resources. It would also fit into the debate about executive function in bilingual children which has shifted from assuming a blanket advantage to a more nuanced picture (Barker & Bialystok, 2019; Bialystok, 2009, 2015). There is existing research which has suggested a link between executive functioning and language processing on locally ambiguous structures (Navarro-Torres, Garcia, Chidambaram, & Kroll, 2019; Teubner-Rhodes, Mishler, Corbett, Andreu, Sanz-Torrent, Trueswell, Novick, 2016; Vuong & Martin, 2014). The proposed mechanism is that the need for reanalysis entails the need to suppress the original misinterpretation and switch to a new one. Consequently, better executive function abilities are expected to lead to better reanalysis. The current evidence for this is suggestive but still not definitive.

The second approach to further the work in this thesis is to expand on the participant groups. For the purposes of these studies, bilingual adults and children were recruited from various first languages. This gives rise to the following challenge; it is unclear what the role of the first language is on processing in English. One should recall that the impact of the L1 is subject to debate in the literature. This is a bigger problem for the study on which-questions. Most of the participants spoke a first language with plural morphology and/or subject verb-agreement. It is not clear whether they process which-questions effectively because their L1 is similar to English in this respect. We have attempted to address this issue in the statistical analysis by allowing random slopes and intercepts for participants. This reduces the impact of individuals and, in particular, outliers. Given that all L1s had only few participants each, both for adults and children, this indirectly reduces the impact of specific L1s on the results. Therefore, a possibility for future research would be to compare bilinguals whose L1 has wh-movement to bilinguals with a wh-in-situ L1 or alternatively bilinguals with an L1s with or without number agreement between subject and verb. This would then provide a more direct test for potential L1 effects. This was not undertaken in the current thesis for reasons of practical limitations; the majority of the world's languages do not lack some or all of the linguistic phenomena tested and those that do, do not have sizeable communities of speakers in the UK. As such, it is logistically easily feasible to recruit a large number of bilinguals with an L1 where these features are radically different to English. A further complication is that we tested bilinguals with a high degree of proficiency in English. It may be argued that the high degree of proficiency in all groups of bilinguals will lead to L2 processing which is qualitatively similar to those of monolinguals (e.g. Hopp, 2015; Pliatsikas & Marinis, 2013a) and that, in bilingual adults and children with a lower degree of proficiency, the results may be different. The studies showed that the bilinguals behaved similarly to monolinguals but still processed language slower. If even highly proficient bilinguals show these differences, it is unlikely that bilinguals with a lower degree of proficiency would not show these effects. In fact, testing bilinguals who are less proficient in English may be confounded by the fact that their limited proficiency hinders processing. In other words, it would be unclear whether any differences observed are due to bilingualism – either in adults or in children – or whether they can simply not do the task. Future work with bilinguals should also address the different types of bilingualism, particularly in the case of children. More specifically, future research on processing should address the distinction between simultaneous vs. early sequential bilingual children.

A second limitation is methodological and concerns only the studies on garden path sentences. The ambiguous sentences (and the subsequent comprehension questions) were in the past tense with the perfective aspect. This was done for a number of practical reasons outlined in the methods for Study 2A. Evidence from the literature suggests that speakers of a language use tense and aspect in real time predictively (Altman & Kamide, 2007). This means the participants may have heard a completed event in the past and may be expecting to see a picture of this event in one of the pictures (i.e. the object in the correct destination, in the example provided in the thesis, the apple in the bag). But the fact that they do not see the expected picture may be perceived as pragmatically odd. It is unclear what impact, if indeed any, this would have on the results and, in particular on the gaze data. It is plausible that it results in participants looking at all pictures equally for longer. Yet this does not appear to be impact the effects of ambiguity as they are observed in the gaze data. The effect of referential context is less consistent in the gaze data and thus, there is greater scope for a potential impact. It may be the case that participants simply override any issues with the pragmatics of the pictures as they are aware that they are doing an experimental task and that they need to select one picture to answer the comprehension question. Therefore, a potential future direction in this line of research would be to control for the tense and aspect of the sentences in the stimuli.

An additional limitation of the studies on garden path sentences is the number of trials per condition (four). While this is consistent with previous studies (e.g. Trueswell et al., 1999; Snedeker & Trueswell, 2004), it potentially raises questions about the likelihood of the experimental design to detect effects that may be present in the data, i.e. statistical power. This is done across studies for a practical reason which is not applicable to the experimental task on which-question processing; participants may become aware of the temporary ambiguity of the sentences and avoid garden-path effects by interpreting the ambiguous prepositional phrase as a modifier to the verb's object rather than as an adjunct to the verb expressing destination. For this reason, a comparatively larger number of filler trials is needed reducing the number of experimental trials that can be used in the task. This study used 20 fillers, broadly in line with the previous studies. Typically, in studies on garden path sentences, the fillers are equal to or may even be double the number of the experimental trials. Despite the study being potentially underpowered, effects of ambiguity are observed in both studies for both accuracy and the gaze data. Smaller effects of referential context are also observed but more sporadically and in an unpredicted direction. This indicates that lack of statistical power is unlikely to be a significant limitation in the experimental design.

It was mentioned in the section outlining the methods for the study on garden path sentences that E-prime did not record the picture the participants selected as an answer to the comprehension question other than whether it was the correct one or not. While accuracy was significantly lower for the ambiguous questions relative to the unambiguous ones across all groups of participants, it does not establish garden-path effects as such. Given the increased looks to the incorrect destination in the gaze data, it is plausible that participants selected the incorrect destination as the answer to the comprehension questions more often in the ambiguous sentences than in the unambiguous sentences. Further studies on garden-path sentences with different manipulations would benefit from adjusting the E-prime script so that this information is recorded.

A further option for future research in processing of garden-path sentences in bilingual children would be to manipulate intonational contours. The current study used sentences recorded at a slow pace with minor pauses between segments. However, research suggests that children as young as 4-6 years old can utilise intonational cues during ambiguity resolution (Snedeker & Yuan, 2008). Children interpreted ambiguous prepositional phrases as either modifying the verb (i.e. instrument) or the noun phrase. This effect was, nonetheless modulated by the lexical cues in the ambiguous sentence towards one interpretation relative to the alternative. This means that the intonational cues were most effective when the lexical cues also biased the parser towards a specific interpretation. However, the sentences in Snedeker & Yuan contained global ambiguity and the difference in pause duration (about 100ms) resulted in different intonational cues. It is unclear whether the intonational cues would be effective in the same manner in the context of the current study where the sentences were locally ambiguous and where a prepositional phrase is strongly preferred as a modifier of the verb "put" rather than a modifier to the object of the verb put irrespective of the duration of the pause or any change in rising/falling intonation.

6.6. Conclusions

In sum, this thesis reports four studies; two with bilingual children and two with bilingual adults, either native bilinguals or L2 learners. Two linguistic features were tested, which-questions, which are syntactically complex and locally ambiguous and garden-path sentences which are locally ambiguous. For the first phenomenon, we tested whether participants could use number mismatch as a disambiguating cue whereas for the second one we tested whether referential context (one or two referents in the visual stimuli) could function as such. The results from the four studies suggest the both bilingual adults and children were equally as able as their monolingual counterparts as an end stage. All groups showed evidence for incremental processing and hence misinterpreted ambiguous sentences but subsequently reanalysed them. However, in terms of speed of processing both bilingual adults and children differed to the monolinguals – being slower. In fact, the bilingual adults patterned with the L2 learners. This difference in speed was most pronounced for real time processing. Lastly, the bilingual groups showed a reduced use of cues to disambiguate in real time, consistent with previous work, but were able to do so at the end of the sentence. This suggests that bilinguals may differ to monolinguals to an extent in terms of processing of sentences but that these differences are observed on a very fine-grained timescale. We interpret these findings not as evidence of a deficiency but hypothesize that it can be attributed to the need to manage the two languages which are consistently active in real time.

APPENDICES

APPENDIX A

Parental Questionnaire for Monolingual Participants

1. General Information about the Child

1.1 Birth Date × ______________________________________

1.2 If place of birth is Not country of residence, date of arrival in country of residence × ______________________

2. Child's early history × Language, etc.

2.1 How old was your child when he/she spoke his/her first word?

____

2.2 How old was your child when he/she first put words together to make short sentences? ____________________

Example × *more water* ; *more milk* ; etc.

2.3 Before your child was three or four years old, were you ever concerned about his/her language? NO or YES

____

2.4 Has your child ever had any hearing problems or frequent ear infections? NO or YES

If Yes, please elaborate × ______________________________________

3. Information about the mother's and the father's education

3.1 Information about the mother's education

Education ×

		Number of years	Further information
Primary school	Yes / No		
Secondary school	Yes / No		
University	Yes / No		
Other professional training	Yes / No		

3.2 Information about the father's education

Education ×

		Number of years	Further information
Primary school	Yes / No		
Secondary school	Yes / No		
University	Yes / No		
Other professional training	Yes / No		

4. We would like to know if any members of the family have/had a history of difficulties in school or difficulties with any aspects of language.

In each cell, please indicate YES or NO ×

	Brother/ sister	Mother	Father	Father's family	Mother's family
Difficulties at school					
Difficulties mainly with reading and spelling					
Repeated one or more grades in school					
Difficulties understanding others when they speak					
Difficulties expressing oneself orally (pronunciation, forming sentences, finding the right word, etc.)					

APPENDIX B

Parental Questionnaire for Bilingual Participants

1. General Information about the Child

1.1 Birth Date: ____________________________________

1.2 If place of birth is Not country of residence, date of arrival in country of residence: _____________________

2. Child's early history: Language, etc.

2.1 How old was your child when he/she spoke his/her first word and in which language was his/her first word?

____

2.2 How old was your child when he/she first put words together to make short sentences and in which language was his/her first sentences? ___________________

Example: *more water*; *more milk*; etc.

2.3 Before your child was three or four years old, were you ever concerned about his/her language? NO or YES

____

2.4 Has your child ever had any hearing problems or frequent ear infections? NO or YES ____________________________________

2.5 What languages does your child speak Now?

English	Other (specify)

2.6 Which language do you think your child feels the most at home in? _______________________________

2.7 Before your child was four years old, how often did he/she hear each language?

	0 Never 0%	1 Rarely 25%	2 Sometimes 50%	3 Usually 75%	4 Always 100%
English					
Other (specify)					

2.8 In what contexts did he/she hear each language? (Check all appropriate cells.)

	English	**Other**
a. Conversations with mother		
b. Conversations with father		
c. Conversations with grand parents		
d. Conversations with babysitter / child minder		
e. Conversations with other adults (specify)		
f. Conversations with siblings		
g. Nursery school/day care center / kindergarten		

3. Current Skills

	English	Other
3.1 Compared to other children the same age, how do you think your child expresses him/herself in …? *0 = Not very well/Not as well as them; 1 = a little less well/a few differences; 2 = (generally) the same; 3 = very well, better*	0 1 2 3	0 1 2 3
3.2 Do you think that your child speaks like a child the same age who only speaks …..? *0 = Not very well/Not as well as them; 1 = a little less well/a few differences; 2 = (generally) the same; 3 = very well, better*	0 1 2 3	0 1 2 3
3.3 Compared to other children the same age, do you think your child has difficulties making correct sentences? *0 = Yes, many difficulties; 1 = some difficulties; 2 = (generally) the same; 3 = No difficulties, better than other children*	0 1 2 3	0 1 2 3

4. Languages used at home

4.1 With parents

	Mother ↔ Child					Father ↔ Child				
	0 Never 0%	1 Rarely 25%	2 Sometimes 50%	3 Usually 75%	4 Always 100%	0 Never 0%	1 Rarely 25%	2 Sometimes 50%	3 Usually 75%	4 Always 100%
English										
Other										

4.2 Is there another adult who regularly takes care of your child? (grandparent, babysitter, etc.) YES or No

Grandparent ↔ Child						Babysitter ↔ Child				
	0 Never 0%	1 Rarely 25%	2 Sometimes 50%	3 Usually 75%	4 Always 100%	0 Never 0%	1 Rarely 25%	2 Sometimes 50%	3 Usually 75%	4 Always 100%
English										
Other										
Other Adult 1 ↔ Child						Other Adult 2 ↔ Child				
	0 Never 0%	1 Rarely 25%	2 Sometimes 50%	3 Usually 75%	4 Always 100%	0 Never 0%	1 Rarely 25%	2 Sometimes 50%	3 Usually 75%	4 Always 100%
English										
Other										

5. Languages spoken outside the home

5.1 What language activities does your child do <u>each week</u> and in what language(s)?

	English			**/Other**		
Activities	0 Never or almost never	1 At least once a week	2 Every day	0 Never or almost never	1 At least once a week	2 Every day
a. Reading (books, magazines, comic books, newspapers)						
b. Television/ movies / cinema						
c. Storytelling						

5.2 What language is spoken between your child and the friends he/she plays with regularly?

Child ↔ Friends					
	0 Never 0%	1 Rarely 25%	2 Sometimes 50%	3 Usually 75%	4 Always 100%
English					
Other					

5.3 What language is spoken with family friends with whom you are in regular contact?

Family Friends					
	0 Never 0%	1 Rarely 25%	2 Sometimes 50%	3 Usually 75%	4 Always 100%
English					
Other (specify) _______________					

6. Information about the mother and the father

6.1 In which country were you born?

____

6.2 If you are currently working, what is the language you use at your work place?

____

6.3 Education ×

	Mother	Father	Number of years
Primary school	Yes / No	Yes / No	M: F:

	Mother	Father	Number of years
Secondary school	Yes / No	Yes / No	M: F:
University	Yes / No	Yes / No	M: F:
Other professional training	Yes / No	Yes / No	M: F:

6.1.4 In your opinion, how well do you speak the following languages?

	0 Only a few words	1 Gets along, but with difficulty	2 Basic abilities (gets along)	3 Well	4 Very well
Polish					
English					
Other					

7. Difficulties

In each cell, please indicate YES or NO:

	Brother/ sister	Mother	Father	Father's family	Mother's family
Difficulties at school					
Difficulties mainly with reading and spelling					
Repeated one or more grades in school					
Difficulties understanding others when they speak					
Difficulties expressing oneself orally (pronunciation, forming sentences, finding the right word, etc.)					

APPENDIX C

Experimental items for Studies 1A and 1B

Subject questions – Number match – Singular First NP

1. Which bear is chasing the camel?
2. Which dog is stroking the owl?
3. Which donkey is carrying the zebra?
4. Which duck is tickling the chicken?
5. Which goat is pushing the cow?
6. Which gorilla is kicking the horse?
7. Which lion is spraying the elephant?
8. Which monkey is licking the lamb?
9. Which rat is kissing the rabbit?
10. Which spider is splashing the squirrel?

Subject questions – Number mismatch – Singular First NP

1. Which bear is chasing the camels?
2. Which dog is stroking the owls?
3. Which donkey is carrying the zebras?
4. Which duck is tickling the chickens?
5. Which goat is pushing the cows?
6. Which gorilla is kicking the horse?
7. Which lion is spraying the elephants?
8. Which monkey is licking the lambs?
9. Which rat is kissing the rabbits?
10. Which spider is splashing the squirrels?

Subject questions – Number match – Plural First NP

1. Which bears are chasing the camels?
2. Which dogs are stroking the owls?
3. Which donkeys are carrying the zebras?
4. Which ducks are tickling the chickens?
5. Which goats are pushing the cows?
6. Which gorillas are kicking the horses?
7. Which lions are spraying the elephants?
8. Which monkeys are licking he lambs?
9. Which rats are kissing the rabbits?
10. Which spiders are splashing the squirrels?

Subject questions – Number mismatch – Plural First NP

1. Which bears are chasing the camel?
2. Which dogs are stroking the owl?
3. Which donkeys are carrying the zebra?
4. Which ducks are tickling the chicken
5. Which goats are pushing the cow?
6. Which gorillas are kicking the horse?
7. Which lions are spraying the elephant?
8. Which monkeys are licking the lamb?
9. Which rats are kissing the rabbit?
10. Which spiders are splashing the squirrel?

Object questions – Number match – Singular First NP

1. Which bear is the camel chasing?
2. Which dog is the owl stroking?
3. Which donkey is the zebra carrying?
4. Which duck is the chicken tickling?
5. Which goat is the cow pushing?
6. Which gorilla is the horse kicking?
7. Which lion is the elephant spraying?
8. Which monkey is the lamb licking?
9. Which rat is the rabbit kissing?
10. Which spider is the squirrel splashing?

Object questions – Number mismatch – Singular First NP

1. Which bear are the camels chasing?
2. Which dog are the owls stroking?
3. Which donkey are the zebras carrying?
4. Which duck are the chickens ticking?
5. Which goat are the cows pushing?
6. Which gorilla are the horses kicking?
7. Which lion are the elephants spraying?
8. Which monkey are the lambs licking?
9. Which rat are the rabbits kissing?
10. Which spider are the squirrels splashing?

Object questions – Number match – Plural First NP

1. Which bears are the camels chasing?
2. Which dogs are the owls stroking?
3. Which donkeys are the zebras carrying?
4. Which ducks are the chickens tickling?
5. Which goats are the cows pushing?
6. Which gorillas are the horses kicking?
7. Which lions are the elephants spraying?
8. Which monkeys are the lambs licking?
9. Which rats are the rabbits kissing?
10. Which spiders are the squirrels splashing?

Object questions – Number mismatch – Plural First NP

1. Which bears is the camel chasing?
2. Which dogs is the owl stroking?
3. Which donkeys is the zebra carrying?
4. Which ducks is the chicken ticking?
5. Which goats is the cow pushing?
6. Which gorillas is the horse kicking?
7. Which lions is the elephant spraying?
8. Which monkeys is the lamb licking?
9. Which rats is the rabbit kissing?
10. Which spiders is the squirrel splashing?

APPENDIX D

Experimental items for Studies 2A and 2B

Set of items 1: Apple, banana, plate, bag
(Un)ambiguous: Peter put the apple (that's) on the plate in the bag before going to school.
Question: What did Peter put in the bag?

Set of items 2: Book, lamp, desk, backpack
(Un)ambiguous: Steve put the book (that's) on the desk in the backpack before going home.
Question: What did Steve put in the backpack?

Set of items 3: Card, dice, table, drawer
(Un)ambiguous: David put the card (that's) on the table in the drawer earlier this morning.
Question: What did David put in the drawer?

Set of items 4: Money, key, chair, wallet
(Un)ambiguous: Matthew put the money (that's) on the chair in the wallet before going out.
Question: What did Matthew put in the wallet?

Set of items 5: Shoe, jumper, mat, wardrobe
(Un)ambiguous: Sarah put the shoe (that's) on the mat in the wardrobe after getting dressed.
Question: What did Sarah put in the wardrobe?

Set of items 6: Sandwich, cheese, napkin, fridge
(Un)ambiguous: Julia put the sandwich (that's) on the napkin in the fridge to keep for tomorrow.
Question: What did Julia put in the fridge?

Set of items 7: Plaster, sweet wrapper, bench, bin
(Un)ambiguous: Clare put the plaster (that's) on the bench in the bin while playing in the park.
Question: What did Clare put in the bin?

Set of items 8: Soap, comb, towel, bath
(Un)ambiguous: Mary put the soap (that's) on the towel in the bath before having a wash.
Question: What did Mary put in the bath?

Set of items 9: Burger, pear, napkin, bin
(Un)ambiguous: James put the burger (that's) on the napkin in the bin because it tasted bad.
Question: What did James put in the bin?

Set of items 10: Ball, kite, bench, backpack
(Un)ambiguous: Ben put the ball (that's) on the bench in the backpack after playing for hours.
Question: What did Ben put in the backpack?

Set of items 11: Coin, credit card, table, wallet
(Un)ambiguous: Tom put the coin (that's) on the table in the wallet while counting the money.
Question: What did Tom put in the wallet?

Set of items 12: Toothbrush, sponge, towel, bath
(Un)ambiguous: Robert put the toothbrush (that's) on the towel in the bath by accident today.
Question: What did Robert put in the bath?

Set of items 13: Sock, glove, mat, drawer
(Un)ambiguous: Katie put the sock that's on the mat in the drawer while tidying up.
Question: What did Katie put in the drawer?

Set of items 14: Hat, scarf, chair, wardrobe
(Un)ambiguous: Emily put the hat (that's) on the chair in the wardrobe while cleaning the bedroom.
Question: What did Emily put in the wardrobe?

Set of items 15: Cake, egg, plate, fridge
(Un)ambiguous: Jenny put the cake (that's) on the plate in the fridge to finish later.
Question: What did Jenny put in the fridge?

Set of items 16: Bread, chocolate, bag, desk
(Un)ambiguous: Susan put the bread (that's) on the desk in the bag to take home for later.
Question: What did Susan put in the bag

APPENDIX E

Models for Study 1A

Fixed effects for accuracy and reaction times

	Accuracy				Reaction times			
Dependent variable	*β*	SE	*z*	*p*	*β*	SE	*z*	*p*
Group	-0.01	0.16	-0.04	0.972	182.71	152.32	1.20	0.235
Structure	-1.07	0.13	-7.93	<0.001	245.50	33.05	7.43	<0.001
Number	0.05	0.10	0.46	0.644	-96.39	33.10	-2.91	0.004
FirstNP	-0.18	0.10	-1.76	0.079	27.59	32.89	0.84	0.402
Group × Structure	-0.03	0.12	-0.24	0.810	53.38	32.94	1.62	0.105
Group × Number	-0.01	0.08	-0.12	0.906	-41.28	32.79	-1.26	0.208
Structure × Number	0.16	0.11	1.43	0.154	24.16	32.87	0.74	0.462
Group × FirstNP	0.11	0.10	1.13	0.258	35.98	32.80	1.10	0.273
Structure × FirstNP	0.16	0.08	1.89	0.058	7.36	32.87	0.22	0.823
Number × FirstNP	-0.02	0.08	-0.19	0.853	-87.23	32.90	-2.65	0.008
Group × Structure × Number	0.00	0.08	0.00	0.999	19.89	32.79	0.61	0.544
Group × Structure × FirstNP	-0.09	0.08	-1.09	0.276	10.94	32.80	0.33	0.739
Group × Number × FirstNP	0.06	0.08	0.67	0.503	-72.46	32.79	-2.21	0.027
Structure × Number × FirstNP	0.02	0.08	0.24	0.812	32.45	32.86	0.99	0.323
Group × Structure × Number × FirstNP	-0.03	0.08	-0.36	0.717	3.12	32.80	0.10	0.924

R script

Accuracy: StimSlide.ACC ~ Group * Structure * Number * FirstNP + (1 | Subject) + (0 + Structure|Subject) + (0 + FirstNP|Subject) + (1|item) + (0 + Structure ||item) + (0 + Number | item) + (0 + FirstNP|item) + (0 + Group × Structure |item) + (0 + Structure × Number | item) + (0 + Group × FirstNP|item)

Reaction times: StimSlide.RT ~ Group*Structure*Number*FirstNP + (1|Subject) + (1|item)

Fixed effects for gaze data – looks to the correct picture

Dependent variable	β	SE	t	p
Time^1	26.38	0.99	26.63	< 0.001
Time^2	0.15	0.98	0.16	0.876
Time^3	-3.29	0.99	-3.33	0.001
Time^4	1.89	0.99	1.92	0.055
Group	-0.03	0.03	-0.95	0.347
Structure	-0.16	0.02	-8.92	<0.0001
Number	0.00	0.02	0.03	0.979
FirstNP	0.00	0.01	-0.34	0.739
Time^1 ×Group	-7.10	0.99	-7.17	<0.0001
Time^2 ×Group	1.72	0.98	1.75	0.08
Time^3 ×Group	0.15	0.99	0.15	0.883
Time^4 ×Group	-1.95	0.99	-1.98	0.048
Time^1 ×Structure	7.33	0.99	7.40	<0.001
Time^2 ×Structure	2.34	0.98	2.38	0.017
Time^3 ×Structure	-2.53	0.99	-2.57	0.010
Time^4 × Structure	3.40	0.99	3.45	0.001
Group × Structure	0.01	0.02	0.36	0.723
Time^1 × Number	-1.05	0.99	-1.06	0.290
Time^2 × Number	-0.93	0.98	-0.94	0.346
Time^3 × Number	-1.66	0.99	-1.68	0.092
Time^4 × Number	1.12	0.99	1.13	0.258
Group × Number	-0.02	0.02	-1.41	0.181
Structure × Number	0.02	0.01	1.62	0.118
Time^1 × FirstNP	0.73	0.99	0.74	0.458
Time^2 × FirstNP	-0.15	0.98	-0.15	0.882
Time^3 × FirstNP	-0.57	0.99	-0.58	0.561
Time^4 × FirstNP	1.47	0.99	1.49	0.137
Group × FirstNP	0.01	0.01	0.43	0.668
Structure × FirstNP	0.00	0.02	0.18	0.858
Number × FirstNP	0.01	0.02	0.86	0.405
Time^1 × Group × Structure	0.37	0.99	0.38	0.706
Time^2 × Group × Structure	-0.39	0.98	-0.40	0.689
Time^3 × Group × Structure	0.13	0.99	0.14	0.893
Time^4 × Group × Structure	0.96	0.99	0.98	0.329
Time^1 × Group × Number	-1.80	0.99	-1.82	0.069
Time^2 × Group × Number	1.94	0.98	1.97	0.049
Time^3 × Group × Number	0.79	0.99	0.81	0.421
Time^4 × Group × Number	-0.20	0.99	-0.20	0.839
Time^1 × Structure × Number	-1.26	0.99	-1.28	0.201
Time^2 × Structure ×Number	2.06	0.98	2.10	0.036
Time^3 × Structure × Number	-0.81	0.99	-0.82	0.415
Time^4 × Structure ×Number	0.24	0.99	0.24	0.810
Group ×Structure ×Number	-0.02	0.01	-1.66	0.102
Time^1 ×Group ×FirstNP	-1.04	0.99	-1.05	0.292
Time^2 ×Group ×FirstNP	1.31	0.98	1.33	0.182
Time^3 ×Group ×FirstNP	0.08	0.99	0.08	0.938
Time^4 ×Group ×FirstNP	1.14	0.98	1.16	0.247
Time^1 ×Structure ×FirstNP	-0.09	0.99	-0.10	0.924
Time^2 ×Structure ×FirstNP	0.83	0.98	0.85	0.396
Time^3 ×Structure ×FirstNP	1.1	0.99	1.12	0.265
Time^4 ×Structure ×FirstNP	-1.37	0.99	-1.39	0.165
Group ×Structure ×FirstNP	-0.01	0.01	-0.95	0.349
Time^1 ×Number ×FirstNP	1.61	0.99	1.63	0.104
Time^2 ×Number ×FirstNP	-0.5	0.98	-0.51	0.609
Time^3 × Number × FirstNP	-2.27	0.99	-2.30	0.022
Time^4 × Number × FirstNP	-1.95	0.99	-1.97	0.048
Group × Number × FirstNP	0.00	0.01	-0.13	0.899
Structure × Number × FirstNP	-0.01	0.02	-0.95	0.356
Time^1 × Group × Structure × Number	-0.68	0.99	-0.69	0.493
Time^2 × Group × Structure × Number	-1.04	0.98	-1.06	0.289
Time^3 × Group × Structure × Number	1.04	0.99	1.05	0.294
Time^4 × Group × Structure × Number	-0.18	0.99	-0.19	0.852
Time^1 × Group × Structure × FirstNP	1.4	0.99	1.41	0.158
Time^2 × Group × Structure × FirstNP	0.42	0.98	0.43	0.669
Time^3 × Group × Structure × FirstNP	-0.14	0.99	-0.14	0.890
Time^4 × Group × Structure × FirstNP	-0.46	0.99	-0.46	0.642
Time^1 × Group × Number × FirstNP	1.25	0.99	1.27	0.206
Time^2 × Group × Number × FirstNP	-0.97	0.98	-0.99	0.324
Time^3 × Group × Number × FirstNP	-1.6	0.99	-1.62	0.105
Time^4 × Group × Number × FirstNP	1.10	0.99	1.11	0.265
Time^1 × Structure × Number × FirstNP	3.23	0.99	3.26	0.001
Time^2 × Structure × Number × FirstNP	0.08	0.98	0.08	0.933
Time^3 × Structure × Number × FirstNP	1.17	0.99	1.19	0.235
Time^4 × Structure × Number × FirstNP	-1.46	0.99	-1.48	0.140
Group × Structure × Number × FirstNP	0.01	0.01	0.55	0.594
Time^1×Group×Structure×Number×FirstNP	1.47	0.99	1.49	0.137
Time^2 ×Group ×Structure×Number×FirstNP	-0.93	0.98	-0.94	0.346
Time^3×Group×Structure×Number×FirstNP	0.75	0.99	0.76	0.448
Time^4×Group×Structure×Number×FirstNP	26.38	0.99	26.63	< 0.001

R script

elog ~ poly(Bin, 4) * Group * Structure * Number * FirstNP + (1 + Structure * Number * FirstNP||Subject) + (1 |item) + (0 + Structure | item) + (0 + Group × Number |item) + (0 + Group × Structure|item) + (0 + Number |item) + (0 + Structure × FirstNP | item) + (0 + Group × Number × FirstNP|item) + (0 + Structure × Number × FirstNP|item) + (0 + Group × Structure × Number × FirstNP | item) + (0 + Structure × Number|item) + (0 + Number × FirstNP |item) + (0 + Group × Structure × FirstNP|item) + (0 + Group | item) + (1 | Trial_ID)

APPENDIX F

Models for Study 1AB

Fixed effects for accuracy

	Contrast 1 ML vs. BL then BL-NS vs. BL-NNS				Contrast 2 NS vs. NNS then ML-NS vs. BL-NS			
Dependent variable	*β*	SE	*z*	*p*	*β*	SE	*z*	*p*
Group1	-0.25	0.19	-1.31	0.191	-0.44	0.19	-2.34	0.020
Group2	-0.32	0.17	-1.89	0.059	-0.03	0.17	-0.18	0.859
Structure	-0.50	0.11	-4.63	<0.001	-0.50	0.11	-4.63	<0.001
Number	-0.02	0.08	-0.21	0.836	-0.01	0.08	-0.14	0.889
FirstNP	-0.12	0.08	-1.52	0.128	-0.11	0.13	-0.79	0.430
Group1 × Structure	-0.26	0.15	-1.69	0.092	-0.19	0.15	-1.28	0.202
Group2 × Structure	-0.06	0.13	-0.44	0.658	-0.16	0.14	-1.20	0.229
Group1 × Number	-0.19	0.12	-1.62	0.106	-0.14	0.11	-1.35	0.177
Group2 × Number	-0.05	0.10	-0.48	0.632	-0.12	0.11	-1.13	0.259
Structure × Number	0.21	0.08	2.58	0.010	0.21	0.08	2.57	0.010
Group1 × FirstNP	-0.08	0.12	-0.69	0.489	-0.02	0.11	-0.18	0.854
Group2 × FirstNP	0.03	0.10	0.27	0.785	-0.08	0.11	-0.71	0.478
Structure × FirstNP	0.02	0.08	0.28	0.784	0.01	0.08	0.18	0.861
Number × FirstNP	-0.10	0.08	-1.18	0.240	-0.09	0.08	-1.16	0.246
Group1 × Structure × Number	-0.12	0.12	-1.02	0.310	-0.01	0.11	-0.09	0.927
Group2 × Structure × Number	0.05	0.10	0.54	0.589	-0.12	0.11	-1.08	0.282
Group1 × Structure × FirstNP	0.12	0.12	0.99	0.323	0.03	0.11	0.30	0.765
Group2 × Structure × FirstNP	-0.03	0.10	-0.29	0.774	0.10	0.11	0.93	0.354
Group1 × Number × FirstNP	0.11	0.12	0.97	0.334	0.11	0.11	1.06	0.289
Group2 × Number × FirstNP	0.05	0.10	0.57	0.569	0.06	0.11	0.54	0.591
Structure × Number × FirstNP	-0.05	0.08	-0.67	0.505	-0.06	0.08	-0.73	0.464
Group1 × Structure × Number × FirstNP	0.00	0.12	0.00	0.999	-0.12	0.11	-1.12	0.262
Group2 × Structure × Number × FirstNP	-0.12	0.10	-1.24	0.214	0.06	0.11	0.54	0.590

R script

Contrast 1: StimSlide.ACC ~ Group * Structure * Number * FirstNP + (1 | Subject) + (0 + Structure | Subject) + (1| item)

Contrast 2: StimSlide.ACC ~ Group * Structure * Number * FirstNP + (1 | Subject) + (1 | item) + (0 + Structure | Subject) + (0 + FirstNP| item)

Fixed effects for reaction times

	Contrast 1 ML vs. BL then BL-NS vs. BL-NNS				Contrast 2 NS vs. NNS then ML-NS vs. BL-NS			
Dependent variable	*β*	SE	*z*	*p*	*β*	SE	*z*	*p*
Group1	292.58	96.61	3.03	0.003	170.65	85.43	2.00	0.047
Group2	46.47	71.15	0.65	0.514	205.60	80.64	2.55	0.012
Structure	90.75	45.25	2.01	0.067	74.50	20.92	3.56	0.001
Number	-13.18	12.11	-1.09	0.276	-17.87	47.52	-0.38	0.715
FirstNP	57.52	11.96	4.81	0.000	73.29	45.33	1.62	0.140
Group1 × Structure	51.04	28.90	1.77	0.081	56.57	29.54	1.92	0.058
Group2 × Structure	30.30	25.89	1.17	0.245	19.83	25.01	0.79	0.430
Group1 × Number	7.54	16.29	0.46	0.643	8.43	20.90	0.40	0.695
Group2 × Number	4.67	14.98	0.31	0.755	2.76	15.21	0.18	0.857
Structure × Number	11.66	12.10	0.96	0.335	-3.90	11.91	-0.33	0.744
Group1 × FirstNP	17.52	16.29	1.08	0.282	-12.02	24.24	-0.50	0.630
Group2 × FirstNP	-22.14	14.98	-1.48	0.139	27.23	27.64	0.99	0.349
Structure × FirstNP	20.65	11.96	1.73	0.084	24.46	11.90	2.06	0.040
Number × FirstNP	-45.40	11.96	-3.80	0.000	-30.23	12.11	-2.50	0.013
Group1 × Structure × Number	10.08	16.29	0.62	0.536	170.65	85.43	2.00	0.047
Group2 × Structure × Number	-10.07	14.98	-0.67	0.501	205.60	80.64	2.55	0.012
Group1 × Structure × FirstNP	8.41	16.29	0.52	0.606	74.50	20.92	3.56	0.001
Group2 × Structure × FirstNP	-15.94	14.98	-1.06	0.287	-17.87	47.52	-0.38	0.715
Group1 × Number × FirstNP	-20.82	16.29	-1.28	0.201	73.29	45.33	1.62	0.140
Group2 × Number × FirstNP	-15.81	14.97	-1.06	0.291	56.57	29.54	1.92	0.058
Structure × Number × FirstNP	5.47	11.96	0.46	0.647	19.83	25.01	0.79	0.430
Group1×Structure×Number×FirstNP	-16.40	16.29	-1.01	0.314	8.43	20.90	0.40	0.695
Group2×Structure×Number×FirstNP	2.53	14.97	0.17	0.866	2.76	15.21	0.18	0.857

R script

Contrast 1: StimSlide.RT ~ Group*Structure*Number*FirstNP + (1|Subject) + (0+Structure|Subject) + (0+Structure|item) + (1|item)

Contrast 2: StimSlide.RT ~ Group*Structure*Number*FirstNP + (1|Subject) + (1|item) + (0+Structure|Subject)

Fixed effects for the gaze data – looks to the target picture

	Contrast 1 ML vs. BL then BL-NS vs. BL-NNS				Contrast 2 NS vs. NNS then ML-NS vs. BL-NS			
Dependent variable	*β*	SE	t	*p*	*β*	SE	t	*p*
Group1	-0.03	0.03	-1.02	0.310	0.00	0.03	0.06	0.949
Group2	0.01	0.02	0.61	0.542	-0.03	0.02	-1.11	0.270
Structure	-0.15	0.02	-9.12	0.000	-0.15	0.02	-9.12	0.000
Number	-0.02	0.02	-0.89	0.395	-0.02	0.02	-0.89	0.395
FirstNP	-0.01	0.02	-0.82	0.429	-0.01	0.02	-0.82	0.429
Time1	42.90	1.12	38.41	0.000	42.90	1.12	38.41	0.000
Time2	-25.24	1.12	-22.58	0.000	-25.24	1.12	-22.58	0.000
Time3	-8.66	1.12	-7.77	0.000	-8.66	1.12	-7.77	0.000
Time4	1.86	1.12	1.67	0.096	1.86	1.12	1.67	0.096
Group1 × Structure	-0.02	0.01	-1.46	0.148	0.00	0.01	-0.37	0.713
Group2 × Structure	0.00	0.01	0.39	0.697	-0.01	0.01	-1.35	0.181
Group1 × Number	0.01	0.01	0.81	0.418	0.01	0.01	0.69	0.491
Group2 × Number	0.00	0.01	0.30	0.769	0.00	0.01	0.50	0.620
Structure × Number	0.00	0.02	-0.17	0.871	0.00	0.02	-0.17	0.871
Group1 × FirstNP	0.00	0.01	-0.13	0.899	-0.01	0.01	-0.95	0.345
Group2 × FirstNP	-0.01	0.01	-0.93	0.354	0.00	0.01	0.36	0.720
Structure × FirstNP	-0.01	0.01	-0.94	0.367	-0.01	0.01	-0.94	0.367
Number × FirstNP	-0.02	0.01	-1.56	0.138	-0.02	0.01	-1.56	0.138
Group1 × Time1	3.47	1.51	2.29	0.022	2.64	1.50	1.76	0.079
Group2 × Time1	0.90	1.42	0.64	0.525	2.15	1.43	1.50	0.133
Group1 × Time2	7.36	1.52	4.85	0.000	0.50	1.51	0.33	0.742
Group2 × Time2	-3.18	1.42	-2.24	0.025	7.11	1.43	4.97	0.000
Group1 × Time3	0.39	1.51	0.26	0.794	-0.35	1.50	-0.24	0.814
Group2 × Time3	-0.55	1.42	-0.39	0.697	0.57	1.43	0.40	0.689
Group1 × Time4	2.20	1.51	1.45	0.147	2.96	1.50	1.97	0.049
Group2 × Time4	1.86	1.42	1.31	0.191	0.72	1.43	0.50	0.617
Structure × Time1	24.96	1.12	22.35	0.000	24.96	1.12	22.35	0.000
Structure × Time2	9.19	1.12	8.23	0.000	9.19	1.12	8.23	0.000
Structure × Time3	0.92	1.12	0.82	0.410	0.92	1.12	0.82	0.410
Structure × Time4	0.15	1.12	0.13	0.895	0.15	1.12	0.13	0.895
Number × Time1	-7.50	1.12	-6.71	0.000	-7.50	1.12	-6.71	0.000
Number × Time2	-0.55	1.12	-0.50	0.620	-0.55	1.12	-0.50	0.620
Number × Time3	0.16	1.12	0.14	0.887	0.16	1.12	0.14	0.887
Number × Time4	1.02	1.12	0.91	0.364	1.02	1.12	0.91	0.364
FirstNP × Time1	-0.12	1.12	-0.10	0.918	-0.12	1.12	-0.10	0.918
FirstNP × Time2	1.27	1.12	1.13	0.257	1.27	1.12	1.13	0.257
FirstNP × Time3	-1.48	1.12	-1.33	0.184	-1.48	1.12	-1.33	0.184
FirstNP × Time4	1.68	1.12	1.50	0.134	1.68	1.12	1.50	0.134
Group1 × Structure × Number	-0.01	0.01	-0.69	0.493	-0.01	0.01	-0.67	0.503
Group2 × Structure × Number	0.00	0.01	-0.34	0.734	0.00	0.01	-0.38	0.708
Group1 × Structure × FirstNP	0.01	0.01	1.60	0.113	0.02	0.01	1.75	0.083
Group2 × Structure × FirstNP	0.01	0.01	0.99	0.325	0.01	0.01	0.78	0.440
Group1 × Number × FirstNP	0.00	0.01	-0.19	0.849	-0.01	0.01	-1.20	0.234
Group2 × Number × FirstNP	-0.01	0.01	-1.16	0.249	0.00	0.01	0.42	0.675
Structure × Number × FirstNP	0.01	0.02	0.60	0.563	0.01	0.02	0.60	0.563
Group1 × Structure × Time1	-4.43	1.51	-2.93	0.003	0.52	1.50	0.35	0.726
Group2 × Structure × Time1	2.74	1.42	1.93	0.054	-4.69	1.43	-3.28	0.001
Group1 × Structure × Time2	-4.88	1.52	-3.22	0.001	0.63	1.50	0.42	0.675
Group2 × Structure × Time2	3.07	1.42	2.16	0.030	-5.20	1.43	-3.63	0.000
Group1 × Structure × Time3	-1.09	1.52	-0.72	0.473	-2.87	1.50	-1.91	0.056
Group2 × Structure × Time3	-2.33	1.42	-1.65	0.100	0.35	1.43	0.24	0.807
Group1 × Structure × Time4	3.58	1.51	2.37	0.018	1.83	1.50	1.22	0.223
Group2 × Structure × Time4	0.04	1.42	0.03	0.979	2.67	1.43	1.86	0.063
Group1 × Number × Time1	-1.08	1.51	-0.72	0.475	-2.38	1.50	-1.59	0.113
Group2 × Number × Time1	-1.84	1.42	-1.29	0.196	0.11	1.43	0.08	0.940
Group1 × Number × Time2	0.25	1.52	0.17	0.868	1.43	1.51	0.95	0.342
Group2 × Number × Time2	1.31	1.42	0.92	0.358	-0.46	1.43	-0.32	0.746
Group1 × Number × Time3	-1.97	1.52	-1.30	0.194	-3.31	1.50	-2.21	0.027
Group2 × Number × Time3	-2.32	1.42	-1.64	0.101	-0.31	1.43	-0.22	0.825
Group1 × Number × Time4	-2.01	1.51	-1.33	0.184	0.64	1.50	0.43	0.670
Group2 × Number × Time4	1.65	1.42	1.16	0.248	-2.33	1.44	-1.63	0.104
Structure × Number × Time1	-0.73	1.12	-0.66	0.511	-0.73	1.12	-0.66	0.511
Structure × Number × Time2	0.93	1.12	0.83	0.405	0.93	1.12	0.83	0.405
Structure × Number × Time3	-3.27	1.12	-2.93	0.003	-3.27	1.12	-2.93	0.003
Structure × Number × Time4	-2.94	1.12	-2.62	0.009	-2.94	1.12	-2.62	0.009
Group1 × FirstNP × Time1	1.53	1.51	1.02	0.310	6.80	1.50	4.54	0.000
Group2 × FirstNP × Time1	6.03	1.42	4.25	0.000	-1.87	1.43	-1.31	0.192

Group1 × FirstNP × Time2	2.53	1.52	1.67	0.095	0.06	1.51	0.04	0.967
Group2 × FirstNP × Time2	-1.20	1.42	-0.85	0.396	2.50	1.43	1.75	0.080
Group1 × FirstNP × Time3	0.01	1.52	0.01	0.996	-1.35	1.50	-0.90	0.370
Group2 × FirstNP × Time3	-1.35	1.42	-0.95	0.341	0.68	1.43	0.48	0.634
Group1 × FirstNP × Time4	-2.59	1.51	-1.71	0.087	-2.87	1.50	-1.91	0.056
Group2 × FirstNP × Time4	-1.58	1.42	-1.11	0.268	-1.16	1.43	-0.81	0.421
Structure × FirstNP × Time1	1.32	1.12	1.18	0.237	1.32	1.12	1.18	0.237
Structure × FirstNP × Time2	0.41	1.12	0.36	0.716	0.41	1.12	0.36	0.716
Structure × FirstNP × Time3	-2.03	1.12	-1.82	0.069	-2.03	1.12	-1.82	0.069
Structure × FirstNP × Time4	-0.80	1.12	-0.71	0.476	-0.80	1.12	-0.71	0.476
Number × FirstNP × Time1	-8.54	1.12	-7.66	0.000	-8.54	1.12	-7.66	0.000
Number × FirstNP × Time2	-1.74	1.12	-1.56	0.120	-1.74	1.12	-1.56	0.120
Number × FirstNP × Time3	2.83	1.12	2.54	0.011	2.83	1.12	2.54	0.011
Number × FirstNP × Time4	1.97	1.12	1.76	0.078	1.97	1.12	1.76	0.078
Group1 × Structure × Number × FirstNP	0.01	0.01	0.62	0.540	0.01	0.01	1.16	0.249
Group2 × Structure × Number × FirstNP	0.01	0.01	0.89	0.374	0.00	0.01	0.05	0.964
Group1 × Structure × Number × Time1	0.99	1.51	0.65	0.513	-1.89	1.50	-1.26	0.207
Group2 × Structure × Number × Time1	-2.38	1.42	-1.68	0.093	1.93	1.43	1.35	0.176
Group1 × Structure × Number × Time2	-1.78	1.52	-1.17	0.241	0.29	1.51	0.20	0.845
Group2 × Structure × Number × Time2	1.18	1.42	0.83	0.405	-1.93	1.43	-1.35	0.178
Group1 × Structure × Number × Time3	-1.59	1.51	-1.05	0.295	4.16	1.50	2.77	0.006
Group2 × Structure × Number × Time3	4.96	1.42	3.50	0.000	-3.67	1.43	-2.57	0.010
Group1 × Structure × Number × Time4	-0.27	1.51	-0.18	0.860	1.87	1.50	1.24	0.214
Group2 × Structure × Number × Time4	2.00	1.42	1.40	0.160	-1.20	1.43	-0.84	0.403
Group1 × Structure × FirstNP × Time1	-0.79	1.51	-0.52	0.601	-0.93	1.50	-0.62	0.534
Group2 × Structure × FirstNP × Time1	-0.54	1.42	-0.38	0.705	-0.33	1.43	-0.23	0.820
Group1 × Structure × FirstNP × Time2	-0.15	1.52	-0.10	0.922	1.31	1.51	0.87	0.386
Group2 × Structure × FirstNP × Time2	1.38	1.42	0.97	0.331	-0.80	1.43	-0.56	0.575
Group1 × Structure × FirstNP × Time3	-0.13	1.52	-0.09	0.932	0.49	1.50	0.33	0.742
Group2 × Structure × FirstNP × Time3	0.56	1.42	0.39	0.693	-0.38	1.43	-0.26	0.792
Group1 × Structure × FirstNP × Time4	2.72	1.52	1.80	0.072	2.25	1.50	1.50	0.135
Group2 × Structure × FirstNP × Time4	0.88	1.42	0.62	0.535	1.60	1.44	1.12	0.265
Group1 × Number × FirstNP × Time1	-3.02	1.51	-2.00	0.045	-3.49	1.50	-2.33	0.020
Group2 × Number × FirstNP × Time1	-1.98	1.42	-1.40	0.162	-1.28	1.43	-0.89	0.372
Group1 × Number × FirstNP × Time2	0.66	1.52	0.43	0.665	-0.06	1.50	-0.04	0.967
Group2 × Number × FirstNP × Time2	-0.39	1.42	-0.28	0.783	0.69	1.43	0.48	0.630
Group1 × Number × FirstNP × Time3	-1.24	1.52	-0.82	0.411	1.25	1.50	0.84	0.404
Group2 × Number × FirstNP × Time3	1.88	1.42	1.33	0.185	-1.87	1.43	-1.31	0.190
Group1 × Number × FirstNP × Time4	0.40	1.51	0.26	0.793	0.59	1.50	0.39	0.696
Group2 × Number × FirstNP × Time4	0.39	1.42	0.27	0.786	0.10	1.43	0.07	0.942
Structure × Number × FirstNP × Time1	4.28	1.12	3.84	0.000	4.28	1.12	3.84	0.000
Structure × Number × FirstNP × Time2	1.04	1.12	0.93	0.352	1.04	1.12	0.93	0.352
Structure × Number × FirstNP × Time3	-1.69	1.12	-1.52	0.129	-1.69	1.12	-1.52	0.129
Structure × Number × FirstNP × Time4	-2.24	1.12	-2.00	0.045	-2.24	1.12	-2.00	0.045
Group1×Structure×Number×FirstNP× Time1	-0.99	1.51	-0.66	0.512	0.21	1.50	0.14	0.890
Group2×Structure×Number×FirstNP× Time1	0.70	1.42	0.50	0.621	-1.09	1.43	-0.77	0.444
Group1×Structure Number×FirstNP×Time2	-0.25	1.52	-0.17	0.869	1.34	1.51	0.89	0.372
Group2×Structure×Number×FirstNP×Time2	1.47	1.42	1.03	0.301	-0.92	1.43	-0.65	0.519
Group1×Structure×Number×FirstNP×Time3	1.60	1.52	1.06	0.290	-0.09	1.50	-0.06	0.954
Group2×Structure×Number×FirstNP×Time3	-0.89	1.42	-0.63	0.530	1.65	1.43	1.16	0.248
Group1×Structure×Number× FirstNP×Time4	3.14	1.52	2.07	0.039	2.74	1.50	1.82	0.068
Group2×Structure×Number×FirstNP×Time4	1.17	1.43	0.82	0.411	1.77	1.44	1.23	0.219

R script

Contrast 1: lmer(elog ~ Group*Structure*Number*FirstNP*poly(Time, 4) + (1+Structure*Number*FirstNP||Subject) + (1+Structure*Number*FirstNP||item) +(1|Trial_ID), weights=1/wts, control = lmerControl(optimizer="bobyqa"),

Contrast 2: elog ~ Group*Structure*Number*FirstNP*poly(Time, 4) + (1+Structure*Number*FirstNP||Subject) + (1+Structure*Number*FirstNP||item) +(1|Trial_ID), weights=1/wts, control = lmerControl(optimizer="bobyqa")

APPENDIX G

Models for Study 1B

Fixed effects for accuracy and reaction times

	Accuracy				Reaction times			
Dependent variable	*β*	SE	*z*	*p*	*β*	SE	*t*	*p*
Group	-0.02	0.24	-0.08	0.934	-5.21	116.39	-0.05	0.964
Ambiguity	0.79	0.11	7.11	<0.001	-63.98	52.48	-1.22	0.223
Context	-0.05	0.10	-0.47	0.636	107.48	58.94	1.82	0.074
Group × Ambiguity	-0.08	0.13	-0.65	0.518	56.37	72.85	0.77	0.451
Group × Context	-0.05	0.10	-0.55	0.583	71.71	58.88	1.22	0.229
Ambiguity × Context	0.00	0.10	0.00	0.998	-46.47	52.15	-0.89	0.373
Group × Ambiguity × Context	0.05	0.10	0.50	0.617	18.12	61.93	0.29	0.774

R script

Accuracy: StimSlide.ACC ~ Group * Ambiguity * Context + (1 | Subject) + (0 + Ambiguity | Subject) + (1 | item) + (0 + Group | item) + (0 + Group × Ambiguity | item)

Reaction times: ReactonTimes ~ Group * Ambiguity * Context + (1 | Subject) + (0 + Context | Subject) + (1 | item) + (0 + Group | item) + (0 + Group × Ambiguity | item) + (0 + Group × Ambiguity × Context || item) + (1 | item_ambiguity), data = subset(Children.edat.data.2, StimSlide.ACC == "1"), control × lmerControl(optimizer = "bobyqa")

Fixed effects for the gaze data 1

Looks to the incorrect destination (the plate) in ID region (“on the plate”)

Dependent variable	*β*	SE	*t*	*p*
Group	0.01	0.02	0.43	0.666
Ambiguity	-0.02	0.02	-1.04	0.299
Context	0.01	0.02	0.40	0.688
Time1	-0.03	0.53	-0.05	0.957
Time2	-0.73	0.52	-1.41	0.160
Group × Ambiguity	0.00	0.02	-0.09	0.931
Group × Context	0.01	0.02	0.46	0.646
Ambiguity × Context	-0.01	0.02	-0.54	0.598
Group × Time1	-1.14	0.52	-2.18	0.030
Group × Time2	-0.03	0.52	-0.06	0.949
Ambiguity × Time1	0.33	0.52	0.64	0.524
Ambiguity × Time2	-0.09	0.52	-0.17	0.867
Context × Time1	-0.53	0.52	-1.01	0.311
Context × Time2	1.61	0.52	3.13	0.002
Group × Ambiguity × Context	-0.01	0.02	-0.70	0.484
Group × Ambiguity × Time1	1.36	0.52	2.60	0.009
Group × Ambiguity × Time2	-0.72	0.52	-1.40	0.163
Group × Context × Time1	-0.38	0.52	-0.72	0.473
Group × Context × Time2	-0.75	0.52	-1.45	0.149
Ambiguity × Context × Time1	0.27	0.52	0.51	0.612
Ambiguity × Context × Time2	0.03	0.52	0.05	0.958
Group × Ambiguity × Context × Time1	-0.40	0.52	-0.77	0.443
Group × Ambiguity × Context × Time2	0.32	0.52	0.62	0.534

R script

elog.id ~ Group* Ambiguity* Context*poly(scaledbin, 2) + (1|subject) + (1|item) + (0+Ambiguity × Context||item)+ (1|subject_trial),
control = lmerControl(optimizer="bobyqa"), weights = 1/wts.id, data = subset(CHWL8, region == "id"))

Fixed effects for the gaze data 2

Looks to the correct destination (the bag) in ID region ("on the plate")

Dependent variable	β	SE	t	p
Group	-0.01	0.02	-0.59	0.557
Ambiguity	0.00	0.03	-0.10	0.923
Context	0.04	0.03	1.26	0.227
Time1	1.13	0.67	1.68	0.093
Time2	0.40	0.66	0.61	0.542
Time3	0.09	0.66	0.13	0.896
Time4	0.09	0.66	0.13	0.893
Group × Ambiguity	0.00	0.03	-0.19	0.853
Group × Context	-0.01	0.02	-0.31	0.756
Ambiguity × Context	-0.02	0.03	-0.96	0.355
Group × Time1	1.01	0.67	1.50	0.133
Group × Time2	-0.32	0.66	-0.49	0.625
Group × Time3	1.14	0.66	1.73	0.084
Group × Time4	0.50	0.65	0.77	0.441
Ambiguity × Time1	4.67	0.67	6.96	0.000
Ambiguity × Time2	1.72	0.66	2.60	0.009
Ambiguity × Time3	-0.82	0.66	-1.24	0.216
Ambiguity × Time4	1.43	0.66	2.18	0.029
Context × Time1	0.73	0.67	1.08	0.279
Context × Time2	1.37	0.66	2.08	0.038
Context × Time3	-1.14	0.66	-1.72	0.085
Context × Time4	0.73	0.66	1.11	0.268
Group × Ambiguity × Context	0.04	0.03	1.36	0.195
Group × Ambiguity × Time1	-0.80	0.67	-1.20	0.232
Group × Ambiguity × Time2	0.74	0.66	1.12	0.263
Group × Ambiguity × Time3	0.77	0.66	1.17	0.240
Group × Ambiguity × Time4	-2.19	0.66	-3.34	0.001
Group × Context × Time1	1.04	0.67	1.55	0.122
Group × Context × Time2	1.27	0.66	1.91	0.056
Group × Context × Time3	-0.31	0.66	-0.47	0.640
Group × Context × Time4	1.39	0.65	2.13	0.034
Ambiguity × Context × Time1	0.73	0.67	1.09	0.276
Ambiguity × Context × Time2	-0.29	0.66	-0.44	0.660
Ambiguity × Context × Time3	1.26	0.66	1.92	0.055
Ambiguity × Context × Time4	-0.63	0.66	-0.97	0.333
Group × Ambiguity × Context × Time1	0.85	0.67	1.27	0.205
Group × Ambiguity × Context × Time2	-0.48	0.66	-0.73	0.465
Group × Ambiguity × Context × Time3	-0.34	0.66	-0.52	0.602
Group × Ambiguity × Context × Time4	0.27	0.66	0.41	0.683

R script

```
elog.cd ~ Group*Ambiguity*Context*poly(Time, 4) + (1|subject)  + (0+Ambiguity|subject) + (0+Context|subject) + (1|item) + (0+Ambiguity|item) + (0+Context|item) + (0+Group × Ambiguity|item) + (0+Group × Ambiguity × Context|item) + (0+Ambiguity × Context|item) + (1|subject_trial) + (1|item_ambiguous), control = lmerControl(optimizer="bobyqa"), weights =   1/wts.cd
```

Fixed effects for the gaze data 3

Looks to the incorrect destination (the plate) in CD region (“in the bag”)

Dependent variable	*β*	SE	*t*	*p*
Group	-0.04	0.01	-2.39	0.017
Ambiguity	-0.02	0.02	-1.11	0.266
Context	0.02	0.02	0.96	0.341
Time1	-0.42	0.45	-0.92	0.356
Time2	0.98	0.45	2.19	0.029
Time3	0.68	0.44	1.55	0.121
Group × Ambiguity	0.01	0.02	0.83	0.408
Group × Context	0.00	0.02	0.25	0.802
Ambiguity × Context	-0.01	0.02	-0.43	0.666
Group × Time1	0.15	0.45	0.34	0.737
Group × Time2	0.87	0.45	1.96	0.051
Group × Time3	0.59	0.44	1.35	0.177
Ambiguity × Time1	0.41	0.45	0.90	0.367
Ambiguity × Time2	-0.50	0.45	-1.12	0.263
Ambiguity × Time3	-0.16	0.44	-0.36	0.717
Context × Time1	-0.40	0.45	-0.88	0.377
Context × Time2	-0.22	0.45	-0.50	0.619
Context × Time3	-0.05	0.44	-0.12	0.903
Group × Ambiguity × Context	0.00	0.02	-0.22	0.828
Group × Ambiguity × Time1	-1.59	0.45	-3.52	0.000
Group × Ambiguity × Time2	0.84	0.45	1.88	0.060
Group × Ambiguity × Time3	-1.24	0.44	-2.82	0.005
Group × Context × Time1	0.34	0.45	0.75	0.456
Group × Context × Time2	-0.95	0.45	-2.11	0.035
Group × Context × Time3	0.88	0.44	2.00	0.045
Ambiguity × Context × Time1	0.59	0.45	1.30	0.195
Ambiguity × Context × Time2	-0.41	0.45	-0.91	0.363
Ambiguity × Context × Time3	-0.74	0.44	-1.69	0.092
Group × Ambiguity × Context × Time1	-0.30	0.45	-0.67	0.502
Group × Ambiguity × Context × Time2	1.24	0.45	2.77	0.006

R script

```
elog.id ~ Group*Ambiguity*Context*poly(Time, 3)  + (0+Context|subject) + (1|item) +  (1|subject_trial) , c
ontrol = lmerControl(optimizer="bobyqa"), weights =   1/wts.id, data = subset(CHWL8, region == "cd"))
```

Fixed effects for the gaze data 4

Looks to the correct destination (the bag) in CD region ("in the bag")

Dependent variable	*β*	SE	*t*	*p*
Group	0.02	0.04	0.61	0.545
Ambiguity	0.09	0.03	3.05	0.003
Context	0.04	0.03	1.39	0.184
Time1	3.63	0.80	4.54	<0.001
Time2	-5.13	0.79	-6.49	<0.001
Group × Ambiguity	0.01	0.03	0.18	0.857
Group × Context	0.01	0.03	0.23	0.816
Ambiguity × Context	-0.01	0.03	-0.23	0.818
Group × Time1	-0.18	0.80	-0.23	0.817
Group × Time2	-0.55	0.79	-0.69	0.489
Ambiguity × Time1	-1.43	0.80	-1.79	0.074
Ambiguity × Time2	0.85	0.79	1.07	0.283
Context × Time1	0.90	0.80	1.12	0.261
Context × Time2	-0.20	0.79	-0.25	0.804
Group × Ambiguity × Context	-0.01	0.03	-0.22	0.832
Group × Ambiguity × Time1	0.39	0.80	0.49	0.625
Group × Ambiguity × Time2	-0.64	0.79	-0.81	0.421
Group × Context × Time1	-0.16	0.80	-0.20	0.842
Group × Context × Time2	-0.86	0.79	-1.09	0.276
Ambiguity × Context × Time1	-2.21	0.80	-2.77	0.006
Ambiguity × Context × Time2	-0.90	0.79	-1.14	0.256
Group × Ambiguity × Context × Time1	0.22	0.80	0.28	0.779
Group × Ambiguity × Context × Time2	0.60	0.79	0.76	0.450

R script

Looks to the correct destination

elog.cd ~ Group * Ambiguity * Context * poly(Time, 2) + (1 | subject) + (0 + Ambiguity | subject) + (1 | item) + (0 + Context | item) + (0 + Group × Ambiguity × Context ||
item) + (0 + Ambiguity × Context | item) + (1 | subject_trial), weights = 1/wts.cd, control= lmerControl(optimizer = "bobyqa")

Fixed effects for the gaze data 5

Looks in the FINAL region ("before going to school")

	Looks to the incorrect destination (the plate)				Looks to the incorrect destination (the bag)			
Dependent variable	*β*	SE	*t*	*p*	*β*	SE	*t*	*p*
Group	-0.01	0.02	-0.57	0.572	-0.06	0.03	-1.72	0.098
Ambiguity	0.01	0.02	0.57	0.570	-0.03	0.03	-1.03	0.306
Context	0.05	0.02	2.81	0.013	0.07	0.03	2.52	0.025
Time1	0.74	0.51	1.45	0.146	-4.84	0.67	-7.21	0.000
Time2	-0.66	0.51	-1.30	0.195	3.89	0.67	5.84	0.000
Group × Ambiguity	0.02	0.02	0.88	0.394	-0.02	0.03	-0.68	0.506
Group × Context	0.00	0.02	0.28	0.782	-0.02	0.02	-1.05	0.294
Ambiguity × Context	0.00	0.02	0.02	0.985	-0.01	0.02	-0.43	0.664
Group × Time1	-0.06	0.51	-0.12	0.908	-1.22	0.67	-1.81	0.070
Group × Time2	-0.33	0.51	-0.66	0.510	0.92	0.67	1.38	0.167
Ambiguity × Time1	-0.01	0.51	-0.02	0.986	1.93	0.67	2.88	0.004
Ambiguity × Time2	-0.56	0.50	-1.11	0.265	-1.51	0.67	-2.26	0.024
Context × Time1	0.99	0.51	1.94	0.053	-1.65	0.67	-2.46	0.014
Context × Time2	-0.91	0.51	-1.80	0.072	0.42	0.67	0.64	0.525
Group × Ambiguity × Context	-0.02	0.02	-1.27	0.209	-0.02	0.03	-0.64	0.531
Group × Ambiguity × Time1	1.33	0.51	2.61	0.009	0.10	0.67	0.15	0.880
Group × Ambiguity × Time2	0.93	0.50	1.85	0.065	-0.08	0.67	-0.12	0.902
Group × Context × Time1	1.23	0.51	2.41	0.016	-0.44	0.67	-0.66	0.512
Group × Context × Time2	0.21	0.51	0.42	0.676	-0.82	0.67	-1.24	0.216
Ambiguity × Context × Time1	-0.09	0.51	-0.18	0.861	-0.21	0.67	-0.31	0.759
Ambiguity × Context × Time2	-0.51	0.50	-1.02	0.309	-0.89	0.67	-1.34	0.182
Group × Ambiguity × Context × Time1	-0.55	0.51	-1.08	0.282	1.65	0.67	2.46	0.014
Group × Ambiguity × Context × Time2	-0.39	0.50	-0.78	0.435	-0.01	0.67	-0.02	0.984

R script

Looks to incorrect destination: elog.id ~ Group*Ambiguity*Context*poly(Time, 2) + (1+Ambiguity*Context||subject) + (1|item) + (0+Context|item) + (0+Group × Ambiguity|item) + (1|subject_trial), control = lmerControl(optimizer="bobyqa"), weights = 1/wts.id, data = subset(CHWL8, region == "final")

Looks to correct destination: elog.cd ~ Group*Ambiguity*Context*poly(Time, 2) + (1|subject) + (0+Ambiguity|subject) + (1|item) + (0+Group|item) + (0+Context|item) + (0+Group × Ambiguity|item) + (0+Group × Ambiguity × Context|item) + (1|subject_trial), control = lmerControl(optimizer="bobyqa"), weights = 1/wts.cd, data = subset(CHWL8, region == "final")

APPENDIX H

Models for Study 1BB

Fixed effects for accuracy

	Contrast 1 ML vs. BL then BL-NS vs. BL-NNS				Contrast 2 NS vs. NNS then ML-NS vs. BL-NS			
Dependent variable	SE	*z*	SE	*p*	SE	*z*	SE	*p*
Group1	-0.09	0.20	-0.45	0.653	-0.08	0.20	-0.38	0.701
Group2	-0.03	0.18	-0.17	0.867	-0.05	0.19	-0.28	0.776
Ambiguity	0.81	0.10	7.83	0.000	0.81	0.10	7.83	0.000
Context	-0.01	0.10	-0.10	0.922	-0.01	0.10	-0.10	0.922
Group1 × Ambiguity	-0.13	0.14	-0.90	0.370	-0.17	0.13	-1.28	0.201
Group2 × Ambiguity	-0.11	0.13	-0.85	0.393	-0.04	0.13	-0.32	0.747
Group1 × Context	-0.05	0.14	-0.37	0.713	-0.11	0.13	-0.80	0.423
Group2 × Context	-0.08	0.12	-0.65	0.519	0.00	0.13	0.01	0.995
Ambiguity × Context	0.03	0.10	0.29	0.775	0.03	0.10	0.29	0.775
Group1 × Ambiguity × Context	-0.11	0.14	-0.76	0.450	-0.12	0.13	-0.91	0.363
Group2 × Ambiguity × Context	-0.07	0.12	-0.54	0.589	-0.05	0.13	-0.36	0.721

R script

CONTRAST 1: (StimSlide.ACC~ Group*Ambiguity*Context + (1|Subject) + (1|item), family = "binomial", control = glmerControl(optimizer="bobyqa")

CONSTAST2: StimSlide.ACC~ Group*Ambiguity*Context + (1|Subject) + (1|item), family = "binomial", control = glmerControl(optimizer="bobyqa")

Fixed effects for reaction times

	Contrast 1 ML vs. BL then BL-NS vs. BL-NNS				Contrast 2 NS vs. NNS then ML-NS vs. BL-NS			
Dependent variable	*β*	SE	*z*	*p*	*β*	SE	*z*	*p*
Group1	141.09	72.56	1.94	0.055	17.72	73.48	0.24	0.810
Group2	-52.52	67.34	-0.78	0.437	132.23	66.80	1.98	0.050
Ambiguity	-179.69	27.53	-6.53	0.000	-179.75	29.39	-6.12	0.000
Context	92.60	27.50	3.37	0.001	92.86	27.37	3.39	0.001
Group1 × Ambiguity	-98.66	37.56	-2.63	0.009	-40.22	40.42	-1.00	0.322
Group2 × Ambiguity	7.65	34.90	0.22	0.827	-78.19	36.98	-2.11	0.037
Group1 × Context	78.25	37.46	2.09	0.037	34.41	37.59	0.92	0.360
Group2 × Context	-5.09	34.87	-0.15	0.884	61.03	34.44	1.77	0.077
Ambiguity × Context	-57.52	27.50	-2.09	0.037	-57.63	27.37	-2.11	0.035
Group1 × Ambiguity × Context	-65.20	37.45	-1.74	0.082	-32.23	37.56	-0.86	0.391
Group2 × Ambiguity × Context	1.03	34.84	0.03	0.977	-49.02	34.44	-1.42	0.155

R script

CONTRAST 1: ReactonTimes~ Group*Ambiguity*Context + (1|Subject) + (1|item), control = lmerControl(optimizer="bobyqa")

CONTRAST 2: ReactonTimes~ Group*Ambiguity*Context + (1|Subject) +(0+Ambiguity|Subject) + (1|item), control = lmerControl(optimizer="bobyqa")

Fixed effects for the gaze data 1

Looks to the incorrect destination (the plate) in the ID region ("on the plate")

	Contrast 1 ML vs. BL then BL-NS vs. BL-NNS				Contrast 2 NS vs. NNS then ML-NS vs. BL-NS			
Dependent variable	*β*	SE	*z*	*p*	*β*	SE	*z*	*p*
Group1	0.05	0.02	2.02	0.047	0.01	0.02	0.43	0.671
Group2	-0.01	0.02	-0.63	0.532	0.04	0.02	1.98	0.051
Ambiguity	-0.05	0.01	-3.97	0.000	-0.05	0.01	-3.97	0.000
Context	-0.02	0.01	-2.11	0.035	-0.02	0.01	-2.11	0.035
Time1	3.32	0.48	6.86	<0.001	3.32	0.48	6.86	<0.001
Time2	0.28	0.48	0.57	0.568	0.28	0.48	0.57	0.568
Group1 × Ambiguity	-0.03	0.02	-1.78	0.078	-0.03	0.02	-1.68	0.097
Group2 × Ambiguity	-0.01	0.02	-0.87	0.385	-0.02	0.02	-1.02	0.310
Group1 × Context	0.00	0.02	0.07	0.946	0.02	0.02	1.29	0.197
Group2 × Context	0.02	0.01	1.38	0.169	-0.01	0.01	-0.64	0.523
Ambiguity × Context	0.00	0.01	-0.43	0.664	0.00	0.01	-0.43	0.664
Group1 × Time1	1.08	0.66	1.63	0.103	-1.23	0.66	-1.85	0.064
Group2 × Time1	-1.77	0.61	-2.89	0.004	1.69	0.61	2.78	0.005
Group1 × Time2	-0.27	0.66	-0.40	0.686	0.34	0.66	0.51	0.612
Group2 × Time2	0.47	0.61	0.77	0.442	-0.43	0.61	-0.72	0.474
Ambiguity × Time1	-1.58	0.48	-3.27	0.001	-1.58	0.48	-3.27	0.001
Ambiguity × Time2	0.76	0.48	1.56	0.118	0.76	0.48	1.56	0.118
Context × Time1	-1.96	0.48	-4.06	<0.001	-1.96	0.48	-4.06	<0.001
Context × Time2	0.71	0.48	1.47	0.142	0.71	0.48	1.47	0.142
Group1 × Ambiguity × Context	0.00	0.02	0.08	0.938	-0.02	0.02	-1.52	0.129
Group2 × Ambiguity × Context	-0.02	0.01	-1.71	0.089	0.01	0.01	0.93	0.354
Group1 × Ambiguity × Time1	1.37	0.66	2.08	0.037	-0.84	0.66	-1.26	0.208
Group2 × Ambiguity × Time1	-1.52	0.61	-2.49	0.013	1.79	0.61	2.95	0.003
Group1 × Ambiguity × Time2	0.65	0.66	0.99	0.324	1.93	0.66	2.91	0.004
Group2 × Ambiguity × Time2	1.61	0.61	2.63	0.009	-0.32	0.61	-0.52	0.600
Group1 × Context × Time1	-0.33	0.66	-0.51	0.614	1.37	0.66	2.06	0.040
Group2 × Context × Time1	1.54	0.61	2.51	0.012	-1.02	0.61	-1.67	0.095
Group1 × Context × Time2	-1.23	0.66	-1.86	0.062	-1.71	0.66	-2.57	0.010
Group2 × Context × Time2	-1.10	0.61	-1.79	0.073	-0.37	0.61	-0.61	0.541
Ambiguity × Context × Time1	0.03	0.48	0.05	0.958	0.03	0.48	0.05	0.958
Ambiguity × Context × Time2	-0.24	0.48	-0.50	0.614	-0.24	0.48	-0.50	0.614
Group1 × Ambiguity × Context × Time1	-0.93	0.66	-1.40	0.161	0.14	0.66	0.21	0.831
Group2 × Ambiguity × Context × Time1	0.60	0.61	0.99	0.323	-1.00	0.61	-1.64	0.101
Group1 × Ambiguity × Context × Time2	-0.22	0.66	-0.33	0.742	-0.13	0.66	-0.19	0.849
Group2 × Ambiguity × Context × Time2	-0.02	0.61	-0.03	0.977	-0.15	0.61	-0.25	0.800

R script

CONTRAST 1: elog.id ~ Group*Ambiguity*Context*poly(Time, 2) + (1|subject) + (0+Ambiguity|subject) + (1|item) + (1|subject_trial), c ontrol = lmerControl(optimizer="bobyqa"), weights = 1/wts.id,
CONTRAST 2: elog.id ~ Group*Ambiguity*Context*poly(Time, 2) + (1|subject) + (0+Ambiguity|subject) + (1|item) + (1|subject_trial), c ontrol = lmerControl(optimizer="bobyqa"), weights = 1/wts.id)

Fixed effects for the gaze data 2

Looks to the correct destination (the bag) in the ID region (“on the plate”)

	Contrast 1 ML vs. BL then BL-NS vs. BL-NNS				Contrast 2 NS vs. NNS then ML-NS vs. BL-NS			
Dependent variable	*β*	SE	*z*	*p*	*β*	SE	*z*	*p*
Group1	0.01	0.02	0.25	0.804	0.01	0.02	0.24	0.812
Group2	0.00	0.02	0.13	0.899	0.00	0.02	0.14	0.888
Ambiguity	0.02	0.01	1.36	0.195	0.02	0.01	1.36	0.195
Context	0.00	0.02	0.18	0.856	0.00	0.02	0.18	0.856
Time	7.90	0.58	13.53	<0.001	7.90	0.58	13.53	<0.001
Group1 × Ambiguity	-0.01	0.02	-0.40	0.689	-0.02	0.02	-1.16	0.245
Group2 × Ambiguity	-0.02	0.02	-1.06	0.292	0.00	0.02	0.21	0.837
Group1 × Context	-0.02	0.02	-1.36	0.173	-0.02	0.02	-1.07	0.286
Group2 × Context	-0.01	0.02	-0.43	0.668	-0.01	0.02	-0.90	0.367
Ambiguity × Context	0.01	0.01	0.43	0.670	0.01	0.01	0.43	0.670
Group1 × Time	2.29	0.80	2.88	0.004	1.74	0.80	2.18	0.029
Group2 × Time	0.60	0.74	0.81	0.419	1.42	0.73	1.94	0.053
Ambiguity × Time	3.89	0.58	6.66	0.000	3.89	0.58	6.66	0.000
Context × Time	-0.44	0.58	-0.76	0.447	-0.44	0.58	-0.76	0.447
Group1 × Ambiguity × Context	0.02	0.02	1.03	0.305	0.02	0.02	1.33	0.186
Group2 × Ambiguity × Context	0.01	0.02	0.90	0.370	0.01	0.02	0.39	0.698
Group1 × Ambiguity × Time	-0.99	0.80	-1.24	0.214	-1.69	0.80	-2.11	0.035
Group2 × Ambiguity × Time	-1.19	0.74	-1.62	0.106	-0.15	0.73	-0.20	0.843
Group1 × Context × Time	1.63	0.80	2.04	0.041	0.96	0.80	1.19	0.233
Group2 × Context × Time	0.14	0.74	0.19	0.848	1.15	0.73	1.56	0.118
Ambiguity × Context × Time	-1.34	0.58	-2.29	0.022	-1.34	0.58	-2.29	0.022
Group1 × Ambiguity × Context × Time	0.84	0.80	1.06	0.290	0.84	0.80	1.04	0.297
Group2 × Ambiguity × Context × Time	0.41	0.74	0.56	0.575	0.43	0.73	0.58	0.563

R script

CONTRAST 1: elog.cd ~ Group * Ambiguity * Context * poly(Time,1) + (1 | subject) + (1 + Ambiguity * Context || item) + (1 | subject _trial), Weights × 1/wts.cd, Control × lmerControl(optimizer = "bobyqa"), data × subset(ADWL8, region == "id")

CONTRAST 2: elog.cd ~ Group*Ambiguity*Context*poly(Time, 1) + (1|subject) + (1+Ambiguity*Context||item) + (1|subject_trial), contr ol = lmerControl(optimizer="bobyqa"), weights = 1/wts.target, data = subset(ADWL8, region == "id"))

Fixed effects for the gaze data 3

Looks to the incorrect destination (the plate) in the CD region ("in the bag")

	Contrast 1 ML vs. BL then BL-NS vs. BL-NNS				Contrast 2 NS vs. NNS then ML-NS vs. BL-NS			
Dependent variable	*β*	SE	*z*	*p*	*β*	SE	*z*	*p*
Group1	0.01	0.02	0.43	0.669	-0.01	0.02	-0.38	0.705
Group2	-0.01	0.02	-0.71	0.487	0.01	0.02	0.71	0.487
Ambiguity	-0.02	0.01	-1.58	0.114	-0.02	0.01	-1.58	0.114
Context	0.03	0.01	2.81	0.006	0.03	0.01	2.81	0.006
Time1	-1.58	0.46	-3.42	0.001	-1.58	0.46	-3.42	0.001
Time2	1.50	0.46	3.27	0.001	1.50	0.46	3.27	0.001
Group1 × Ambiguity	0.01	0.01	0.41	0.684	0.02	0.01	1.11	0.266
Group2 × Ambiguity	0.01	0.01	1.00	0.319	0.00	0.01	-0.18	0.860
Group1 × Context	0.00	0.02	0.14	0.889	0.00	0.02	0.06	0.955
Group2 × Context	0.00	0.01	-0.01	0.989	0.00	0.01	0.12	0.903
Ambiguity × Context	0.00	0.01	-0.07	0.943	0.00	0.01	-0.07	0.943
Group1 × Time1	0.06	0.62	0.09	0.929	0.16	0.64	0.25	0.804
Group2 × Time1	0.13	0.59	0.22	0.824	-0.02	0.58	-0.04	0.968
Group1 × Time2	1.03	0.62	1.67	0.096	-0.17	0.63	-0.28	0.782
Group2 × Time2	-0.69	0.58	-1.18	0.237	1.11	0.57	1.95	0.051
Ambiguity × Time1	0.53	0.46	1.15	0.251	0.53	0.46	1.15	0.251
Ambiguity × Time2	-0.38	0.46	-0.83	0.405	-0.38	0.46	-0.83	0.405
Context × Time1	1.75	0.46	3.80	0.000	1.75	0.46	3.80	0.000
Context × Time2	-0.90	0.46	-1.97	0.049	-0.90	0.46	-1.97	0.049
Group1 × Ambiguity × Context	-0.01	0.01	-0.98	0.328	-0.01	0.01	-0.56	0.574
Group2 × Ambiguity × Context	0.00	0.01	-0.09	0.925	-0.01	0.01	-0.75	0.451
Group1 × Ambiguity × Time1	0.11	0.62	0.18	0.855	0.02	0.64	0.03	0.977
Group2 × Ambiguity × Time1	-0.04	0.59	-0.07	0.948	0.10	0.57	0.18	0.856
Group1 × Ambiguity × Time2	-0.20	0.62	-0.32	0.750	-0.73	0.63	-1.16	0.246
Group2 × Ambiguity × Time2	-0.63	0.58	-1.09	0.276	0.17	0.57	0.30	0.766
Group1 × Context × Time1	-0.03	0.62	-0.05	0.961	-1.48	0.64	-2.33	0.020
Group2 × Context × Time1	-1.47	0.59	-2.50	0.012	0.71	0.57	1.24	0.216
Group1 × Context × Time2	-0.21	0.62	-0.35	0.729	0.76	0.63	1.20	0.231
Group2 × Context × Time2	0.86	0.58	1.48	0.139	-0.59	0.57	-1.04	0.300
Ambiguity × Context × Time1	-0.89	0.46	-1.93	0.054	-0.89	0.46	-1.93	0.054
Ambiguity × Context × Time2	0.29	0.46	0.63	0.529	0.29	0.46	0.63	0.529
Group1×Ambiguity× Context ×Time1	0.70	0.62	1.13	0.259	1.58	0.64	2.48	0.013
Group2×Ambiguity× Context × Time1	1.23	0.59	2.09	0.037	-0.09	0.57	-0.15	0.881
Group1×Ambiguity× Context × Time2	-0.62	0.62	-1.01	0.313	-0.15	0.63	-0.24	0.809
Group2×Ambiguity× Context× Time2	0.16	0.58	0.27	0.786	-0.55	0.57	-0.96	0.338

R script

```
CONTRAST 1:  elog.id ~ Group*Ambiguity*Context*poly(Time, 2) + (1|subject) + (0+Context|subject) + (1|item) +(0+Group|item)  + (1|s
ubject_trial), control = lmerControl(optimizer="bobyqa"), weights =   1/wts.id,
              data = subset(ADWL8, region == "cd")

CONTRAST 2:   elog.id ~ Group*Ambiguity*Context*poly(Time, 2) + (1|subject) + (0+Context|subject) + (1|item) +(0+Group|item)  + (1|
subject_trial), control = lmerControl(optimizer="bobyqa"), weights =   1/wts.id
              data = subset(ADWL8, region == "cd")
```

Fixed effects for the gaze data 4
Looks to the correct destination (the bag) in the CD region ("in the bag")

	Contrast 1 ML vs. BL then BL-NS vs. BL-NNS				Contrast 2 NS vs. NNS then ML-NS vs. BL-NS			
Dependent variable	*β*	SE	*z*	*p*	*β*	SE	*z*	*p*
Group1	0.04	0.04	1.10	0.276	0.08	0.04	1.86	0.065
Group2	0.05	0.04	1.46	0.148	0.01	0.04	0.16	0.871
Ambiguity	0.04	0.03	1.46	0.159	0.04	0.03	1.46	0.159
Context	0.00	0.02	-0.15	0.878	0.00	0.02	-0.15	0.878
Time1	8.79	0.79	11.14	< 0.001	8.79	0.79	11.14	< 0.001
Time2	-9.20	0.78	-11.76	< 0.001	-9.20	0.78	-11.76	< 0.001
Group1 × Ambiguity	0.00	0.03	0.06	0.957	-0.04	0.03	-1.39	0.167
Group2 × Ambiguity	-0.04	0.03	-1.56	0.123	0.02	0.03	0.84	0.405
Group1 × Context	-0.01	0.02	-0.41	0.679	0.04	0.02	1.67	0.094
Group2 × Context	0.05	0.02	2.06	0.040	-0.03	0.02	-1.38	0.167
Ambiguity × Context	0.01	0.03	0.32	0.754	0.01	0.03	0.32	0.754
Group1 × Time1	-1.85	1.06	-1.74	0.083	-0.88	1.09	-0.81	0.417
Group2 × Time1	0.04	1.00	0.04	0.968	-1.41	0.98	-1.43	0.153
Group1 × Time2	-0.48	1.06	-0.46	0.649	-2.30	1.08	-2.13	0.033
Group2 × Time2	-2.06	1.00	-2.07	0.038	0.67	0.98	0.69	0.491
Ambiguity × Time1	-0.36	0.79	-0.46	0.648	-0.36	0.79	-0.46	0.648
Ambiguity × Time2	-0.18	0.78	-0.23	0.819	-0.18	0.78	-0.23	0.819
Context × Time1	1.22	0.79	1.55	0.121	1.22	0.79	1.55	0.121
Context × Time2	1.16	0.78	1.49	0.137	1.16	0.78	1.49	0.137
Group1 × Ambiguity × Context	0.00	0.03	-0.08	0.935	0.01	0.04	0.24	0.812
Group2 × Ambiguity × Context	0.01	0.03	0.32	0.753	-0.01	0.03	-0.25	0.804
Group1 × Ambiguity × Time1	-0.18	1.06	-0.17	0.863	1.98	1.09	1.82	0.069
Group2 × Ambiguity × Time1	2.07	1.00	2.06	0.039	-1.17	0.98	-1.19	0.234
Group1 × Ambiguity × Time2	-0.17	1.06	-0.16	0.870	-0.02	1.08	-0.01	0.988
Group2 × Ambiguity × Time2	0.07	1.00	0.07	0.944	-0.16	0.98	-0.17	0.866
Group1 × Context × Time1	0.38	1.06	0.36	0.718	-2.41	1.09	-2.21	0.027
Group2 × Context × Time1	-2.60	1.00	-2.59	0.010	1.59	0.98	1.61	0.107
Group1 × Context × Time2	0.45	1.06	0.43	0.668	-1.26	1.08	-1.17	0.244
Group2 × Context × Time2	-1.48	1.00	-1.49	0.136	1.08	0.98	1.11	0.267
Ambiguity × Context × Time1	2.71	0.79	3.44	0.001	2.71	0.79	3.44	0.001
Ambiguity × Context × Time2	-0.51	0.78	-0.66	0.513	-0.51	0.78	-0.66	0.513
Group1 × Ambiguity × Context × Time1	-0.04	1.07	-0.03	0.973	1.73	1.09	1.59	0.113
Group2 × Ambiguity × Context × Time1	1.75	1.00	1.74	0.082	-0.90	0.98	-0.91	0.361
Group1 × Ambiguity × Context × Time2	0.64	1.06	0.61	0.544	0.79	1.08	0.73	0.464
Group2 × Ambiguity × Context × Time2	0.47	1.00	0.47	0.636	0.25	0.98	0.25	0.802

R scrip

CONTRAST 1: elog.cd ~ Group*Ambiguity*Context*poly(Time, 2) + (1|subject) + (0+Ambiguity|subject) + (0+Ambiguity × Context|subject) + (1|item) + (0+Group × Ambiguity × Context||item) + (0+Ambiguity|item) + (1|subject_trial) + (1|item_ambiguous), control = lmerControl(optimizer="bobyqa"), weights = 1/wts.cd, data = subset(ADWL8, region == "cd")

CONTRAST 2: elog.cd ~ Group*Ambiguity*Context*poly(Time, 2) + (1|subject) + (0+Ambiguity|subject) + (0+Ambiguity × Context|subject) + (1|item) + (0+Group × Ambiguity × Context||item) + (0+Ambiguity|item) + (1|subject_trial) + (1|item_ambiguous), control = lmerControl(optimizer="bobyqa"), weights = 1/wts.cd, data = subset(ADWL8, region == "cd")

Fixed effects for the gaze data 5

Looks to the incorrect destination (the plate) in the FINAL region ("before going to school")

	Contrast 1 ML vs. BL then BL-NS vs. BL-NNS				Contrast 2 NS vs. NNS then ML-NS vs. BL-NS			
Dependent variable	*β*	SE	*z*	*p*	*β*	SE	*z*	*p*
Group1	0.01	0.02	0.38	0.702	0.01	0.02	0.38	0.702
Group2	-0.03	0.02	-2.17	0.032	-0.03	0.02	-2.17	0.032
Ambiguity	0.00	0.01	-0.15	0.877	0.00	0.01	-0.15	0.877
Context	0.03	0.01	2.73	0.006	0.03	0.01	2.73	0.006
Time1	1.19	0.47	2.53	0.012	1.19	0.47	2.53	0.012
Time2	-1.36	0.47	-2.89	0.004	-1.36	0.47	-2.89	0.004
Time3	0.76	0.47	1.64	0.102	0.76	0.47	1.64	0.102
Group1 × Ambiguity	0.01	0.02	0.45	0.652	0.01	0.02	0.45	0.652
Group2 × Ambiguity	-0.01	0.01	-0.90	0.369	-0.01	0.01	-0.90	0.369
Group1 × Context	0.04	0.02	2.27	0.023	0.04	0.02	2.27	0.023
Group2 × Context	0.01	0.01	0.66	0.509	0.01	0.01	0.66	0.509
Ambiguity × Context	0.01	0.01	1.01	0.311	0.01	0.01	1.01	0.311
Group1 × Time1	0.06	0.64	0.10	0.921	0.06	0.64	0.10	0.921
Group2 × Time1	0.47	0.60	0.79	0.429	0.47	0.60	0.79	0.429
Group1 × Time2	1.06	0.64	1.67	0.096	1.06	0.64	1.67	0.096
Group2 × Time2	-0.42	0.60	-0.70	0.484	-0.42	0.60	-0.70	0.484
Group1 × Time3	0.61	0.63	0.97	0.334	0.61	0.63	0.97	0.334
Group2 × Time3	0.81	0.59	1.37	0.171	0.81	0.59	1.37	0.171
Ambiguity × Time1	0.14	0.47	0.29	0.771	0.14	0.47	0.29	0.771
Ambiguity × Time2	0.11	0.47	0.24	0.814	0.11	0.47	0.24	0.814
Ambiguity × Time3	0.95	0.47	2.05	0.041	0.95	0.47	2.05	0.041
Context × Time1	0.00	0.47	0.00	0.999	0.00	0.47	0.00	0.999
Context × Time2	-0.68	0.47	-1.45	0.146	-0.68	0.47	-1.45	0.146
Context × Time3	-0.02	0.47	-0.05	0.959	-0.02	0.47	-0.05	0.959
Group1 × Ambiguity × Context	0.00	0.02	-0.18	0.861	0.00	0.02	-0.18	0.861
Group2 × Ambiguity × Context	-0.01	0.01	-0.96	0.340	-0.01	0.01	-0.96	0.340
Group1 × Ambiguity × Time1	0.61	0.64	0.95	0.345	0.61	0.64	0.95	0.345
Group2 × Ambiguity × Time1	0.41	0.60	0.68	0.496	0.41	0.60	0.68	0.496
Group1 × Ambiguity × Time2	-1.08	0.64	-1.70	0.090	-1.08	0.64	-1.70	0.090
Group2 × Ambiguity × Time2	-0.21	0.60	-0.36	0.720	-0.21	0.60	-0.36	0.720
Group1 × Ambiguity × Time3	-0.48	0.63	-0.76	0.445	-0.48	0.63	-0.76	0.445
Group2 × Ambiguity × Time3	0.61	0.59	1.04	0.300	0.61	0.59	1.04	0.300
Group1 × Context × Time1	-0.55	0.64	-0.86	0.388	-0.55	0.64	-0.86	0.388
Group2 × Context × Time1	-0.74	0.60	-1.23	0.218	-0.74	0.60	-1.23	0.218
Group1 × Context × Time2	0.56	0.64	0.88	0.377	0.56	0.64	0.88	0.377
Group2 × Context × Time2	-1.67	0.60	-2.80	0.005	-1.67	0.60	-2.80	0.005
Group1 × Context × Time3	-0.04	0.63	-0.06	0.950	-0.04	0.63	-0.06	0.950
Group2 × Context × Time3	1.46	0.59	2.47	0.014	1.46	0.59	2.47	0.014
Ambiguity × Context × Time1	-0.61	0.47	-1.31	0.192	-0.61	0.47	-1.31	0.192
Ambiguity × Context × Time2	-0.47	0.47	-1.01	0.312	-0.47	0.47	-1.01	0.312
Ambiguity × Context × Time3	0.37	0.47	0.79	0.428	0.37	0.47	0.79	0.428
Group1 × Ambiguity × Context × Time1	1.34	0.64	2.09	0.036	1.34	0.64	2.09	0.036
Group2 × Ambiguity × Context × Time1	1.06	0.60	1.77	0.078	1.06	0.60	1.77	0.078
Group1 × Ambiguity × Context × Time2	-0.28	0.64	-0.44	0.659	-0.28	0.64	-0.44	0.659
Group2 × Ambiguity × Context × Time2	-0.67	0.60	-1.12	0.262	-0.67	0.60	-1.12	0.262
Group1 × Ambiguity × Context × Time3	-0.16	0.63	-0.25	0.805	-0.16	0.63	-0.25	0.805
Group2 × Ambiguity × Context × Time3	0.39	0.59	0.66	0.507	0.39	0.59	0.66	0.507

R script

CONTRAST 1: elog.id ~ Group*Ambiguity*Context*poly(Time, 3) + (1|subject) + (1|item) + (1|subject_trial), control = lmerControl(optimizer="bobyqa"), weights = 1/wts.id, data = subset(ADWL8, region == "final")

CONTRAST 2: elog.id ~ Group*Ambiguity*Context*poly(Time, 3) + (1|subject) + (1|item) + (1|subject_trial), control = lmerControl(optimizer="bobyqa"), weights = 1/wts.id, data = subset(ADWL8, region == "final")

Fixed effects for the gaze data 6

Looks to the correct destination (the bag) in the FINAL region ("before going to school")

	Contrast 1 ML vs. BL then BL-NS vs. BL-NNS				Contrast 2 NS vs. NNS then ML-NS vs. BL-NS			
Dependent variable	*β*	SE	*z*	*p*	*β*	SE	*z*	*p*
Group1	-0.02	0.04	-0.44	0.658	-0.03	0.08	-0.44	0.663
Group2	0.02	0.04	0.61	0.546	0.05	0.07	0.67	0.502
Ambiguity	0.05	0.02	3.01	0.003	0.05	0.02	2.60	0.022
Context	0.02	0.02	1.05	0.296	0.02	0.02	1.09	0.281
Time1	-9.20	0.70	-13.10	< 0.001	-9.22	0.70	-13.11	< 0.001
Time2	3.25	0.70	4.65	0.000				
Group1 × Ambiguity	0.04	0.02	1.77	0.078	0.08	0.05	1.72	0.087
Group2 × Ambiguity	0.00	0.02	-0.22	0.829	-0.07	0.05	-1.53	0.129
Group1 × Context	0.00	0.03	-0.10	0.921	-0.01	0.06	-0.12	0.902
Group2 × Context	-0.01	0.03	-0.45	0.657	-0.01	0.05	-0.11	0.911
Ambiguity × Context	0.00	0.02	0.11	0.914	0.00	0.02	0.10	0.919
Group1 × Time1	1.32	0.95	1.39	0.166	2.67	1.91	1.40	0.162
Group2 × Time1	-0.43	0.89	-0.48	0.634	-2.46	1.77	-1.39	0.163
Group1 × Time2	-2.23	0.95	-2.36	0.019				
Group2 × Time2	1.90	0.89	2.14	0.032				
Ambiguity × Time1	-0.76	0.70	-1.09	0.277	-0.71	0.70	-1.01	0.311
Ambiguity × Time2	-0.51	0.70	-0.73	0.465				
Context × Time1	-2.31	0.70	-3.30	0.001	-2.30	0.70	-3.27	0.001
Context × Time2	0.54	0.70	0.77	0.443				
Group1 × Ambiguity × Context	-0.04	0.03	-1.57	0.119	-0.08	0.05	-1.56	0.121
Group2 × Ambiguity × Context	0.03	0.03	1.14	0.256	0.09	0.05	1.86	0.066
Group1 × Ambiguity × Time1	-0.17	0.95	-0.18	0.856	-0.50	1.91	-0.26	0.793
Group2 × Ambiguity × Time1	-0.94	0.89	-1.06	0.291	-0.56	1.77	-0.32	0.750
Group1 × Ambiguity × Time2	1.09	0.95	1.15	0.249				
Group2 × Ambiguity × Time2	-0.58	0.89	-0.65	0.514				
Group1 × Context × Time1	2.58	0.95	2.70	0.007	5.00	1.91	2.62	0.009
Group2 × Context × Time1	-0.05	0.89	-0.05	0.958	-3.81	1.77	-2.16	0.031
Group1 × Context × Time2	-0.66	0.95	-0.70	0.484				
Group2 × Context × Time2	1.07	0.89	1.21	0.227				
Ambiguity × Context × Time1	-1.23	0.70	-1.75	0.080	-1.22	0.70	-1.73	0.084
Ambiguity × Context × Time2	0.64	0.70	0.92	0.358				
Group1 × Ambiguity × Context × Time1	-1.01	0.95	-1.06	0.289	-2.03	1.91	-1.07	0.287
Group2 × Ambiguity × Context × Time1	-1.94	0.89	-2.18	0.030	-0.40	1.77	-0.22	0.823
Group1 × Ambiguity × Context × Time2	-2.60	0.95	-2.75	0.006				
Group2 × Ambiguity × Context × Time2	-0.22	0.89	-0.25	0.800				

R script

CONTRAST 1: elog.cd ~ Group*Ambiguity*Context*poly(Time, 2) + (1|subject) + (0+Context|subject) + (0+Ambiguity:Context||subject) + (1|item) + (1|subject_trial), control = lmerControl(optimizer="bobyqa"), weights = 1/wts.cd)

CONTRAST 2: elog.cd ~ Group*Ambiguity*Context*poly(Time, 1) + (1+Ambiguity*Context||subject) + (1|item) + (1|subject_trial) + (1|item_ambiguous), control = lmerControl(optimizer="bobyqa"), weights = 1/wts.cd,)

APPENDIX I

Group comparisons for baseline language measures administered to children

Task	Monolingual score (SD)	Bilingual score (SD)	Statistic	Effect size	Significance
		Receptive language			
C&F-raw	50.11 (3.19)	48.03 (4.23)	U = 392.000	z = -2.103	p = 0.35*
C&F-scaled	11.83 (1.69)	10.77 (2.14)	U =3 86.000	z = -2.197	p = .028*
TROG-2 raw	16.53 (1.73)	15.93 (2.50)	U = 489.000	z = 0.667	p = 0.505
TROG-2 standard	103.86 (8.24)	101.73 (11.41)	F(1,64)=.771	η2p = .012	p = 0.383
		Expressive language			
WS-raw	30.25 (1.63)	29.47 (4.10)	F(1,65) = 1.104	η2p=.017	p = 0.297
WS-scaled	12.10 (0.99)	12.63 (1.19)	U = 25.500	z = -1.394	p = 0.203
FS-raw	49.83 (4.63)	47.97 (6.27)	U = 454.500	z = -1.106	p = 0.269
FS-scaled	12.47 (2.09)	12.23 (2.96)	U= 551.500,	z = -.083	p = 0.934
		Vocabulary			
RTWF-raw	44.70 (3.25)	40.32 (4.41)	U = 242.500	z = -4.469	p<.001***
RTWF-std	107.87 (8.87)	97.21 (12.45)	U = 296.000	z = -3.845	p<.001***
Word classes – expressive (raw)	13.00 (3.66)	12.50 (4.37)	F(1,63) = 0.252	η2p= .004	p = 0.617
Word classes – expressive (scaled)	13.31 (2.88)	12.24(2.72)	U = 377.500	z = -1.928	p = 0.054
Word classes – receptive (raw)	14.34 (3.93)	13.60 (4.42)	F(1,63) = 0.515	η2p = .0048	p = 0.476
Word classes – receptive (scaled)	12.58 (3.46)	10.77 (3.44)	F(1,63) = 4.535	η2p = .066	p = 0.037*
Word classes – total score (raw)	25.16 (7.26)	22.60 (6.64)	F(1,63) = 2.228	η2p = .033	p = 0.140
		PSTM			
CNRep-raw	33.00 (3.99)	34.59 (2.60)	U=457.500	z = -1.625	p = .104
CNRep-scaled	103.51 (16.76)	109.50 (10.60)	U=464.500	z = -1.542	p = .123
		Recalling Sentences			
RS-raw	78.51 (8.26)	68.45 (13.81)	U=288.500,	z =-3.266	p = .001**
RS-scaled	13.28 (1.97)	11.19 (3.18)	U=212.000	z = -3.124	p =. 002**
		CELF-4 Composite Scores (standardised)			
CLS	116.36 (8.42)	110.83 (10.99)	U=380.000	z = -2.065	p = .039*
RLI	113.56 (13.07)	105.68 (11.70)	F(1,63)=6.525	η2p = .202	p =.013*
ELI	117.67 (8.66)	114.38 (11.80)	U=445.500	z = -1.222	p = 0.222
LMI	115.77 (7.23)	109.73 (11.50)	U=202.000	z = -1.95	p = 0.051

ANOVAs were used for normally distributed data - Mann Whitney when they were not

References

Adani, F., Forgiarini, M., Guasti, M. T. & van der Lely, H. K. J. (2014). Number dissimilarities facilitate the comprehension of relative clause in children affected by (Grammatical) Specific Language Impairment. *Journal of Child Language* 41(4), 811–41.

Altmann, G. T., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of memory and language*, 57(4), 502-518.

Atkinson, E., Wagers, M. W., Lidz, J., Phillips, C., & Omaki, A. (2018). Developing incrementality in filler-gap dependency processing. *Cognition*, 179, 132-149.

Avrutin, S. (2000). Comprehension of discourse-linked and non-discourse-linked questions by children and Broca's aphasics. In Grodzinsky, Yosef, Shapiro, Lee & Swinney, David (eds). *Language and brain: representation and processing* (pp. 295–313). San Diego, CA: Academic Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modelling with crossed random effects for subjects and items. *Journal of memory and language*, *59*(4), 390-412.

Barker, R. M., & Bialystok, E. (2019). Processing differences between monolingual and bilingual young adults on an emotion n-back task. *Brain and cognition*, 134, 29-43.

Barr, D. J. (2008). Analyzing 'visual world'eyetracking data using multilevel logistic regression. *Journal of memory and language*, *59*(4), 457-474.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, 68(3), 255-278.

Bates, D., Maechler, D., Bolker, B. & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1), 1–48.

Bates E, & MacWhinney B. (1982). Functionalist approaches to grammar. In: Gleitman L, Wanner E, (Eds.) *Language acquisition: the state of the art* (pp 173-218). Cambridge University Press.

Belletti, A., Bennati, E., & Sorace, A. (2007). Theoretical and developmental issues in the syntax of subjects: Evidence from near-native Italian. *Natural Language & Linguistic Theory*, 25(4), 657.

Bentea, A., & Durrleman, S. (2018). Not committing has its advantages: facilitating children's comprehension of object filler–gap dependencies is one of them. *Journal of child language*, 45(1), 35-71.

Bernolet, S., Hartsuiker, R. J., & Pickering, M. J. (2007). Shared syntactic representations in bilinguals: Evidence for the role of word-order repetition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(5), 931.

Bever, T.G. (1970). The cognitive bias for linguistic structures. In: Hayes JR, (Ed.). *Cognition and the development of language* (pp 279-362). New York: Wiley.

Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and cognition*, 12(1), 3-11

Bialystok, E. (2015). Bilingualism and the development of executive function: The role of attention. *Child development perspectives,* 9(2), 117-121.

Birdsong, D. (Ed.). (1999). Whys and Why Nots of the Critical Period Hypothesis for Second Language Acquisition. In D. Birdsong (Ed.) S*econd Language Acquisition and the Critical Period* (pp1-22). Mahwah, NJ: Erlbaum.

Blom, E., & Vasić, N. (2011). The production and processing of determiner–noun agreement in child L2 Dutch. *Linguistic Approaches to Bilingualism*, *1*(3), 265-290.

Bock, K. (1995). Producing agreement. *Current Directions in Psychological Science*, 4(2), 56-61.

Blumenfeld, H. K., & Marian, V. (2007). Constraints on parallel activation in bilingual spoken language processing: Examining proficiency and lexical status using eye-tracking. *Language and cognitive processes*, 22(5), 633-660.

de Bruin, A., Della Sala, S., & Bak, T. H. (2016). The effects of language use on lexical processing in bilinguals. *Language, Cognition and Neuroscience*, 31(8), 967-974

Chan, A. Y. (2004). Syntactic transfer: Evidence from the interlanguage of Hong Kong Chinese ESL learners. *The Modern Language Journal*, 88(1), 56-74.

Choi, Y., & Trueswell, J. C. (2010). Children’s (in) ability to recover from garden paths in a verb-final language: Evidence for developing control in sentence processing. *Journal of Experimental Child Psychology*, *106*(1), 41-61.

Chomsky, N. (1995). The Minimalist Program. Cambridge, MA: MIT Press.

Chondrogianni, V., & Marinis, T. (2012). Production and processing asymmetries in the acquisition of tense morphology by sequential bilingual children. *Bilingualism: Language and Cognition*, *15*(1), 5-21.

Chondrogianni, V. & Marinis, T. (2016). L2 children do not fluctuate: Production and on-line processing of indefinite articles in Turkish-speaking child learners of English. In: Haznedar, B. & Ketrez, F. N. (Eds.). *The Acquisition of Turkish in Childhood*. John Benjamins. [*Trends in Language Acquisition Research 20*], 361-388.

Chondrogianni, V., Marinis, T., Edwards, S., & Blom, E. (2015a). Production and on-line comprehension of definite articles and clitic pronouns by Greek sequential bilingual children and monolingual children with specific language impairment. Applied Psycholinguistics, 36(5), 1155-1191.

Chondrogianni, V., Vasić, N., Marinis, T., & Blom, E. (2015b). Production and on-line comprehension of definiteness in English and Dutch by monolingual and sequential bilingual children. *Second language research*, *31*(3), 309-341.

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive psychology*, 42(4), 368-407.

Clahsen, H., & Felser, C. (2006a). How native-like is non-native language processing? *Trends in Cognitive Sciences,* 10,564–570.

Clahsen, H., & Felser, C. (2006b). Grammatical processing in language learners. *Applied Psycholinguistics*, 27, 3–42.

Clahsen, H., & Felser, C. (2006c). Continuity and shallow structures in language processing. *Applied Psycholinguistics*, 27(1), 107-126.

Clahsen, H., & Felser, C. (2018). Some notes on the shallow structure hypothesis. *Studies in Second Language Acquisition*, 40(3), 693-706.

Colomé, À. (2001). Lexical activation in bilinguals' speech production: Language-specific or language-independent?. *Journal of memory and language*, 45(4), 721-736.

Contemori, C., Carlson, M., & Marinis, T. (2018). On-line processing of English which-questions by children and adults: a visual world paradigm study. *Journal of child language*, 45(2), 415-441.

Contemori, C. & Marinis, T. (2014). The impact of number mismatch and passives on the real-time processing of relative clauses. *Journal of Child Language* 41(3), 658–89.

Costa, A., & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of memory and Language*, 50(4), 491-511.

Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The cognate facilitation effect: implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(5), 1283.

Crain, S., & Fodor, J. D. (1985). How can grammars help parsers? In DR Doaety, Lauri Karttunen, & Arnold M. Zicky (Eds.) *Natural Language Parsing: Psychological, Computational, and Theoretical Perspectives*, 94-128.

Cunnings, I. (2017). Parsing and working memory in bilingual sentence processing. *Bilingualism: Language and Cognition*, 20(4), 659-678.

Cunnings, I., & Felser, C. (2014). Plausibility and filled gap effects in native and non-native sentence processing. Poster presented at AMLaP 2014, University of Edinburgh, UK.

Cunnings, I., Batterham, C., Felser, C., & Clahsen, H. (2010). 'Constraints on L2 learners' processing of wh- dependencies: Evidence from eye-movements' in VanPatten, B. & Jegerski, J. (Eds.), *Second Language Processing and Parsing I*, pp 87–110. Amsterdam, John Benjamins.

Dallas, A., & Kaan, E. (2008). Second Language Processing of Filler-Gap Dependencies by Late Learners. *Language and Linguistics Compass*, 2(3), 372-388.

Dallas, A., DeDe, G., & Nicol, J. (2013). An Event-Related Potential (ERP) Investigation of Filler-Gap Processing in Native and Second Language Speakers. *Language Learning*, *63*(4), 766-799.

De Cat, C., Gusnanto, A., & Serratrice, L. (2018). Identifying a threshold for the executive function advantage in bilingual children. *Studies in Second Language Acquisition*, 40(1), 119-151.

Dekydtspotter, L., Sprouse, R. A., & Anderson, B. (1997). The interpretive interface in L2 acquisition: The process-result distinction in English-French interlanguage grammars. *Language Acquisition*, 6(4), 297-332.

Dekydtspotter, L., Sprouse, R. A., & Swanson, K. A. (2001). Reflexes of mental architecture in second-language acquisition: The interpretation of combien extractions in English-French interlanguage. *Language Acquisition*, 9(3), 175-227.

Dekydtspotter, L., Sprouse, R.A., & Thyre, R. (2001). The Interpretation of Quantification at a Distance in English-French Interlanguage: Domain specificity and second-language acquisition. Language Acquisition 8(4), 265–320.

Dekydtspotter, L., Schwartz, B. D., & Sprouse, R. A. (2006). The comparative fallacy in L2 processing research. In *Proceedings of the 8th generative approaches to second language acquisition conference (GASLA 2006): The Banff conference* (pp. 33-40).

de Houwer, A. (1995). Bilingual language acquisition. In P. Fletcher & B. MacWhinney (Eds.), The handbook of child language. (pp. 219–250). Oxford: Blackwell.

De Houwer, A. (2009). *Bilingual first language acquisition*. Multilingual Matters.

DeLuca, V., Rothman, J., Bialystok, E., & Pliatsikas, C. (2019). Redefining bilingualism as a spectrum of experiences that differentially affects brain structure and function. *Proceedings of the National Academy of Sciences*, 116(15), 7565-7574.

Desmet, T., & Declercq, M. (2006). Cross-linguistic priming of syntactic hierarchical configuration information. *Journal of Memory and Language*, 54(4), 610-632.

de Villiers, J. G., H. B. Tager-Flusberg, Hakuta, K. & Cohen, M. (1979). Children’s comprehension of relative clauses. *Journal of Psycholinguistic Research* 8(5), 499-518.

De Vincenzi, M., Arduino, L. S., Ciccarelli, L. & Job, R. (1999). Parsing strategies in children: comprehension of interrogative sentences. In S. Bagnara, (Ed.), *Proceedings of European Conference on Cognitive Science* (pp. 301–308). Rome: Instituto di Psicologia del CNR.

Deevy, P. & Leonard, L. (2004). The comprehension of Wh-questions in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 47, 802–15.

Diessel, H. (2004). *The acquisition of complex sentences*. Cambridge: Cambridge University Press.

Dijkstra, T., Grainger, J., & Van Heuven, W. J. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and language*, 41(4), 496-518.

Dillon, B., Staub, A., Levy, J., & Clifton Jr, C. (2017). Which noun phrases is the verb supposed to agree with?: Object agreement in American English. *Language*, 93(1), 65-96.

Donkers, J., Hoeks, J. C., & Stowe, L. A. (2013). D-linking or set-restriction? Processing Which-questions in Dutch. *Language and Cognitive Processes*, 28(1-2), 9-28.

Dryer, S.M. (2013a). Position of Interrogative Phrases in Content Questions. In: Dryer, Matthew S. & Haspelmath, Martin (eds.) *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://wals.info/chapter/93, Accessed on 2019-07-02.)

Dryer, S.M. (2013b). Order of Relative Clause and Noun. In: Dryer, Matthew S. & Haspelmath, Martin (eds.) *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://wals.info/chapter/90, Accessed on 2019-05-31.)

Dussias, P. E. (2003). Syntactic ambiguity resolution in L2 learners: Some effects of bilinguality on L1 and L2 processing strategies. *Studies in Second Language Acquisition*, 25(4), 529-557.

Dussias, P. E. (2004). Parsing a first language like a second: The erosion of L1 parsing strategies in Spanish-English bilinguals. *International Journal of Bilingualism*, 8(3), 355-371.

Dussias P. E., & Pinar P. (2010). Effects of reading span and plausibility in the reanalysis of wh-gaps by Chinese-English second language speakers. Second Language Research, 26, 443–472.

Eckman, Fred R. 1977. Markedness and the Contrastive Analysis Hypothesis. *Language Learning* 27 (2). 315-330.

Ervin-Tripp, Susan (1970) “Discourse Agreement: How Children Answer Questions”, in J.R. Hayes (ed.) *Cognition and the Development of Language* (pp.79-106). Wiley: New York.

Epstein, S. D., Flynn, S., & Martohardjono, G. (1996). Second language acquisition: Theoretical and experimental issues in contemporary research. *Behavioral and Brain Sciences*, 19(4), 677-714.

Felser, C., Cunnings, I., Batterham, C., & Clahsen, H. (2012). The timing of island effects in nonnative sentence processing. *Studies in Second Language Acquisition*, 34(1), 67-98.

Felser, C., Roberts, L., Marinis, T., & Gross, R. (2003). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics*, 24(3), 453-489

Felser, C., Marinis, T., & Clahsen, H. (2003). Children's processing of ambiguous sentences: A study of relative clause attachment. *Language Acquisition*, 11(3), 127-163.

Felser, C., Roberts, L., Gross, R., & Marinis, T. (2003). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics*, 24, 453-489.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive psychology*, 47(2), 164-203.

Ferreira, F., & Clifton Jr, C. (1986). The independence of syntactic processing. *Journal of memory and language*, 25(3), 348-368.

Ferreira, F., & Henderson, J. M. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language*, 30(6), 725-745.

Fodor, J. A., Bever T.G., & Garrett, M. (1974). *The psychology of Language: An introduction to Psycholinguistics and Generative Grammar*. New York: MCRAW-Hill.

Fodor, J. D., & Ferreira, F. (1998). *Reanalysis in sentence processing*. Dordrecht, The Netherlands: Kluwer

Fodor, J., & Frazier, L. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, 6, 291-325.

Fodor, J.D. & Inoue, A. (1994). The diagnosis and cure of garden paths. *Journal of Psycholinguistic Research*, 23, 407-434

Franck, J., Colonna, S., & Rizzi, L. (2015). Task-dependency and structure-dependency in number interference effects in sentence comprehension. *Frontiers in psychology*, 6, 349.

Frazier, L. & Clifton, C. Jr. (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes* 4(2), 93–126.

Frazier, L. & Flores d'Arcais, G. B. (1989). Filler-driven parsing: a study of gap filling in Dutch. *Journal of Memory and Language* 28, 331–344.

Frazier, L. (1987a). Sentence processing: A tutorial. In M. Coltheart (Ed.), *Attention & performance XII* (pp. 559-585). Hillsdale, NJ: Erlbaum.

Frazier, L. (1987b). Syntactic processing: evidence from Dutch. *Natural Language & Linguistic Theory*, 5(4), 519-559.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psycholog*y, 14(1), 78-210.

Frazier, L., & Rayner, K. (1990). Taking on semantic commitments: Processing multiple meanings vs. multiple senses. *Journal of memory and language*, 29(2), 181-200.

Frenck-Mestre C., & Pynte J. (1997). Syntactic ambiguity resolution while reading in second and native languages. *Quarterly Journal of Experimental Psychology*, 50A, 119 -148.

Friedmann, N., & Novogrodsky, R. (2004). The acquisition of relative clause comprehension in Hebrew: A study of SLI and normal development. *Journal of Child Language*, 31, 661-681.

Friedmann, N., Belletti, A. & Rizzi, L. (2009). Relativized relatives: types of intervention in the acquisition of A-bar dependencies, *Lingua* 119(1), 67–88.

Garrett, M. (2000). Remarks on the architecture of language processing systems. In Y. Grodzinsky, & L. Shapiro (Eds.), *Language and the Brain: Representation and Processing* (pp. 31-69). San Diego, CA: Academic Press.

Gavruseva, E. (2002). Is there primacy of aspect in child L2 English?. *Bilingualism: Language and Cognition*, 5(2), 109-130.

Genesee, Fred & Elena Nicoladis. 2007. Bilingual first-language acquisition. In E. Hoff & P. McCardle (eds.), *Handbook of language development* (pp. 324–344). Oxford: Blackwell.

Gertner, Y., Fisher, C., & Eisengart, J. (2006). Learning words and rules: Abstract knowledge of word order in early sentence comprehension. *Psychological science*, 17(8), 684-691.

Gerth, S., Otto, C., Felser, C., & Nam, Y. (2017). Strength of garden-path effects in native and non-native speakers' processing of object–subject ambiguities. *International Journal of Bilingualism*, *21*(2), 125-144.

Gibson, E. (1998). Syntactic complexity: Locality of syntactic dependencies. *Cognition*, 68, 1–75.

Gibson, E., & Warren, T. (2004). Reading-Time Evidence for Intermediate Linguistic Structure in Long-Distance Dependencies. *Syntax*, 7(1), 55-78.

Goodluck, H. (2005). D(iscourse)-linking and question formation: comprehension effects in children and Broca's aphasics. In Di Sciullo, A. M. (eds.), *UG and external systems: language, brain and computation* (pp. 185–192). Amsterdam: John Benjamins Publishing Company.

Goodluck, H. (2010). Object extraction is not subject to Child Relativized Minimality. *Lingua*, *120*(6), 1516-1521.

Gollan, T. H., Slattery, T. J., Goldenberg, D., Van Assche, E., Duyck, W., & Rayner, K. (2011). Frequency drives lexical access in reading but not in speaking: The frequency-lag hypothesis. *Journal of Experimental Psychology: General*, 140(2), 186.

Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of experimental psychology: learning, memory, and cognition*, 27(6), 1411.

Gordon, P. C., Hendrick, R., & Levine, W. H. (2002). Memory-load interference in syntactic processing. *Psychological science*, 13(5), 425-430.

Gordon, P., Hendrick, R. & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language* 51, 97–114.

Green, D. W. (1986). Control, activation, and resource: A framework and a model for the control of speech in bilinguals. *Brain and language*, 27(2), 210-223.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and cognition*, 1(2), 67-81.

Grey, S., Sanz, C., Morgan-Short, K., & Ullman, M. T. (2018). Bilingual and monolingual adults learning an additional language: ERPs reveal differences in syntactic processing. *Bilingualism: Language and Cognition*, *21*(5), 970-994.

Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentenial complexity. *Cognitive science*, 29(2), 261-290.

Grosjean, F. (1997). Processing mixed language: Issues, findings, and models. *Tutorials in bilingualism: Psycholinguistic perspectives*, 225-254.

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and cognition*, 1(2), 131-149.

Grosjean, F. (2006). The bilingual is not two monolinguals in one person. *World Englishes: Critical concepts in linguistics*, *4*, 202-213.

Jacob, G., & Felser, C. (2016). Reanalysis and semantic persistence in native and non-native garden-path recovery. *The Quarterly Journal of Experimental Psychology*, *69*(5), 907-925.

Hale, J. (2001). A probabilistic early parser as a psycholinguistic model. In *Proceedings of NAACL* (Vol. 2, pp. 159–166).

Harley, H., & Ritter, E. (2002b). Structuring the bundle: A universal morphosyntactic feature geometry. In H. Simon & H. Weise (Eds.), *Pronouns-grammar and representation* (pp. 23–39). Amsterdam, Netherlands: John Benjamins.

Hartsuiker, R. J., Pickering, M. J., & Veltkamp, E. (2004). Is syntax separate or shared between languages? Cross-linguistic syntactic priming in Spanish-English bilinguals. *Psychological science*, 15(6), 409-414.

Haspelmath, M. (2006). Against markedness (and what to replace it with). *Journal of linguistics*, *42*(1), 25-70.

Hawkins, R., & Chen, C. Y. H. (1997). The partial availability of Universal Grammar in second language acquisition: The ‘failed functional features hypothesis’. *Second Language Research*, 13(3), 187-226.

Hawkins, R., & Hattori, H. (2006). Interpretation of English multiple wh-questions by Japanese speakers: A missing uninterpretable feature account. *Second Language Research*, 22(3), 269-301.

Haznedar, B. (2001). The acquisition of the IP system in child L2 English. *Studies in second language acquisition*, 23(1), 1-39.

Haznedar, B. and Schwartz, B.D. 1997: Are there optional infinitives in child L2 acquisition? In Hughes, E., Hughes, M. and Greenhill, A., (Eds.), *Proceedings of the 21st Annual BUCLD*. Somervillle, MA: Cascadilla Press, 257–68.

Hernandez, A., Bates, E. & Avila, L. (1994). On-line sentence interpretation in Spanish–English bilinguals: What does it mean to be "in between"? *Applied Psycholinguistics*, 15, 417–446.

Hickok, G., & Avrutin, S. (1996). Comprehension of wh-questions in two Broca's aphasics. *Brain and language*, 52(2), 314-327.

Hirsch, C. & Wexler, K. (2006). Children's passives and their resulting interpretation. *The Proceedings of GALANA 1, UConn Occasional Papers in Linguistics* 4: 125-36.

Hoff, E. (2013). Interpreting the early language trajectories of children from low-SES and language minority homes: implications for closing achievement gaps. *Developmental psychology*, *49*(1), 4-14.

Hopp H. (2006). Syntactic features and reanalysis in near-native processing. *Second Language Research*, 22, 369 –397.

Hopp, H. (2010). Ultimate attainment in L2 inflectional morphology: Performance similarities between non-native and native speakers. *Lingua*,120, 901–931.

Hopp, H. (2014). Working memory effects in the L2 processing of ambiguous relative clauses. *Language Acquisition*, 21(3), 250-278.

Hopp, H. (2015). Individual differences in the second language processing of object–subject ambiguities. *Applied Psycholinguistics*, *36*(2), 129-173.

Hopp, H. (2017). The processing of English which-questions in adult L2 learners: Effects of L1 transfer and proficiency. *Zeitschrift für Sprachwissenschaft*, *36*(1), 107-134.

Hurewitz, F., Brown-Schmidt, S., Thorpe, K., Gleitman, L. R., & Trueswell, J. C. (2000). One frog, two frog, red frog, blue frog: Factors affecting children's syntactic choices in production and comprehension. *Journal of psycholinguistic research*, *29*(6), 597-626.

Incera, S., & McLennan, C. T. (2018). Bilingualism and age are continuous variables that influence executive function. *Aging, Neuropsychology, and Cognition*, 25(3), 443-463.

Jacob, G., & Felser, C. (2016). Reanalysis and semantic persistence in native and non-native garden-path recovery. *The Quarterly Journal of Experimental Psychology, 69*(5), 907-925.

Jared, D., & Kroll, J. F. (2001). Do bilinguals activate phonological representations in one or both of their languages when naming words? *Journal of memory and language*, 44(1), 2-31.

Jegerski J. (2012). The processing of temporary subject-object ambiguities in native and near-native Mexican Spanish. *Bilingualism: Language and Cognition*, 15, 721 –735.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive psychology*, 21(1), 60-99.

Juffs, A. (1998). Main verb versus reduced relative clause ambiguity resolution in L2 sentence processing. *Language Learning*, 48, 107-147.

Juffs A. (2005). The influence of first language on the processing of wh -movement in English as a second language. *Second Language Research*, 21 (2), 121–151.

Juffs A., & Harrington M. (1995). Parsing effects in second language sentence processing. *Studies in Second Language Acquisition*, 17 (4), 483 –516.

Just, M & Carpenter, P. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99, 122-149.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and cognitive processes*, 15(2), 159-201.

Kaltsa, M., Tsimpli, I. M., Marinis, T., & Stavrou, M. (2016). Processing coordinate subject-verb agreement in L1 and L2 Greek. *Frontiers in psychology*, 7, 648.

Karimi, H., & Ferreira, F. (2016). Good-enough linguistic representations and online cognitive equilibrium in language processing. *The Quarterly Journal of Experimental Psychology*, 69(5), 1013-1040.

Kidd, E., & Bavin, E. L. (2005). Lexical and referential cues to sentence interpretation: An investigation of children's interpretations of ambiguous sentences. *Journal of Child Language*, *32*(4), 855-876.

Kidd, E., Brandt, S., Lieven, E., & Tomasello, M. (2007). Object relatives made easy: A cross-linguistic comparison of the constraints influencing young children's processing of relative clauses. *Language and cognitive processes*, 22(6), 860-897.

Kidd, E., Stewart, A. J., & Serratrice, L. (2011). Children do not overcome lexical biases where adults do: The role of the referential scene in garden-path recovery. *Journal of Child Language*, *38*(1), 222-234.

Kim, E., Baek, S., & Tremblay, A. (2015). The role of island constraints in second language sentence processing. *Language Acquisition*, *22*(4), 384-416.

King, J.W. & Just, M.A. (1991). Individual differences in syntactic processing: The role of working Memory. *Journal of Memory and Language*, 30, 580–602.

Lapata, M., Keller, F., & Im Walde, S. S. (2001). Verb frame frequency as a predictor of verb bias. *Journal of Psycholinguistic Research*, 30(4), 419-435.

Lardiere, D. (1998). Case and tense in the ‘fossilized’steady state. *Second language research*, 14(1), 1-26.

Lardiere, D. (2017). *Ultimate attainment in second language acquisition: A case study*. Routledge.

Lardiere, D. (2009). Some thoughts on the contrastive analysis of features in second language acquisition. *Second language research*, 25(2), 173-227.

Lemmerth, N., & Hopp, H. (2019). Gender processing in simultaneous and successive bilingual children: cross-linguistic lexical and syntactic influences. *Language Acquisition*, 26(1), 21-45.

Lempert, H. & Kinsbourne, M. (1980). Preschool children's sentence comprehension: Strategies with respect to word order. *Journal of Child Language*, 7, 371-79.

Levy, R. (2008). Expectation-based syntactic comprehension. Cognition 106, 1126-77.

Lew-Williams, C. (2017). Specific Referential Contexts Shape Efficiency in Second Language Processing: Three Eye-Tracking Experiments With 6-and 10-Year-Old Children in Spanish Immersion Schools. *Annual Review of Applied Linguistics*, 37, 128-147.

Li, P. (1996). Spoken word recognition of code-switched words by Chinese–English bilinguals. *Journal of memory and language*, 35(6), 757-774.

Loebell, H., & Bock, K. (2003). Structural priming across languages. *Linguistics*, 41(5; ISSU 387), 791-824.

Love, T. E. (2007). The processing of non-canonically ordered constituents in long distance dependencies by pre-school children: A real-time investigation. *Journal of Psycholinguistic research*, 36(3), 191-206.

Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology*, *25*(5), 605-621.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101(4), 676-703.

MacKay, I. R., & Flege, J. E. (2004). Effects of the age of second language learning on the duration of first and second language sentences: The role of suppression. *Applied Psycholinguistics*, 25(3), 373-396.

MacWhinney, B. J. (2001). Psycholinguistics: Overview. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social & behavioural sciences* (Vol. 18, pp. 12343–12349). Amsterdam: Elsevier.

MacWhinney, B. & Bates, E. (1989). The *crosslinguistic study of sentence processing*. New York: Cambridge University Press.

Marian, V., & Spivey, M. (2003). Competing activation in bilingual language processing: Within-and between-language competition. *Bilingualism: Language and Cognition*, 6(2), 97-115.

Marinis, T. (2003). Psycholinguistic techniques in second language acquisition research. *Second language research*, 19(2), 144-161.

Marinis, T. (2007). On-line processing of passives in L1 and L2 children. In Belikova, A., Meroni, L. & Umeda, M. (eds.). *Proceedings of the 2nd conference on generative approaches to language acquisition North America (GALANA)* (pp. 265-276). Somerville, MA: Cascadilla Proceedings Project.

Marinis, T. (2008). On-line processing of sentences involving reflexive and non-reflexive pronouns in L1 and L2 children. In: Gavarró, A. & Freitas, M. J. (Eds.) *Language acquisition and development. Proceedings of GALA 2007*. Cambridge Scholars Publishing, 348-358.

Marinis, T. (2010). On-line sentence processing methods in typical and atypical populations. In Unsworth, S., & Blom, E (Eds.) *Experimental methods in language acquisition research [Language Learning and Language Teaching]* (Vol. 27, pp. 139-162). John Benjamins, Amsterdam/Philadelphia.

Marinis, T. (2013). Online Psycholinguistic Methods in Second Language Acquisition. In: Chapelle, C. A. (Ed). *The Encyclopedia of Applied Linguistic*s. Wiley-Blackwell, 4307–4315.

Marinis, T. (2015). Providing some more pieces to the puzzle: L2 adults, L2 children and children with specific language impairment. *Linguistic Approaches to Bilingualism*, 5(4), 499-504.

Marinis, T. (2018). Cross-modal priming in bilingual sentence processing. *Bilingualism: Language and Cognition*, 21(3), 456-461.

Marinis, T. & Cunnings, I. (2018). Using psycholinguistic techniques in a second language teaching setting. In: Wright, C., Piske, T., & Young-Scholten, M. (Eds.). *Mind Matters in SLA. (Second language acquisition; 26)*. pp. 185-202. Bristol, Multilingual Matters.

Marinis, T., Roberts, L., & Felser, C. (2003). The role of lexical biases in relative clause ambiguity resolution. Unpublished manuscript, University of Essex.

Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language sentence processing. *Studies in Second Language Acquisition*, *27*(1), 53-78.

Marinis, T., & Saddy, D. (2013). Parsing the passive: Comparing children with specific language impairment to sequential bilingual children. *Language Acquisition*, *20*(2), 155-179.

Marinis, T., & van der Lely, H. K. (2007). On-line processing of wh-questions in children with G-SLI and typically developing children. *International Journal of Language & Communication Disorders*, 42(5), 557-582.

Martin, C. D., Costa, A., Dering, B., Hoshino, N., Wu, Y. J., & Thierry, G. (2012). Effects of speed of word processing on semantic access: The case of bilingualism. *Brain and language*, 120(1), 61-65.

Matin, E., Shao, K. & Boff, K. (1993). Saccadic overhead: information processing time with and without saccades. *Perception & Psychophysics* 53, 372–80.

McDonald, J. L. (2000). Grammaticality judgments in a second language: Influences of age of acquisition and native language. *Applied psycholinguistics,* 21(3), 395-423.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological review*, 101(4), 676-703.

Meijer, P. J., & Fox Tree, J. E. (2003). Building syntactic structures in speaking: A bilingual exploration. *Experimental Psychology*, 50(3), 184.

Meisel, J. M. (1997). The acquisition of the syntax of negation in French and German. Contrasting first and second language development. *Second Language Research*, 13(3), 227-263.

Meisel, J. M. (2009). Second language acquisition in early childhood. *Zeitschrift für Sprachwissenschaft*, 28(1), 5-34.

Meroni, L., & Crain, S. (2003). On not being led down the kindergarten path. In *Proceedings of the 27th Boston University Conference on language development* (pp. 531-544). Somerville, MA: Cascadilla Press.

Mirman, D. (2014). *Growth Curve Analysis and Visualization Using R*. Chapman and Hall / CRC.

Mirman, D., Dixon, J.A., & Magnuson, J.S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, 59(4), 475-494.

Nakamura, C., Arai, M., & Mazuka, R. (2012). Immediate use of prosody and context in predicting a syntactic structure. *Cognition*, 125(2), 317-323.

Navarro-Torres, C. A., Garcia, D. L., Chidambaram, V., & Kroll, J. F. (2019). Cognitive Control Facilitates Attentional Disengagement during Second Language Comprehension. *Brain sciences*, *9*(5), 95.

Novick, J. M., Kan, I. P., Trueswell, J. C., & Thompson-Schill, S. L. (2009). A case for conflict across multiple domains: memory and language impairments following damage to ventrolateral prefrontal cortex. *Cognitive neuropsychology*, 26(6), 527-567.

Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cognitive, Affective, & Behavioral Neuroscience*, 5(3), 263-281.

Omaki, A., & Schulz, B. (2011). Filler-gap dependencies and island constraints in second-language sentence processing. *Studies in Second Language Acquisition*, *33*(4), 563-588.

Omaki, A., Davidson White, I., Goro, T., Lidz, J., & Phillips, C. (2014). No fear of commitment: Children's incremental interpretation in English and Japanese wh-questions. *Language Learning and Development*, 10(3), 206-233.

Osterhout, L., Holcomb, P. J., & Swinney, D. A. (1994). Brain potentials elicited by garden-path sentences: evidence of the application of verb information during parsing. *Journal of experimental psychology: Learning, memory, and cognition*, 20(4), 786.

Ouhalla, J. (1994). *Introducing transformational grammar: From principles and parameters to minimalism*. London: Arnold.

Pan, H. Y., & Felser, C. (2011). Referential context effects in L2 ambiguity resolution: Evidence from self-paced reading. *Lingua*, 121(2), 221-236.

Pan, H. Y., Schimke, S., & Felser, C. (2015). Referential context effects in non-native relative clause ambiguity resolution. *International Journal of Bilingualism*, 19(3), 298-313.

Papadopoulou, D., & Clahsen, H. (2003). Parsing strategies in L1 and L2 sentence processing: A study of relative clause attachment in Greek. *Studies in Second Language Acquisition*, 25, 501-528.

Papangeli, A. & Marinis, T. (2010). Επεξεργασία δομικά αμφίσημων προτάσεων στην Ελληνική ως Γ1 και ως Γ2. [Processing of structurally ambiguous sentences in Greek as L1 and L2]. In: *Proceedings of the Annual Meeting of the Department of Linguistics, School of Philology, Faculty of Philosophy, Aristotle University of Thessaloniki, May 2–3, 2009*. Thessaloniki: Institute of Modern Greek Studies [Institute Manoli Triantafillidi], 477-486.

Phillips, C., Kazanina, N., & Abada, S. H. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research*, 22(3), 407-428.

Pickering, M. & Barry, G. (1991). Sentence processing without empty categories. *Language and Cognitive Processes* 6, 229-259.

Pliatsikas, C., & Marinis, T. (2013a). Rule-based morphological processing in a second language: a behavioural investigation. Language Studies Working Papers, 5(1), 3-12.

Pliatsikas, C., & Marinis, T. (2013b). Processing of regular and irregular past tense morphology in highly proficient second language learners of English: A self-paced reading study. *Applied Psycholinguistics*, 34(5), 943-970.

Pliatsikas, C., & Marinis, T. (2013c). Processing empty categories in a second language: When naturalistic exposure fills the (intermediate) gap. *Bilingualism: Language and Cognition*, *16*(1), 167-182.

Pollard, C. & Sag, I.A. (1994). *Head-driven phrase structure grammar*. Chicago: University of Chicago Press.

Pozzan, L., & Trueswell, J. C. (2016). Second language processing and revision of garden-path sentences: a visual word study. *Bilingualism: Language and Cognition*, *19*(3), 636-643.

Prévost, P., & White, L. (2000). Missing surface inflection or impairment in second language acquisition? Evidence from tense and agreement. *Second language research*, 16(2), 103-133.

Proverbio, A. M., Čok, B., & Zani, A. (2002). Electrophysiological measures of language processing in bilinguals. *Journal of Cognitive Neuroscience*, 14(7), 994-1017.

Radford, A. 2004. *English syntax: An introduction*. Cambridge, UK: Cambridge University Press.

Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(6), 1290.

Rizzi, L. (1990). *Relativized minimality*. The MIT Press.

Rizzi, L. (2004). Locality and the left periphery. In A. Belletti (ed.), *Structures and beyond: the cartography of syntactic structures* (3) 223–251. New York: Oxford University Press.

Roberts, P. M., Garcia, L. J., Desrochers, A., & Hernandez, D. (2002). English performance of proficient bilingual adults on the Boston Naming Test. *Aphasiology*, 16(4-6), 635-645.

Roberts, L., & Felser, C. (2011). Plausibility and recovery from garden paths in second language sentence processing. *Applied Psycholinguistics*, *32*(2), 299-331.

Roberts, L., Marinis, T., Felser, C., & Clahsen, H. (2007). Antecedent priming at trace positions in children's sentence processing. *Journal of psycholinguistic research*, 36(2), 175-188.

Roesch, A. D., & Chondrogianni, V. (2016). "Which mouse kissed the frog?" Effects of age of onset, length of exposure, and knowledge of case marking on the comprehension of wh-questions in German-speaking simultaneous and early sequential bilingual children. *Journal of child language*, *43*(3), 635-661.

Roland, D., Dick, F. & Elman, J. L. (2007). Frequency of English grammatical structures: a corpus analysis. *Journal of Memory and Language*, 57, 348-79.

Sag, I.A, & Fodor, J.D. (1995). Extraction without traces. In R. Aranovich, W. Byrne, S. Preuss, & M. Senturia, (eds.), *Proceedings of the 13th annual meeting of the West Coast Conference on Formal Linguistics*. (pp. 365–384). Stanford, CA: CSLI Publications.

Salis, C., & Edwards, S. (2008). Comprehension of wh-questions and declarative sentences in agrammatic aphasia: The set partition hypothesis. *Journal of Neurolinguistics*, 21(5), 375-399.

Schlesewsky, M., Fanselow, G., Kliegl, R. & Krems, J. (2000). The subject preference in the processing of locally ambiguous wh-questions in German. In B. Hemforth & L. Konieczny (eds), *German sentence processing*, 65–93. Dordrecht: Kluwer.

Schneider, W., Eschman, A. & Zuccolotto, A. (2002). *E-Prime user's guide*. Pittsburgh, PA: University of Pittsburgh Psychology Software Tools, Inc.

Schoonbaert, S., Hartsuiker, R. J., & Pickering, M. J. (2007). The representation of lexical and syntactic information in bilinguals: Evidence from syntactic priming. *Journal of Memory and Language*, 56(2), 153-171.

Schwartz, B. (2006). Invited plenary speaker. Illuminating child L2 developmental puzzles. *Generative Approaches to Language Acquisition North America (GALANA) 2*, McGill University, Montreal.

Schwartz, B. D., & Sprouse, R. A. (1996). L2 cognitive states and the full transfer/full access model. *Second language research,* 12(1), 40-72.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological review*, 96(4), 523.

Serratrice, L. (2018) Becoming bilingual in early childhood. In: De Houwer, A. and Ortega, L. (eds.) *The Cambridge Handbook of Bilingualism*. Cambridge University Press, Cambridge, UK. ISBN 9781316831922, pp 15-35.

Slabakova, R. (2003). Semantic evidence for functional categories in interlanguage grammars. *Second Language Research*, 19(1), 42-75.

Slabakova, R. (2009). What is Easy and What is Hard to Acquire in a Second Language? In In M. Bowles, T. Ionin, S. Montrul & A. Tremblay (eds), *Proceedings of the 10th Generative Approaches to Second Language Acquisition Conference (GASLA 2009)*. Cascadilla Proceedings Project: Somerville, USA, pp. 280-294.

Slattery, T. J., Sturt, P., Christianson, K., Yoshida, M., & Ferreira, F. (2013). Lingering misinterpretations of garden path sentences arise from competing syntactic representations. *Journal of Memory and Language*, *69*(2), 104-120.

Snedeker, J., & Yuan, S. (2008). Effects of prosodic and lexical constraints on parsing in young children (and adults). *Journal of memory and language*, 58(2), 574-608.

Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive psychology*, 49(3), 238-299.

Sorace, A. (2011). Pinning down the concept of 'interface' in bilingualism. *Linguistic Approaches to Bilingualism*, 1, 1– 33.

Sorace, A., & Filiaci, F. (2006). Anaphora resolution in near-native speakers of Italian. *Second Language Research*, 22(3), 339-368.

Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive psychology*, 45(4), 447-481.

Staub, A. (2010). Eye movements and processing difficulty in object relative clauses. *Cognition* 116, 71–86.

Stowe, A. (1986). Parsing wh-constructions: evidence for on-line gap location. *Language and Cognitive Processes* 1, 227-45.

Stromswold, K. (1995a). The Acquisition of Subject and Object Wh-Questions. Language *Acquisition* 4(1/2), 5-48.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217), 1632-1634.

Teubner-Rhodes, S. E., Mishler, A., Corbett, R., Andreu, L., Sanz-Torrent, M., Trueswell, J. C., & Novick, J. M. (2016). The effects of bilingualism on conflict monitoring, cognitive control, and garden-path recovery. *Cognition*, *150*, 213-231.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of memory and language*, 33(3), 285-318.

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(3), 528.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, 73(2), 89-134.

Tsimpli, I. M. (2003). Clitics and articles in L2 Greek. In Liceras, J. M., Zobl, H. & Goodluck, H. (eds.), *Proceedings of the 6th Generative Approaches to Second Language Acquisition Conference (GASLA 2002)*, pp. 331–339. Somerville, MA: Cascadilla Proceedings Project.

Tsimpli, I. M. (1997). Resumptive features and L2A: a minimalist account. In *Boston University Conference on Language Development* (Vol. 21, pp. 639-55).

Tsimpli, I. M., & Dimitrakopoulou, M. (2007). The interpretability hypothesis: Evidence from wh-interrogatives in second language acquisition. *Second Language Research*, 23(2), 215-242.

Tsimpli, I. M., & Mastropavlou, M. (2008). Feature interpretability in L2 acquisition and SLI: Greek clitics and determiners. In Goodluck, H., Liceras, J. & Zobl, H. (eds.), *The role of formal features in second language acquisition*, pp. 143–183. London: Routledge.

Ullman, M. (2001). The neural basis of lexicon and grammar: The declarative/procedural model. *Bilingualism: Language and Cognition*, 4, 105–122.

Ullman, M. (2005). A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model. In Sanz, C. (ed.), *Mind and Context in Adult Second Language Acquisition*, pp. 141–178. Washington, Georgetown University Press.

Ullman, M. (2015). The declarative/procedural model: A neurobiologically motivated theory of first and second language. In Van Patten, B. & Williams, J. (eds.), *Theories in Second Language Acquisition: An Introduction* (2nd edition). New York: Routledge.

Unsworth, S. (2005). *Child L2, adult L2, child L1: Differences and similarities*. Unpublished PhD Dissertation, Utrecht University.

Van Dyke, J. & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language* 55, 157-66.

Vainikka, A., & Young-Scholten, M. (1994). Direct access to X'-theory: evidence from Korean and Turkish adults learning German. *Language acquisition studies in generative grammar*, 31(4), 71-89.

Vainikka, A., & Young-Scholten, M. (1996a). *Gradual development of L2 phrase structure. Second language research*, 12(1), 7-39.

Vainikka, A., & Young-Scholten, M. (1996). *The early stages in adult L2 syntax: Additional evidence from Romance speakers*. Second Language Research, 12(2), 140-176.

Vainikka, A., & Young-Scholten, M. (1998a). The initial state in the L2 acquisition of phrase structure. In Flynn, S., Martohardjono, G. &O'Neil, W. (eds.), *The generative study of second language acquisition*, pp. 17–88. Somerville, MA: Laurence Erlbaum.

Vainikka, A., & Young-Scholten, M. (1998b). Morphosyntactic triggers in adult SLA. In Beck, M.-L. (ed.), *Morphology and its interfaces in second language knowledge*, pp. 89–114. Amsterdam: John Benjamins.

Vainikka, A., & Young-Scholten, M. (2010). All acquisition begins with the projection of a bare verb phrase. *Applied Psycholinguistics*, 31(2), 332-339.

Vasić, N., & Blom, W. B. T. (2011). Production and processing of determiners in Turkish-Dutch child L2 learners. In *BUCLD35: proceedings of the 35th annual Boston University Conference on Language Development* (pp. 616-627). Cascadilla Press.

Vasić, N., Chondrogianni, V., Marinis, T., & Blom, W. B. T. (2012). Processing of gender in Turkish-Dutch and Turkish-Greek child L2 learners. In *BUCLD36: proceedings of the 36th annual Boston University Conference on Language Development* (pp. 646-659). Cascadilla Press.

Vuong, L. C., & Martin, R. C. (2014). Domain-specific executive control and the revision of misinterpretations in sentence comprehension. *Language, Cognition and Neuroscience, 29*(3), 312-325.

Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, 61(2), 206-237.

Warren, T., White, S. J., & Reichle, E. D. (2009). Investigating the causes of wrap-up effects: Evidence from eye movements and E–Z Reader. Cognition, 111(1), 132-137.

Weighall, A. R. (2008). The kindergarten path effect revisited: Children's use of context in processing structural ambiguities. *Journal of Experimental Child Psychology*, *99*(2), 75-95.

Weerman, F., Bisschop, J., & Punt, L. (2006). L1 and L2 acquisition of Dutch adjectival inflection. *ACLC Working Papers*, 1(1), 5-36.

Werker, J. F., & Byers-Heinlein, K. (2008). Bilingualism in infancy: First steps in perception and comprehension. *Trends in cognitive sciences*, 12(4), 144-151.

White, L. (1987). Markedness and second language acquisition: The question of transfer. *Studies in Second Language Acquisition*, 9, 261-286.

White, L. (2003). Fossilization in steady state L2 grammars: Persistent problems with inflectional morphology. *Bilingualism: language and cognition*, 6(2), 129-141.

White, L. (2009). Grammatical theory: Interfaces and L2 knowledge. *The new handbook of second language acquisition*, 49-68.

Wilhelm, A. & Hanna, K. (1992). "On the Acquisition of WH- questions", *Calgary Working Papers in Linguistics* 15, 89-98.

Williams, J. N. (2006). Incremental interpretation in second language sentence processing. *Bilingualism: Language and Cognition*, 9(1), 71-88.

Williams, J. N., Möbius, P., & Kim, C. (2001). Native and non-native processing of English wh-questions: Parsing strategies and plausibility constraints. *Applied Psycholinguistics*, 22(4), 509-540.

Woodard, K., Pozzan, L., & Trueswell, J. C. (2016). Taking your own path: Individual differences in executive function and language processing skills in child learners. *Journal of experimental child psychology*, 141, 187-209.

Yoshinaga, N. (1996). *Wh-questions: a comparative study of their form and acquisition in English and Japanese*. Unpublished doctoral dissertation, University of Hawaii.