University of Reading

Online processing of which-questions in bilingual children: evidence from eye-tracking

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Pontikas, G., Cunnings, I. ORCID: https://orcid.org/0000-0002-5318-0186 and Marinis, T. (2023) Online processing of which-questions in bilingual children: evidence from eye-tracking. Journal of Child Language, 50 (5). pp. 1082-1118. ISSN 0305-0009 doi: 10.1017/S0305000922000253 Available at https://centaur.reading.ac.uk/105406/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1017/S0305000922000253

Publisher: Cambridge University Press

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

University of Reading

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Journal of Child Language (2022), 1–37
doi:10.1017/S0305000922000253

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

Online processing of which-questions in bilingual children: Evidence from eye-tracking

George PONTIKAS[1,*], Ian CUNNINGS[1] and Theodoros MARINIS[1,2]

[1]School of Psychology and Clinical Languages Sciences, University of Reading, Reading, UK
[2]Department of Linguistics, University of Konstanz, Germany
Corresponding author: George Pontikas, PhD. School of Psychology and Clinical Language Sciences University of Reading, Reading, UK. E-mail: g.pontikas@reading.ac.uk

(Received 12 November 2020; revised 14 March 2022; accepted 08 April 2022)

Abstract

An emergent debate surrounds the nature of language processing in bilingual children as an extension of broader questions about their morphosyntactic development in comparison to monolinguals, with the picture so far being nuanced. This paper adds to this debate by investigating the processing of morphosyntactically complex which–questions (e.g., Which bear is chasing the camel?) using the visual world paradigm and is the first study to examine the online processing of such questions in bilingual children. For both groups, object which-questions were more difficult than subject which-questions, due to an initial misinterpretation that needed to be reanalysed. Both groups were aided by number mismatch between the two nouns in the sentence, especially in object which-questions. Our findings are in line with previous studies that have shown a slower processing speed in bilingual children relative to monolinguals but qualitatively similar patterns.

Keywords: sentence processing; childhood bilingualism; eye-tracking; wh-questions; morphosyntax

Introduction

Research on language acquisition in bilingual children has shown that they may perform less well than monolingual children in elicitation tasks tapping morphosyntax. However, most studies showing a gap between bilingual and monolingual children's morphosyntactic abilities have employed production tasks (e.g., Paradis, 2005; Paradis, Rice, Crago & Marquis, 2008; Unsworth, 2007). Over the last decade, studies examining comprehension instead of or in addition to production have enriched the literature and have so far brought varied findings. An emerging consensus is that bilingual children have knowledge of syntax and morphology, and that differences in production are attributable to production costs (Chondrogianni & Marinis, 2012, 2016). Studies employing grammatical violation paradigms to test comprehension have revealed that bilingual children show similar processing patterns as monolingual children and are sensitive to grammatical violations (e.g., Chondrogianni & Marinis, 2012, 2016). At the same time, they have

© The Author(s), 2022. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0305000922000253 Published online by Cambridge University Press

shown that bilingual children process language at a slower rate compared to monolingual children, suggesting less efficient processing.

The existing literature on bilingual comprehension has however not provided direct insight into how bilingual children interpret sentences in real-time. A hallmark of incremental sentence interpretation is that initially assigned interpretations may ultimately turn out to be incorrect. Such misinterpretations require revision and re-interpretation for successful comprehension (Trueswell, Sekerina, Hill & Logrip, 1999), which may be facilitated by various factors. Many existing studies examining online processing in bilingual children have adopted violation paradigms (Chondrogianni & Marinis, 2012, 2016). Whilst such tasks provide insight into sensitivity to different morpho-syntactic violations, they do not tap into how a sentence is interpreted during incremental processing. To address this gap, we utilised the visual world eye-tracking paradigm, and temporary ambiguity in wh- questions, to examine incremental interpretation during comprehension. Furthermore, as it has been shown that morphosyntactic information can have a facilitatory effect during processing for adults and monolingual children by aiding disambiguation (e.g., Contemori, Carlson & Marinis, 2018; Schouwenaars, Hendriks & Ruigendijk, 2018), we also test whether this applies to bilingual children. In particular, we manipulated number agreement between the two NPs of a wh-question to assess if and how number agreement is utilised to facilitate processing of wh-questions in bilingual children.

There are several reasons why sentence processing in bilingual children may differ to that of their monolingual counterparts. Firstly, the input they receive in any language will inevitably be different to that of monolinguals in terms of quantity and quality of exposure, age of onset, among others. One possible consequence of this might be that bilingual children process language more slowly during comprehension than monolingual children. Indeed, the available evidence from ungrammaticality detection indicates that bilingual children are slower than monolinguals although they face a similar slowdown to monolinguals with phrase-level ungrammaticality (Chondrogianni & Marinis, 2016). Secondly, while slower processing does not necessarily implicate qualitatively different processing patterns between monolingual and bilingual children, it is a possibility. For example, bilingual children might not compute a syntactic representation of what they hear as quickly as monolinguals, or they may be unable to integrate the available linguistic/non-linguistic information quickly enough to guide processing while computing this representation. Slower processing in this sense might have knock-on effects that lead to different processing profiles between monolingual and bilingual children. The available evidence from off-line comprehension studies in how bilingual children process which-questions and utilise linguistic information to facilitate processing has indicated qualitative differences between bilingual and monolingual children at end-stage comprehension, at least for certain linguistic features such as case (Roesch & Chondrogianni, 2016). However, real time processing of which-questions in bilingual children and its timecourse remains unexplored.

Sentence processing and the use of morphosyntactic information in bilingual children

Research in morphosyntactic processing in bilingual children is scarce. A number of studies have examined grammatical violations in bilingual children using self-paced listening and word-monitoring tasks, and have investigated the processing of tense, definite and indefinite articles, clitic pronouns, and gender agreement (Chondrogianni

https://doi.org/10.1017/S0305000922000253 Published online by Cambridge University Press

& Marinis, 2012, 2016; Chondrogianni et al., 2015; Vasić et al., 2012). Most studies indicate nativelike sensitivity to grammatical violations as evidenced by a slowdown at the ungrammatical segments in a sentence for both bilingual and monolingual children. One exception to these findings comes from Vasić et al. (2012) who report a lack of sensitivity to grammatical gender violations in Dutch. This is interpreted as indicating difficulties with lexical knowledge (grammatical gender) rather than difficulties in processing grammatical cues.

One study to test comprehension of German wh-questions as well as the use of morphosyntactic information as a facilitatory cue is Roesch and Chondrogianni (2016), who used a picture selection task with simultaneous and early sequential bilinguals. This study examined the use of case as a disambiguating cue where the position thereof was manipulated (initial, final and both). The impact of case differed across groups; monolinguals made consistent use of the cue and accuracy was higher when case disambiguated whether an NP was a subject or object; the simultaneous bilinguals did so only in the initial position while the sequential bilingual did not do so at all. This points to a reduced use of at least certain disambiguating cues, which may be further contingent on early exposure in bilingual children.

The few studies currently available which use eye-tracking to investigate real-time use of morphosyntax in bilinguals have examined the use of grammatical gender marking on determiners to predict the upcoming noun. Lew-Williams (2017) found that school aged English–Spanish bilingual children were able to utilise number but not grammatical gender to predict upcoming nouns. This study, however, tested children in an immersion context where the L2 was not the majority language. Prediction based on grammatical gender was also investigated by Lemmerth and Hopp (2019) in simultaneous and sequential German–Russian bilinguals of the same age. While simultaneous bilinguals showed consistent effects of gender, as did monolingual controls, sequential bilinguals only showed similar effects for nouns with the same gender in Russian and German.

These results suggest reduced use of some types of morphosyntactic information to guide ambiguity resolution and prediction in some groups of bilingual children. It should be noted however that case and gender are fundamentally different morphosyntactic cues to number, which is examined in this study. Gender, for example, is idiosyncratic to particular lexical items and requires acquisition on a largely item-by-item basis, while case functions as a signal to word order and thematic roles. Number on the other hand is not lexically idiosyncratic, and is instead tightly related to conceptual information. As such, comparisons between our study on number and existing research on case and gender are only indirect. While bilingual's children use of case and gender have been examined in offline and online tasks, to date we are unaware of any existing study that has examined bilingual children's use of number agreement during real-time sentence processing.

In sum, research in real time morphosyntactic processing in bilingual children has focused on morphology and has not so far expanded to filler-gap dependences which has been limited to off-line comprehension. Thus, cognitive mechanisms such as incrementality and revision/re-interpretation in bilingual children during sentence processing remain essentially unexplored. Available evidence indicates similar but slower patterns of processing, but it is unclear whether bilingual children can utilise morphosyntactic information in real time in the same way as monolingual children to facilitate comprehension.

https://doi.org/10.1017/S0305000922000253 Published online by Cambridge University Press

The (psycho)linguistics of wh-questions

Wh-questions and, more generally, filler-gap dependences have been widely investigated in first language acquisition and first language processing in adults (e.g., Deevy & Leonard, 2004; Frazier, 1987; Gibson, 1998; Goodluck, 2005; Grodner & Gibson, 2005; Rizzi, 2004). In such dependencies, a link needs to be made between dislocated overt and null elements (fillers and gaps respectively in generativist terminology), which can stretch over a number of words and syntactic constituents. Subject wh-questions are constructions where the dislocated element is the subject of the verb whereas object wh-questions are those where the object of the verb is dislocated, as in (1a) and (2a) respectively.

(1a) Which donkey is carrying the zebra? (subject which-question)
(1b) Which donkey is the zebra carrying? (object which-question)

For English and many other languages, object-extracted wh-questions exhibit greater difficulty than subject-extracted questions, manifested in lower comprehension accuracy and/or slower response times; this has been termed as the "subject-object" asymmetry (Frazier, 1987; Stowe, 1989). One reason that object wh-questions cause difficulty, that is most pertinent to this study, is that the initial noun phrase ('which donkey') is initially preferentially assumed to be a subject. Due to the absence of overt case morphology, the examples (1a) and (1b) are locally ambiguous until after the auxiliary verb. It is only after the first word after the auxiliary – in which-questions, the second NP – that the ambiguity is resolved. However, the parser does not wait until the second NP and instead will begin to construct a syntactic representation and an interpretation of the sentence immediately and incrementally. Initially, the parser prefers interpreting ambiguous sentences as in (1a) and (1b) as a subject-question. The increased difficulty for sentences such as (1b) arises from the mismatch between the preferred syntactic structure and interpretation and the ultimately correct syntactic representation. For object-questions, this interpretation will need to be revised after the second NP.

This discrepancy can be explained under numerous theoretical accounts, the testing of which is not the primary aim of this paper. For example, the active filler hypothesis (Clifton & Frazier, 1989) predicts that gaps are filled at the first available possibility, in an effort to keep the syntactic structure as simple as possible. Thus, both (1a) and (1b) are predicted to initially be interpreted as subject wh- questions, as the subject gap position becomes available first during incremental processing. Reanalysis will be required in (1b) when the sentence is disambiguated to the correct object gap interpretation. Under alternative probabilistic or experience-based accounts (Hale, 2001; Levy, 2008; Roland, Dick & Elman., 2007), speakers of English are predicted to adopt the heuristic of initially interpreting the first NP as the subject of the wh-question due to their exposure. Such accounts assume that (1a) and (1b) are initially interpreted as subject questions because these are more likely to be encountered. The difficulty hence arises from the mismatch between this expectation and ultimately the correct syntactic representation. Assuming monolingual and bilingual speakers are able to compute the appropriate structure, the active filler hypothesis would predict an initial subject-bias for both monolingual and bilingual children. Under probabilistic/experience-based accounts, a subject first preference might also emerge in both monolingual and bilingual children as subject/agent first structures are more frequent in English than object first ones.

The tendency towards a subject-bias is well attested for filler-gap dependences in English even in typical adult L1 speakers using a variety of methodologies. This can be

https://doi.org/10.1017/S0305000922000253 Published online by Cambridge University Press

evidenced from both behavioural measures, such as slower reaction times for object questions (Stowe, 1989; see also, Grodner & Gibson, 2005 for relative clauses), but also electrophysiological measures (Phillips, Kazanina & Abada, 2005) as well as eye-tracking measures (Staub, 2010). Similar results have been obtained for other languages, such as German (Schlesewsky, Fanselow, Kliegl & Krems, 2000) and Dutch, (Frazier & Flores d'Arcais, 1989).

Difficulty in processing object extracted wh-questions could also be attributed to locality effects predicted under a specific theoretical account within the generativist framework – namely, Relativized Minimality (Rizzi, 1990, 2004; for empirical evidence in children, see Friedmann et al., 2009). 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, thus, yielding competition effects. This entails a significant prediction: that the intervening element will have this effect in sentences where the subject and object share the same number, as in (2a) and (2b), but not where they are different, as in (2c) and (2d), where the mismatch in number will function as a cue and aid comprehension.

(2a) Which donkey is the zebra carrying?
(2b) Which donkeys are the zebras carrying?
(2c) Which donkey are the zebras carrying?
(2d) Which donkeys is the zebra carrying?

A secondary aim of this study was to test this prediction in conjunction with the question as to whether bilingual children can utilise cues (in this case, morphosyntactic in nature) to facilitate processing.

Wh-questions in children

There is ample research on the acquisition of wh-questions in children, but the majority has focused on production or offline end stage comprehension tasks. There is a substantially smaller body of evidence for online language processing. Wh-questions have been shown to be a challenging linguistic structure to be acquired in children. 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 the comprehension of object which-questions have been shown to persist until early school years (Deevy & Leonard, 2004; Goodluck, 2005). 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), and Hebrew (Friedmann, Belletti & Rizzi, 2009).

Omaki, Davidson White, Goro, Lidz and Phillips (2014) examined the comprehension of embedded wh-questions in L1 English and L1 Japanese 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 (3).

https://doi.org/10.1017/S0305000922000253 Published online by Cambridge University Press

(3a) "Where did Lizzie say that she was gonna catch butterflies?"
(3b) "Where did Lizzie tell someone that she was gonna catch butterflies?".
(3c) "Where did Lizzie say to someone that she was gonna catch butterflies?".

The main clause interpretation for (3) – where the wh-phrase was attached to the verb "say/tell" rather than "catch" – was more frequent in the responses of both adults and children for English, suggesting a similar bias in adults and children to placing the filler at the earliest possible gap. For Japanese there was a preference for an embedded clause interpretation of sentences such as (4).

(4) 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. 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 more often a main clause interpretation whereas the children did not. This suggests that the children had difficulty re-analysing the sentence. Omaki et al. shows incremental processing and difficulties recovering from garden-path effects, but these conclusions are extrapolated on the basis of off-line findings[1].

In the first relevant eye-tracking study, Atkinson, Wagers, Lidz, Phillips and Omaki (2018) investigated locally ambiguous filler-gap dependences such as "*Can you tell me what Emily was eating the cake with?*" with children and found a bias towards filling the gap as early as possible for adults and 6-year-old children but not for 5-year-olds. The results are consistent with the existing literature which suggests that children process language incrementally.

To our knowledge, the first study to investigate processing of which-questions in children and how they utilise cues to aid processing using online measures is Contemori et al. (2018). The study used the visual world eye-tracking paradigm with L1 English children aged 5-7 years. Participants heard subject and object wh-questions like (1a) and (1b) while looking at a picture and needed to answer a comprehension question by clicking on the picture corresponding to answer. Analysis of accuracy and gaze data showed a persistent disadvantage for object questions for the children. The gaze data showed similar processing mechanisms for children as with adults. For object which-questions looks to the picture corresponding to the sentence heard initially decreased below chance and then increased. This change indicates that object-questions were initially interpreted as subject questions and only after the disambiguating second NP did the hearer reanalyse the questions and build a different syntactic representation.

[1]For similar findings from a study using eye-tracking but for a different type of locally ambiguous sentence, see Trueswell et al., (1999).

https://doi.org/10.1017/S0305000922000253 Published online by Cambridge University Press

Accuracy for object questions remained below that of subject questions indicating persistent difficulty with reanalysis consistent with Omaki et al. (2014).

Contemori et al. also tested the use of morphosyntactic cues to facilitate processing by manipulating the number of the two noun phrases so that it would either be the same (match) as in (5a) or different (mismatch) as in (5b).

(5a) Object-match: Which cow is the goat pushing?
(5b) Object-mismatch: Which cow are the goats pushing?

It was expected that (5b) would be easier to process than (5a), as the number mismatch between the auxiliary and the first NP in (5b) provides an early cue for disambiguation before the second NP[2]. This is also predicted under Relativized Minimality as the feature mismatch means that the second NP/intervener cannot function as a potential filler to the gap. Contemori et al. found that the mismatch in number resulted in higher comprehension accuracy for object questions. In terms of gaze data, a faster increase in looks to the picture corresponding to the sentence when the two NPs had a different number than when they were the same was taken to reflect the fact that number mismatch functioned as a facilitatory cue during real time processing.

Finally, Schouwenaars et al. (2018) investigated the role of case and number agreement cues in subject and object which-questions in monolingual German children using the visual world paradigm. They found a similar pattern of looks as in Contemori et al. with object-questions initially misinterpreted and morphosyntactic cues aiding reanalysis. However, the children were slower to revise an initial misinterpretation for object questions where disambiguation was aided by number agreement only (the two NPs differed in number) relative to when there was disambiguation from both number and case.

The current study

The current study builds on the visual-world eye-racking study from Contemori et al. (2018). Adopting the same research paradigm, we examine the processing of wh-questions in bilingual and monolingual children. To our knowledge, this is the first eye-tracking study on wh-questions in bilingual children. The aims were to examine which-question processing in bilingual children relative to monolingual children and potential differences between subject- and object-questions. We further explore the timecourse of processing to investigate whether there is incrementality in syntactic processing in bilingual children as has been established for monolinguals. Moreover, we examine 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.

An additional contribution of this paper is both methodological and theoretical. A limitation of Contemori et al. was the use of an incomplete paradigm as only which-questions where the number of the first noun phrase was exclusively singular were used (i.e., SG-SG for the match condition, SG-PL for the mismatch condition). However,

[2]The experimental design also included subject questions, but the number mismatch should only be relevant to facilitating processing of object questions where the mismatch between first NP and auxiliary can act as a disambiguating cue.

https://doi.org/10.1017/S0305000922000253 Published online by Cambridge University Press

wh-questions with a plural first noun phrase were not included (i.e., PL-SG and PL-PL for match and mismatch respectively). The plural number has been described as the marked number option relative to the singular in linguistics. Marked features have been associated with additional complexity in linguistics or difficulty in acquisition (Harley & Ritter, 2002; Haspelmath, 2006). They have also been associated with increased difficulty in processing as attested by an increased occurrence of attraction errors – albeit in a different type of syntactic dependency, subject-verb agreement (Wagers, Lau & Phillips, 2009). Therefore, it is unclear whether the effect of number mismatch for questions with a plural first NP will be the same as for questions with a singular first NP. To address this, this study also manipulated the number of the first noun phrase to be either singular or plural (henceforth "First NP") across all previous conditions. We tested older children (8-11 years) relative to Contemori et al. in order to expand the results to older monolingual children and to ensure that the experimental paradigm was useable with bilingual older children.

Our research questions are:

1. Do bilingual children differ to monolingual children in their ultimate interpretation of which-questions i.e., is there evidence that both groups misinterpret object wh- questions.
2. How does number (mis)match influence offline comprehension? Is the effect of number (mis)match modulated by the number of the first NP?
3. Do bilingual children initially misinterpret object wh-questions as subject questions and does the timecourse of recovery differ between monolingual and bilingual children?
4. How does number (mis)match influence real time comprehension of wh-questions in bilingual children? Is the effect of number mismatch modulated by the number of the first NP?

Research questions 1 and 2 can be addressed based on the end-result accuracy data and reaction times. Research questions 3 and 4 are examined based on the gaze data.

Given previous research suggests bilingual children process sentences more slowly than their monolingual peers (Chondrogianni & Marinis, 2012, 2016; Chondrogianni et al., 2015; Vasić et al., 2012), it is plausible that there will be differences between the two groups in this study in terms of processing. These may emerge for the reaction times and/or the gaze data. Previous research has, however, relied on reaction time data from self-paced listening studies. In this respect the timecourse of processing and bilinguals' real time interpretation of sentences remains largely unknown. The evidence for slower processing found in previous studies may reflect either an overall slower but qualitatively similar processing mechanism or a slower processing mechanism alongside qualitative differences. Slower but qualitatively similar processing will be evidenced by a similar but delayed trajectory of looks towards the target image for the bilingual children relative to the monolinguals (i.e., same curve shape with a time-delayed overlay). If slower processing results in qualitative differences, bilingual children may not be able to compute a syntactic representation quickly enough in real time and thus not misinterpret it. If bilinguals initially misinterpret object questions as subject questions, looks to target for object questions will drop below chance and increase thereafter as in Contemori et al. This is not expected to be the case for subject questions where looks will increase from the beginning and will plateau earlier.

https://doi.org/10.1017/S0305000922000253 Published online by Cambridge University Press

Previous work on bilinguals' use of morphological cues has indicated a more nuanced and potentially reduced use of at least some morphosyntactic cues (case and gender), in both offline comprehension (Roesch & Chondrogianni, 2016) but also real-time processing (Lemmerth & Hopp, 2019; Lew-Williams, 2017). However, it is unclear whether bilingual children will be able to utilise number mismatch in the same way as monolingual children. Contemori et al. (2018) showed an effect of number mismatch for disambiguating object questions for both comprehension accuracy and real time processing where number mismatch resulted in higher accuracy and a faster increase in looks to target in comparison to object questions where the number matched. If bilingual children have difficulty integrating morphosyntactic information quickly enough during processing, they may be insensitive to number mismatch unlike monolingual children. However, if bilingual children 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. The aforementioned models of language processing do not make explicit predictions about the effect of number, but it is expected that the unmarked forms will be easier to process for both groups.

Method

Participants

A total of 68 children from Grades 3-6 participated in this study: 37 monolingual children aged 7;10-11;6 (M=9;7, SD=1;1, 16 girls and 21 boys) and 31 bilingual/multilingual children 7;4-11;5 (M= 9;6, SD=1;2, 17 girls and 14 boys) who were recruited from the same schools in the UK. None of the children had a history of language impairment or learning difficulty. All bilingual children had a minimum exposure to English of two years. All children undertook a series of baseline assessments including CELF-4 (Concepts & Following Directions, Word Classes, Formulated Sentences, Recalling Sentences), TROG-2, Renfrew Test of Word Finding, CNRep, Raven's Coloured Progressive Matrices. All children scored within age-appropriate norms. As a group, the bilingual children underperformed the monolingual children on several measures of language but not on others. The results from the between group comparisons CELF-4 composite scores are summarised in Table 1 (for an overview of results from baseline

Table 1. Group comparisons for baseline language measures administered to children

Task	Mean score for monolingual children (SD)	Mean score for bilingual children (SD)	Test statistic	Effect size	Significance
		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	η^2p = .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

CLS: Core Language Score; RLI: Receptive Language Score; ELI: Expressive Language Score; LMI: Language Memory Index (composite scores from CELF-4); ANOVAs were used for normally distributed data – Mann Whitney tests when they were not.

https://doi.org/10.1017/S0305000922000253 Published online by Cambridge University Press

Table 2. Demographics specific to bilingual children

	N
Language spoken at home:	
French	6
Greek	3
Chinese or Serbian or Spanish	2 each
Polish/Serbian or Russian/Georgian or Turkish/Urdu	1 each
Arabic/Flemish/Italian/Latvian/Lithuanian/ Macedonian/Slovak/Surashtra/Tibetan/Urdu	1 each
Country of birth:	
UK	18
France	4
Canada/Georgia/Italy/Libya/Lithuania/Spain/Russia/ Autonomous Region of Tibet (China)	1 each
Not provided	1

measures, see Appendix B). For comparisons, we used both raw and standard scores where available.

The children's language history was carefully documented through the use of the PABIQ questionnaire and brief semi-structured interviews. Background information about language development and use as well as parental education was collected. In terms of their linguistic background, the multilingual children came from a variety of backgrounds; these are summarised in Tables 2 and 3. The majority of bilingual children were classed as English dominant based on the PABIQ questionnaire and one third was rated as balanced in terms of language proficiency. Almost all children used English more often in community and educational settings but language use at home was evenly divided between English dominant, L1 dominant and balanced. Sample size did not permit splitting the bilingual children into subgroups. However, measures of exposure to English, language proficiency and bilingual dominance were used individually as covariates in separate models to the bilingual children's data to control for individual variation in performance. These measures were not significant and did not improve model fit[3]. All monolingual children were born in the UK except two who were born in Australia and grew up with only English spoken in the home and in their environment. All but 2 bilingual children spoke L1s which overtly marked plurality in nouns based on the World Atlas of Language Structures (Dryer, 2013).

Ethical approval 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 email or word of mouth. Separate information sheets and consent forms were completed by children and parents.

[3]Interactions between fixed effects and covariates were not included as the more complex model failed to converge.

https://doi.org/10.1017/S0305000922000253 Published online by Cambridge University Press

Table 3. Parental report – Bilingual language profile: Age of Onset (AoO), Length of Exposure (LoE) to English for bilingual children and dominance (mean and range in years; SD in months)

Reported AoO and LoE		Around birth [N = 20]	Early (<4 years) [N = 4]	Late (>5 years) [N = 5]
AoO	Mean	0	2;6	6;8
	SD	0	0.48	11.93
	Range	0	2;0-3;0	5;6-8;0
LoE	Mean	9;3	6;8	3;4
	SD	13.85	5.74	17.42
	Range	7;4-10;3	6;2-7;4	2;2-5;6

Dominance profile based on parental ratings	English dominant (N)	Balanced (N)	L1 dominant (N)
Language proficiency	19	10	1
Use at home	10	8	12
Use outside home	26	1	3
Total language exposure	22	1	7

Relevant information in the parental questionnaire was not provided by the parents of two participants with regards to AoO and LoE, hence the discrepancy in the numbers.
For the calculation of LoE, chronological age was used for children who were exposed to English at birth. For children who were exposed to English after birth Age of Onset was subtracted from chronological age.
For the dominance scores, mean and SD are not calculated as the measure is ordinal. Instead counts are reported. Children with scores <–2 are considered English dominant; score >2 as L1 dominant and scores between –2 and +2 are considered balanced for the particular measure. These are calculated based on parental ratings.

Design

The study used a visual world eye-tracking task. Participants heard a which-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 patient in the other and vice versa. In each trial, one picture 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.

The first within-subjects variable was the type of the which-question (subject vs. object). The second within-subjects variable was the number of the two noun phrases so that the two could be either the same or different (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, as exemplified in Table 4. The between-subjects variable was language group (monolinguals vs. bilinguals).

Materials

For each trial, one which-question and two pictures were used. 80 which-questions were created by forming which-questions with ten lexical sets across all conditions. Each set of

https://doi.org/10.1017/S0305000922000253 Published online by Cambridge University Press

Table 4. Sample experimental stimuli by condition

Conditions	Subject Questions	Object Questions
Number Match – First NP Singular (SG-SG)	Which bear is chasing the camel?	Which bear is the camel chasing?
Number Match – First NP Plural (PL-PL)	Which bears are chasing the camels?	Which bears are the camels chasing?
Number Mismatch – First NP Singular (SG-PL)	Which bear is chasing the camels?	Which bears is the camel chasing?
Number Mismatch – First NP Plural (PL-SG)	Which bears are chasing the camel?	Which bear are the camels chasing?

lexical items involved a transitive verb and two animals. The transitive verbs were action verbs in the active and were semantically reversible. This way all sentences in the experimental trials were semantically reversible. The verbs used were the same as in Contemori et al. (2018); they were high frequency verbs with an age of acquisition of five years and under according to the MRC Psycholinguistic Database. Contemori et al. compared the frequencies of the nouns used and found no differences, see Appendix A for a full list of sentences. Stimuli were digitally recorded by a male L1 speaker of British English. The sentences were recorded as a single sentence rather than cross-spliced to preserve natural intonation. Trial sentences were reviewed and those with poor audio quality, clicks and abrupt changes to rhythm and intonation were re-recorded before the task was administered to participants. There were no fillers because the inclusion of fillers would have increased the length of the experiment and would have risked loss of attention.

The visual stimuli were derived from Contemori et al. (2018), but additional pictures for the novel conditions in this study were created by copying, to ensure maximal visual similarity. The size and visual features of target and competitor were similar to the greatest degree possible. An example of the visual stimuli is shown in Figure 1, where 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 left. For the trials with the mismatch condition, the singular and plural entity was the same in both pictures. 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.

Trials were pseudo-randomised so that each set of nouns occurred at varying intervals from 2 to 19 intervening trials (mean = 9.18, SD = 3.78). Furthermore, no trials were permitted to follow trials of the same condition although adjacent trials with the same level of a single variable (e.g., a subject question followed by a subject question) were permitted and occurred in around half the trials. A single list was used for this study;

Figure 1. Sample visual stimuli for Subject & Object questions with SG-SG NP pairing.

https://doi.org/10.1017/S0305000922000253 Published online by Cambridge University Press

therefore, a random intercept of trials was initially allowed to control for effects of a single trial occurring in a fixed place. This however was removed as it contributed little to the variance and did not improve the model fit.

Procedure

The experiment took place in a quiet room with the participants wearing headphones. A Tobii X120 (Tobii Technology AB, Sweden) eye-tracker measured the participants' eye-gaze, tracking eye position with a resolution of 120Hz. 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). Testing started with a 5-point calibration procedure. The experimenter (first author) judged the quality of the calibration by examining the calibration plot for the five points. Quality of calibration was judged as adequate when the eye-tracker captured the participant's looks at all 5 points and there was limited drag in line with the guidance provided in the Tobii X120 manual. 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. This also functioned as a calibration check, as the fixation would only register if adequately calibrated. Participants heard a question over a set of headphones and saw two pictures on each side of the screen. Following the question, a cursor appeared on the 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, but the mouse click was not allowed until after the audio file had finished. 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.

Analyses

Accuracy, reaction time, and gaze data were collected and analysed. As the duration of the trials with subject questions was longer than that of ones with object questions (mean = 2,727ms, SD = 126 vs. mean = 2,533ms, SD = 122), the participants' reaction times were defined as the difference between the time the participant needed to click on the selected picture and the duration of the trial. There were no negative times.

For the gaze data, 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 as in Contemori et al. This timepoint allows one to capture effects of misinterpretation and is the earliest point at which number mismatch can disambiguate. It is expected that looks will initially be approximately equal for both pictures as the participants explore the visual stimuli and will subsequently increase for the picture consistent with a subject-biased interpretation. We did not time lock the second NP as this would miss any effect of misinterpretation occurring at the point of the structural ambiguity and would be less likely to reflect incremental and subconscious processing in the latter time bins.

A window of 200ms was allowed for the time it takes to program a saccadic eye-movement (Matin, Shao & Boff, 1993), such that eye-movements were analysed for a

https://doi.org/10.1017/S0305000922000253 Published online by Cambridge University Press

period of 2 seconds (200-2200ms post auxiliary). Incorrect trials were removed from the analyses consistent with Contemori et al. and standard practice with this type of data[4]. This resulted in the loss of around 5% of trials with subject questions and 15% of trials containing object questions due to lower comprehension accuracy in the latter. The time period examined was divided into ten equal bins of 200ms. For each bin, the proportion of looks to target relative to competitor was calculated. These proportions were quasi-logit transformed to compute the empirical logit which better handles cases where the probability is high of low (Barr, 2008).

The analysis was conducted using logistic mixed effects models for accuracy, linear mixed effects models for reaction time and a growth curve spline function for the gaze data – in line with Contemori et al. – 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). To control for the age range and the variability in the children's language skills, age and the Core Language Score (CLS) from the CELF-4 were entered as a continuous variable into the model. Interactions with other variables were not included for purposes of model convergence. The reasoning for selecting the CLS is two-fold: firstly, the CLS is a composite score which best reflects a child's linguistic competence as it is comprised from the scores of several diverse tasks. Secondly, adding scores from numerous baseline assessments as predictors or covariates requires larger datasets for a model to converge and may not be meaningful, as the scores on individual tasks may be correlated as they reflect the same aspect of a child's linguistic competence. As bilingual proficiency may be influenced by length of exposure (henceforth LoE) and language dominance, we fitted a second model to the data for bilingual children only using length of exposure and two dominance scores calculated based on responses in the parental questionnaire alongside the fixed effects. Dominance was defined as the difference in the proficiency scores in the two languages and exposure was defined as the difference in composite scores for exposure to each of the two languages across numerous settings (e.g., school, home, friends). For these two measures, a negative score indicated dominance in English and greater exposure to English, respectively. Age and language proficiency or dominance scores were entered only as main effects into the models, as including interaction terms of background measures with the independent variables generally deteriorated the model fit.

For the gaze data, a growth curve model was fitted to looks to the correct picture to capture change as a non-linear function of time (Mirman, Dixon & Magnuson, 2008; Mirman, 2014). Time was coded as a restricted – or natural – cubic spline with 4 equidistant knots creating three different components[5] (Harrell, 2001). This type of transformation captures the non-linear change in time as the independent variable is transformed to include a linear, a quadratic and a cubic component. The use of a spline function adds further flexibility to the non-linear modelling of change over time by allowing the function and its parameters to differ across the components. In this type of modelling, significant main effects and interactions on the intercept term signify overall differences irrespective of time as in more conventional models. Significant main effects and interaction on the spline's components, i.e., those which involve time, signify that the

[4]Trials were originally analysed as a whole with weaker effects and then separately for correct and incorrect trials; as the patterns observed in the incorrect trials were not robust, these were assumed to be noise and as such excluded from further analyses.

[5]Component 1: Time = 0 to x1, Component 2: x2 – x1, Component 3: End of data analysed – x2.

https://doi.org/10.1017/S0305000922000253 Published online by Cambridge University Press

shape of the growth curve varies between the different levels of the independent variable, e.g., faster or slower growth rate. We conservatively focus on those effects that were significant on a minimum of two of the three components of the spline, as these would reflect the most consistent patterns in the participants' behaviour.

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, as defined as 200ms bin number, was scaled in order to conduct the growth curve analysis. 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 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 for each time bin, 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. The raw data and code for each analysis can be found at https://osf.io/4w693/.

Results

To examine RQ1 and RQ2, we analysed the comprehension accuracy and reaction time data. An overview of the results for the accuracy data and the reaction times can be found in Table 5, followed by the results for the models in Table 6.

Accuracy and reaction times

Table 5. Accuracy as a percentage and reaction times by condition for each group (95% bootstrapped CIs in square brackets)

	Accuracy (%)		Reaction times (ms.)	
Condition	Monolinguals	Bilinguals	Monolinguals	Bilinguals
Subject questions – Number match – First NP Singular	98.5%	98.6%	2232	2433
	[97.0 – 99.7]	[97.1 – 99.6]	[2063 – 2414]	[2188 – 2698]
Subject questions – Number match – First NP Plural	96.6%	97.8%	2089	2612
	[94.5 – 98.5]	[96.0 – 99.3]	[1923 – 2266]	[2145 – 3208]
Subject questions – Number mismatch – First NP Singular	98.5%	97.8%	2030	2212
	[96.9 – 99.7]	[96.0 – 99.3]	[1860 – 2216]	[1995 – 2442]

https://doi.org/10.1017/S0305000922000253 Published online by Cambridge University Press

Table 5. *(Continued)*

Condition	Accuracy (%)		Reaction times (ms.)	
	Monolinguals	Bilinguals	Monolinguals	Bilinguals
Subject questions – Number mismatch – First NP Plural	94.2%	97.3%	1993	2016
	[91.2 – 96.9]	[95.0 – 99.0]	[1855 – 2139]	[1794 -2272]
Object questions – Number match – First NP Singular	84.5%	83.5%	2673	2756
	[80.5 – 88.2]	[79.1 – 87.9]	[2455 – 2908]	[2416 – 3153]
Object questions – Number match – First NP Plural	83.4%	83.3%	2696	3313
	[79.4 – 87.4]	[78.9 – 87.6]	[2471 – 2936]	[2913 – 3755]
Object questions – Number mismatch – First NP Singular	89.2%	87.3%	2673	2995
	[85.4 – 92.7]	[82.7 – 91.4]	[2439 – 2924]	[2575 -3462]
Object questions – Number mismatch – First NP Plural	87.9%	87.4%^	2596	2883
	[84.2 – 91.2]	[83.4 – 91.5]	[2355 – 2880]	[2581 – 3212]

Table 6. Fixed effects for the accuracy and reaction time data

Variable	Accuracy				Reaction Times			
	β	SE	z value	p	β	SE	t value	p
	3.26	0.16	19.80	<0.0001	2565.06	140.70	18.23	<0.0001
Group	0.11	0.14	0.77	0.44	209.94	137.72	1.52	0.133
Structure	−1.06	0.08	−13.70	<0.0001	333.10	32.45	10.26	<0.0001
Number	0.03	0.08	0.33	0.74	−98.96	32.46	−3.05	0.002
First NP	−0.19	0.07	−2.50	0.01	8.60	32.30	0.27	0.79
Age	0.49	0.13	3.84	0.0001	−565.60	132.44	−4.27	<0.0001
CLS	0.44	0.12	3.57	0.0004	9.57	137.39	0.07	0.945
Group x Structure	−0.09	0.08	−1.11	0.269	48.67	32.37	1.5	0.132
Group x Number	−0.01	0.08	−0.18	0.860	−35.59	32.37	−1.10	0.270
Group x First NP	0.12	0.08	1.46	0.145	29.06	32.24	0.90	0.367
Structure x Number	0.18	0.08	2.29	0.022	37.33	32.29	1.16	0.248
Structure x First NP	0.17	0.08	2.11	0.035	25.33	32.29	0.79	0.433
Number x First NP	−0.03	0.08	−0.33	0.743	−51.88	32.31	−1.61	0.108
Group x Structure x Number	−0.01	0.08	−0.07	0.941	26.05	32.33	0.81	0.419
Group x Structure x First NP	−0.09	0.08	−1.07	0.280	17.69	32.24	0.55	0.583

https://doi.org/10.1017/S0305000922000253 Published online by Cambridge University Press

Table 6. *(Continued)*

	Accuracy				Reaction Times			
Variable	β	SE	z value	p	β	SE	t value	p
Group x Number x First NP	0.05	0.08	0.68	0.495	−66.02	32.23	−2.05	0.041
Structure x Number x First NP	0.02	0.08	0.29	0.772	−23.62	32.28	−0.732	0.465
Group x Structure x Number x First NP	−0.04	0.08	−0.56	0.574	7.875	32.23	0.24	0.807
For bilingual children only	β	SE	z value	p	β	SE	t value	p
LoE	−0.074	0.24	−0.30	0.761	308.11	332.03	0.93	0.363
Dominance	−0.239	0.28	−0.87	0.387	−5.46	395.14	−0.01	0.989
Exposure	0.198	0.29	0.68	0.495	36.03	406.92	0.09	0.930

Accuracy data

There was a significant main effect of syntactic structure, with lower accuracy for object than subject questions. Neither the main effect of group or number match, nor any interactions with group, were significant. There was a marginally significant main effect of the number of the first NP, with higher accuracy for questions with a singular first NP and a significant interaction of structure by number of first NP and an interaction of structure by number match. Overall accuracy also significantly improved with age and language proficiency. Given that subject and object questions have shown differential effects in previous studies (e.g., Contemori et al., 2018) and the significant structure by number (mis)match interaction separate analyses were carried out for subject and object questions. A main effect of number match was found for object questions where questions with a mismatch in number between the NPs resulted in higher accuracy than when there was a match. This was not found for the subject questions (Table 8 in the Appendix). For the bilingual children, accuracy did not improve with length of exposure, quantity of exposure or increased dominance in English and did not improve the model fit.

Reaction times

There was a significant effect of structure, with slower reaction times for object questions and an effect of number (mis)match with faster reaction times when there was a mismatch between the number of the two NPs. Neither the effect of group or number of first NP were significant. Reaction times became faster with age but there was no significant effect of language proficiency unlike with the response accuracy. The only significant interaction was the group by number by first NP number interaction. As separate models for monolingual and bilingual children yielded no further effects, this interaction is not discussed further (see Table 9 in Appendices for the output). Length of exposure, language

https://doi.org/10.1017/S0305000922000253 Published online by Cambridge University Press

https://doi.org/10.1017/S0305000922000253 Published online by Cambridge University Press

Table 7. Fixed effects for gaze data

Variable	β	SE	t	p	Variable	β	SE	t	p
Group	0.03	0.04	0.78	0.438	Time_1	0.30	0.02	14.14	<0.0001
Structure	−0.19	0.02	−9.25	<0.0001	Time_2	0.47	0.04	13.07	<0.0001
Number	0.02	0.02	1.20	0.231	Time_3	0.37	0.02	23.81	<0.0001
First NP	0.01	0.02	0.26	0.795	Group x Time_1	−0.09	0.02	−4.21	<0.0001
Age	0.05	0.03	1.45	0.153	Group x Time_2	−0.16	0.04	−4.49	<0.0001
CLS	0.01	0.03	0.32	0.753	Group x Time_3	−0.08	0.02	−5.28	<0.0001
Group x Structure	0.00	0.02	−0.05	0.962	Structure x Time_1	0.09	0.02	4.05	<0.0001
Group x Number	−0.01	0.02	−0.53	0.599	Structure x Time_2	0.04	0.04	1.21	0.225
Group x First NP	0.03	0.02	1.53	0.128	Structure x Time_3	0.11	0.02	6.93	<0.0001
Structure x Number	0.07	0.02	4.17	<0.0001	Number x Time_1	0.03	0.02	1.27	0.204
Structure x First NP	−0.01	0.02	−0.40	0.691	Number x Time_2	−0.05	0.04	−1.49	0.135
Number x First NP	0.02	0.02	1.28	0.200	Number x Time_3	−0.03	0.02	−1.81	0.070
Group x Structure x Number	−0.02	0.02	−1.37	0.170	First NP x Time_1	0.02	0.02	0.91	0.363
Group x Structure x First NP	−0.01	0.02	−0.80	0.426	First NP x Time_2	−0.02	0.04	−0.66	0.506
Group x Number x First NP	−0.01	0.02	−0.51	0.612	First NP x Time_3	0.02	0.02	1.46	0.143
Structure x Number x First NP	−0.07	0.02	−4.31	<0.0001					
Group x Structure x Number x First NP	−0.01	0.02	−0.65	0.515					

dominance and quantity of exposure to English were not significant predictors of reaction times in the bilingual children.

Gaze data

To examine RQ3 and RQ4, regarding the processing of which-questions in real time, we analysed the gaze data. We first present a visual overview of the data as well as an overview of the model output (Table 7, the full model can be found in Appendix C). Subsequently, we outline the significant effects found. We first report the significant main effects and interactions on the intercept, i.e., those which do not involve time. This reflects overall aggregate differences irrespective of time and do not speak to the trajectory of looks to target. We then report main effects and interactions on the spline. The latter show differences in the shape of the curve and are interpreted as differences in the shape of the curve, i.e., change over time. For effects on the splines, we report up to two-way interactions for reasons of conciseness and as these are the most readily interpretable, although the full list of fixed effects can be found in the appendices (Table 10).

Effects on intercept term

There was a significant effect of structure with fewer looks to the target picture for object questions than for subject questions overall. There was no significant difference in total looks towards the target between the bilingual and monolingual children, nor was there a significant effect of number (mis)match and/or first NP number. However, age and language proficiency were not significant predictors of the total amount of looks towards the target. For the bilingual children, length of exposure to English, quantity of English exposure and language dominance were not significant predictors of looks to the correct picture.

Interactions on intercept term

There were significant interactions between structure and number, and structure and number and First NP which did not interact with group. To further explore the significant interactions, models were fitted to subject and object questions separately (Table 11). Focusing on the effects of match, an effect of number match was found for object questions, but this did not reach significance for the subject questions. This suggests an effect of number mismatch in facilitating processing of object questions but not subject questions.

Effects on spline/growth curves

There was a significant effect of time on looks to the target picture on all three components of the spline suggesting that the amount of looks to the target changed continuously. There was a significant main effect of group on all three components of the spline. This suggests differences in trajectory of looks between monolingual and bilingual children with the bilingual children showing less pronounced increases in looks to target. There was also a significant effect of structure on the first and the third component of the spline. However, the impact of number mismatch was weak with only a trend for number

https://doi.org/10.1017/S0305000922000253 Published online by Cambridge University Press

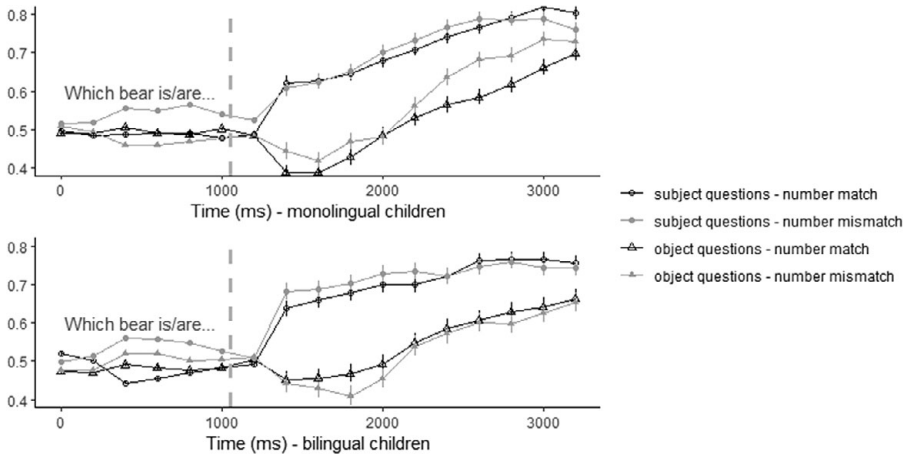

Figure 2. Looks to target as a proportion over time by group and structure (slashed vertical line indicates point of disambiguation).

mismatch in the third component of the spline. There was no effect of first NP number on any components.

Interactions on spline/growth curves

The only significant two-way interactions were an interaction of group by number (mis) match on the first component of the spline and an interaction of structure by number (mismatch) again on the first component of the spline. Visual inspection of the data (Figure 2) suggests an increase in looks to target for subjects following chance performance prior to the auxiliary verb and an initial decrease for objects followed by a slower increase. This reflects an initial misinterpretation of the latter as subject questions and a reanalysis of the structure upon disambiguation. In terms of between group differences in the trajectory of looks, visual inspection of the data suggests a generally slower increase in looks to target for the bilingual children, i.e., a less steep slope or flatter slope. This results in greater differences in looks between the two groups after about 2,000ms post onset of the which-question[6]. For the object questions in particular, looks in the bilingual children do not appear to drop as substantially below chance as they do for the monolinguals suggesting that the garden-path effects and the immediate re-interpretation of the ambiguous question may be taking place over a more protracted period.

Discussion

The present study is among the first to examine sentence processing in bilingual children using the visual world paradigm. Subject and object which-questions were utilised to examine incrementality and timecourse of processing alongside the utilisation of morpho

[6]As the average duration of subject and object questions is 2,500ms and 2,700ms respectively, this would be towards the end of the which-question (second NP or lexical VP) and shortly after its end.

https://doi.org/10.1017/S0305000922000253 Published online by Cambridge University Press

syntactic cues to aid disambiguation, in this case number mismatch between the two NPs. The results show that bilingual children did not significantly underperform the monolinguals in any of the metrics and that both groups showed increased difficulty with object questions. Unlike previous studies that relied on offline comprehension questions, our use of the visual-world paradigm allows us to claim that in both groups this was due to the initial misinterpretation of the ambiguous first NP as a subject NP, thus providing clear evidence of incremental processing in bilingual children. The bilingual children differed from the monolinguals in that they had a more gradual increase in looks to the target picture in the gaze data.

RQ1: Do bilingual children differ to monolingual children in their ultimate interpretation of which-questions?

In both groups, object which-questions were more difficult than subject which-questions. This is reflected in both lower comprehension accuracy and slower reaction times when the question was comprehended correctly. This is in line with a vast body of literature suggesting a subject-object asymmetry in filler-gap dependences (e.g., Contemori et al., 2018; Deevy & Leonard, 2004; Friedmann, Belletti & Rizzi, 2009; Goodluck, 2005; Grodner & Gibson, 2005; Stavrakaki, 2006; Stowe, 1989). As illustrated in the gaze data, this is due to an initial misinterpretation of object questions as subject questions and the subsequent need to reanalyse them after the first parse is untenable. However, the bilingual children were neither significantly slower nor significantly less accurate than their monolingual peers in answering the comprehension questions. This suggests that bilingual children show similar performance to monolingual children overall; both groups show a subject-bias with ambiguous NPs and both groups were usually able to recover from induced garden-path effects and accurately reanalyse the sentence upon disambiguation. The fact that comprehension accuracy, i.e., the response after hearing the complete question, is lower for object relative to subject questions suggests that reanalysis is challenging for the parser and may not always be successful. This is in line with previous studies with children (Trueswell et al., 1999) and has been attributed to lingering misinterpretations observed even in adults (Slattery et al., 2013). Bilingual children did not have significantly greater difficulty with object questions in comparison to the monolinguals as the interaction with group and structure was not significant for accuracy and reaction times. The above suggest that the processing mechanisms for sentences are similar in both bilingual and monolingual children.

The current study is in line with previous studies on sentence processing in bilingual children that have also shown no differences in comprehension accuracy (e.g., Chondrogianni & Marinis, 2012; Chondrogianni et al., 2015; Vasić et al., 2012). Contrary to these studies, we did not find slower reaction times for the bilinguals. However, as previous studies used self-paced listening, the reaction times obtained reflect segment-by-segment real-time processing of a sentence whereas in this study reaction times reflected the time-taken to respond to the question after participants had heard the entire sentence.

Accuracy is higher for the monolingual children in this study relative to the children in Contemori et al. (80-85% vs. 63%). This discrepancy could be due to the fact that the children in this study are older than in Contemori et al. Therefore, as the parser matures, the child's ability to successfully revise misinterpretations becomes more robust leading to higher comprehension accuracy once the participant has heard the entire which-question.

https://doi.org/10.1017/S0305000922000253 Published online by Cambridge University Press

RQ2: How does number (mis)match influence offline comprehension?

Number mismatch was found to have a facilitatory effect for object questions but not for subject questions. This was consistent for both bilingual and monolingual children as evidenced by the absence of group by number (mis)match interactions. These results are similar to those reported by Contemori et al. (2018). This is in line with Relativised Minimality. The reason for the benefit of number mismatch exclusively for object questions is that it is redundant for subject questions; it is only for object questions that the initial interpretation will be erroneous due to the subject bias. Moreover, as with Contemori et al., accuracy for subject questions in this study showed ceiling effects, thus making any additional benefit from morphosyntactic features redundant.

Roesch and Chondrogianni (2016) showed that bilingual children could utilise case to facilitate object which-question processing in German. However, they found that this was limited to simultaneous and early sequential bilinguals and not found in late sequential bilinguals. We did not analyse our data by type of exposure due to the small sample sizes that this would entail. However, we contend that the difference between our study and the study by Roesch and Chondrogianni (2016) may relate to the morphosyntactic features tested.

Reaction times showed an overall benefit of number mismatch irrespective of question type. This is unexpected and would not be predicted by Relativised Minimality, nor would it be related to a subject-biased initial interpretation of ambiguous NPs. It cannot be compared to the results in other studies (Contemori et al., 2018; Roesch & Chondrogianni, 2016) as these do not report results from reaction times.

RQ3: Do bilingual children initially misinterpret object wh-questions as subject questions and does the timecourse of recovery differ between monolingual and bilingual children?

The use of eye-tracking is the novel component of this line of research into sentence processing in bilingual children as it allows us to better understand the cognitive mechanisms involved in processing. The results from this study show that both bilingual and monolingual children initially misinterpret object questions as subject questions. For both groups, looks to the target are initially at around chance suggesting the parser has not yet committed to a specific interpretation. Shortly after the auxiliary verb, there is a decline in looks to the target for object questions, reflecting an increase in looks to the competitor which corresponds to a subject-reading of the ambiguous NP. This was also clearly found in Contemori et al. (2018) for younger children and adults and also with Schouwenaars et al. (2018) for L1 speakers of German in a comparable age range.

The findings are consistent with both structure-based accounts (e.g., Active Filler hypothesis, Frazier & Clifton, 1989) and probabilistic accounts (e.g., Levy, 2008) of the subject-object asymmetry found in filler-gap dependences. Although our design did not attempt to tease apart these different accounts, our results indicate that erroneous initial parses and garden-path effects in bilingual children are the result of their sentence processing being incremental similarly to monolingual children.

Where bilingual and monolingual children differ is in the timecourse of real-time processing of which-questions as evidenced by the effects of group on all components of the spline. For object questions, the looks to the target remain at a low point for longer than in monolingual children (400-600ms vs. 200-400ms, see Figure 2). As the reorientation of looks to the target is taken to signify a re-interpretation of the ambiguous

https://doi.org/10.1017/S0305000922000253 Published online by Cambridge University Press

sentence and a recovery, we interpret this discrepancy in the timings of the increase in looks as a form of slower processing. The second difference between the two groups is in the subsequent increase in looks thereafter. In the bilingual children, the increase in looks to the target is less steep than in the monolingual children after they begin to reorient their looks. Visual inspection of the mean proportion of looks and the standard error suggests this results in significantly fewer looks to target for the bilinguals relative to the monolinguals after 2,000ms after the onset of the question. Differences in steepness of the growth curves are taken to reflect differences in speed of processing consistent with previous work in growth curve analysis (Mirman 2014; Mirman et al., 2008). In this sense, the results from this study are conceptually similar with other studies using self-paced listening which show equal accuracy but slower speed of processing in bilingual children (e.g., Chondrogianni & Marinis, 2012; Chondrogianni et al., 2015; Vasić et al., 2012).

Rather than reflecting qualitatively different patterns of processing, a tentative explanation for this difference in speed is that it results from two linguistic systems remaining active during bilingual sentence comprehension. Slower processing has been shown for bilinguals for both lexical processing (e.g., Blumenfeld & Marian, 2007; de Bruin, Della Sala & Bak, 2016) but also for sentence processing during production (e.g., Bernolet, Hartsuiker & Pickering 2007; Desmet & Declercq, 2006; Loebell & Bock, 2003; Hartsuiker, Pickering & Veltkamp, 2004). Alternatively, one could attribute these differences in processing speed to differences in the input, due to the bilingual speakers presumably having less input in English than the monolingual children. However, note these differences were observed even after language proficiency was controlled for in the analyses. Moreover, measures related to the input bilinguals received in English (LoE and dominance in use/proficiency) did not consistently predict the bilinguals' performance. Further research is needed to tease apart these two potential accounts of the observed differences in processing speed. Note importantly however, that the absence of a group by structure interaction indicates no additional processing burden for object questions for the bilingual children relative to their monolingual peers.

RQ4: How does number (mis)match influence real time comprehension of wh-questions in bilingual children?

Evidence for the effect of number match facilitating processing in real time was moderate. There were overall more looks to the target when there was a mismatch in number between the two NPs in the question than when the number matched for the object questions but no such effect for the subject questions. This indicates that number mismatch had a faciliatory effect on the processing of which-questions as was also found for the off-line measures from this study and also in Contemori et al. (2018) and Roesch and Chondrogianni (2016). This finding is again in line with Relativised Minimality. While number mismatch appears to aid monolinguals more than bilinguals (Figure 2), there was no significant group by number match (by structure) interaction.

Contemori et al. found more looks to target for number match for objects but not subjects – similarly to the present study – and also, a faster increase in looks to target in the children for object questions with number mismatch than for those where the number matched. This was not found for subject questions, nor was it found for the adults. In this sense, the children from the present study, who are older, behaved similarly to the adult controls in the study by Contemori et al. This could be associated with developmental changes in the parser's capacity; the children in this current study – aged 8-11 years – had a more adult-like parser than those in Contemori et al, – aged 5-7 years. Visual inspection of the modelled data in

https://doi.org/10.1017/S0305000922000253 Published online by Cambridge University Press

Contemori et al. shows that the children's looks to the target have a noticeably less steep increase relative to both the adult control data and the gaze data from this study. In fact, for the number match condition looks to target do not rise significantly above chance at any point for the object questions. This is not the case for the adult data in Contemori et al. and also the participants in the current study, where looks to target increase beyond chance for the object questions and show a similar sine-like pattern of decrease and increase.

In German, Schouwenaars et al. (2018) found that children utilised case to disambiguate wh-questions in real-time and could also utilise number agreement in the presence of disambiguating case information. However, number agreement alone did not facilitate disambiguation (note this condition is the one most comparable to the number mismatch manipulation in this study). Schouwenaars et al. argue the advantage for case is due to the fact that it is marked on the first NP, thus acting as an early disambiguating cue (whereas for object questions, number agreement is marked on both the verb and the second NP). Roesch and Chondrogianni (2016) also find an advantage for early cues in disambiguation. Case is not overtly marked on nouns in English and therefore number agreement is the only disambiguating cue available. Therefore, number agreement may have greater functional value as a cue in sentence processing in English than for languages with more complex overt inflectional morphology. This could thereby explain the differences in findings between Schouwenaars et al. and the present study.

The findings from this study differ from other eye-tracking studies on use of morphosyntactic cues during sentence processing in bilingual children (Lemmerth & Hopp, 2019; Lew-Williams, 2017). These have given a more nuanced picture with cue utilisation being more contingent on the exposure to the second/additional language or the properties of the language per se. We believe there are two explanations for this. Firstly, the children in this study are mostly simultaneous or early sequential bilinguals who show more monolingual like patterns as is the case with the simultaneous bilinguals in Lemmerth & Hopp. Greater divergence from monolingual patterns is observed in the sequential bilinguals in Lemmerth & Hopp and in Lew-Williams where the children are L2 learners. The second explanation lies in the nature of what is tested. Lemmerth and Hopp as well as Lew-Williams tested predictive processing and how this can be facilitated through gender marking. It is the case both that the latter is highly lexical, and that its acquisition is therefore expected to need a large quantity of input. There is no reason to expect this is the case for the incrementality in sentence processing and the need for revision in the case of experiencing garden-path effects.

The effects of first NP number

One limitation in Contemori et al. is that all trials mismatch with number mismatch had a singular first NP and a plural second one. We manipulated the number of the first NP to be either singular or plural by extending the paradigm from Contemori et al. Our findings suggest that this limitation did not impact the findings from Contemori et al. in terms of the effect of number mismatch as a facilitatory cue during processing. We did observe an effect for off-line comprehension accuracy, where accuracy was lower for trials with a plural first NP. This may be related to markedness; the plural NP is the marked form and may thus be harder. Alternatively, this difficulty may be attributed to the fact that singular which NPs are simply more felicitous as they require the selection of a single entity from a choice of two rather than a set of two entities from a choice of two sets. We note that this effect is observed only in the case of the response accuracy, and not eye-movement data or

https://doi.org/10.1017/S0305000922000253 Published online by Cambridge University Press

reaction times. Although this may suggest that this effect occurred during the question answering phrase, rather than during online processing of the critical sentences, we are cautious in drawing strong conclusions here given the effect was observed in only one measure.

Heterogeneity in bilingualism

One potential limitation of the current study is the heterogeneity of the population in terms of age, proficiency in English and linguistic background. To address this variability, we fitted the models with age and language proficiency (Core Language Score from CELF-4) as covariates for both bilingual and monolingual children. Age and language proficiency were significant predictors of performance in the anticipated direction. However, the main effects observed were significant even after age and language proficiency were controlled for. To address variability specifically in the bilingual children, we fitted the models with the bilingual data with length of exposure, English language proficiency dominance and English language exposure as covariates. These were not found to be significant predictors of performance and again did not improve the model fit. This is in contrast to previous studies which have shown that a younger age of onset leads to more nativelike acquisition (e.g., Roesch & Chondrogianni, 2016 for wh-questions; Lemmerth & Hopp, 2019 for predictive processing; Chondrogianni & Marinis, 2011 for vocabulary) as does greater exposure to a language (e.g., Peña, Bedore, Shivabasappa & Niu, 2020; Chondrogianni & Marinis, 2011). An explanation for our findings may be that bilingual children have received an adequate quantity of input to enable them to acquire and successfully process object which-questions. This would be consistent with the non-linear effects found in Peña et al. (2020). Moreover, under the tentative hypothesis that processing in an additional language is indeed slower due to the bilingual child having two linguistic systems instead of reduced proficiency and/or exposure, then factors of bilingual experience may be less significant predictors of the child's processing performance.

A further challenge is that unlike in other studies on sentence processing in bilingual children (e.g., Russian–German in Lemmerth & Hopp, English–Spanish in Lew-Williams, French–German in Roesch & Chondrogianni), we recruited children with varying language backgrounds. Given the varied nature of our bilingual sample, it was not possible to investigate cross-linguistic effects in our study. However, according to the World Atlas of Language Structure (Dryer, 2013), the bilingual participants' L1s are all similar to English in terms of the linguistic features used in this study, i.e., there is subject-verb agreement, number marking on nominals and wh-question fronting. While we thus do not draw any strong conclusions about how cross-linguistic influence may have affected our results, examining how it may influence processing and comprehension of wh-questions in English would be a useful avenue of further research.

Conclusion

We examined the online processing of which-questions in bilingual children. The results show that bilingual children did not underperform monolingual children in terms of overall accuracy or overall reaction times. Moreover, they looked at 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 in the timecourse of processing,

https://doi.org/10.1017/S0305000922000253 Published online by Cambridge University Press

with slower processing in the bilingual group. However, these differences were found on a fine-grained timescale with the end result not being different to monolingual children. Qualitatively, the bilingual children did not differ significantly from the monolinguals. They had greater difficulty with object relative to subject questions in the same way as monolingual children as evidenced by the absence of significant 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 the same impact in bilingual children suggesting similar processing mechanisms.

Acknowledgments. We are indebted to all our participants for their willingness and time, their parents who consented to their participation and to the schools which helped us recruit participants and test them on their premises. Recruitment of participants was supported financially by the School of Psychology and Clinical Language Sciences, University of Reading (SPCLS). Ethics approval was granted by the SPCLS Ethics Committee (2015-115-IT). George Pontikas: Conceptualization, Methodology, Data-Collection, Data-Extraction, Data Analysis, Visualization, Software, Writing- Original draft preparation, Writing-Reviewing and Editing. Ian Cunnings: Conceptualization, Methodology, Data-Analysis, Visualization, Software, Writing- Reviewing and Editing, Supervision. Theodoros Marinis: Methodology, Data-Analysis, Supervision, Writing- Reviewing and Editing.

References

Atkinson, E., **Wagers, M. W.**, **Lidz, J.**, **Phillips, C.**, & **Omaki, A.** (2018). Developing incrementality in filler-gap dependency processing. *Cognition*, **179**, 132–149.

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.

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.

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.

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.

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: B. Haznedar & F.N. Ketrez, (eds.), *The Acquisition of Turkish in Childhood*. [*Trends in Language Acquisition Research 20*], (pp. 361–388). John Benjamins.

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**, 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.

https://doi.org/10.1017/S0305000922000253 Published online by Cambridge University Press

Clifton, C., & **Frazier, L.** (1989). Comprehending sentences with long-distance dependencies. In G. M. Carlson & M. K. Tanenhaus (Eds.), *Linguistic Structure in Language Processing* (273–317). Dordrecht: Kluwer Academic Publishers.

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.

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.

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.

Desmet, T., & **Declercq, M.** (2006). Cross-linguistic priming of syntactic hierarchical configuration information. *Journal of Memory and Language*, **54**(4), 610–632.

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.

Dryer, S. M. (2013). Position of Interrogative Phrases in Content Questions. In: M.S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Frazier, L. (1987). Syntactic processing: evidence from Dutch. *Natural Language & Linguistic Theory*, **5**(4), 519–559.

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.

Friedmann, N., **Belletti, A.**, & **Rizzi, L.** (2009). Relativized relatives: types of intervention in the acquisition of A-bar dependencies, *Lingua*, **119**(1), 67–88.

Gibson, E. (1998). Syntactic complexity: Locality of syntactic dependencies. *Cognition*, **68**, 1–75.

Goodluck, H. (2005). D(iscourse)-linking and question formation: comprehension effects in children and Broca's aphasics. In A.M., Di Sciullo (Ed.), *UG and external systems: language, brain and computation*. Amsterdam: John Benjamins Publishing Company (pp. 185–192).

Grodner, D., & **Gibson, E.** (2005). Consequences of the serial nature of linguistic input for sentenial complexity. *Cognitive science*, **29**(2), 261–290.

Hale, J. (2001). A probabilistic early parser as a psycholinguistic model. In *Proceedings of NAACL* (Vol. 2, pp. 159–166).

Harley, H., & **Ritter, E.** (2002). 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.

Harrell, F. E. (2001). *Regression modeling strategies*. Springer, New York, NY.

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.

Jakubowicz, C., & **Gutierrez, J.** (2007, February). Elicited production and comprehension of root wh-questions in French and Basque. In Presentation at the COST Meeting Cross linguistically robust stage of children's linguistic performance, Berlin.

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.

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.

Loebell, H., & **Bock, K.** (2003). Structural priming across languages. *Linguistics*, **41**(5; ISSU 387), 791–824.

Matin, E., **Shao, K.**, & **Boff, K.** (1993). Saccadic overhead: information processing time with and without saccades. *Perception & Psychophysics*, **53**, 372–80.

Mirman, D. (2014). *Growth Curve Analysis and Visualization Using R*. Chapman and Hall / CRC.

https://doi.org/10.1017/S0305000922000253 Published online by Cambridge University Press

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.

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.

Paradis, J. (2005). Grammatical morphology in children learning English as a second language. *Language, Speech, and Hearing Services in Schools*, **36**, 172–187.

Paradis, J., **Rice, M. L.**, **Crago, M.**, & **Marquis, J.** (2008). Distinguishing child second language from first language and specific language impairment. *Applied Psycholinguistics*, **29**, 689–722.

Peña, E. D., **Bedore, L. M.**, **Shivabasappa, P.**, & **Niu, L.** (2020). Effects of divided input on bilingual children with language impairment. *International Journal of Bilingualism*, **24**(1), 62–78.

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.

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.

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.

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*, (pp. 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.

Schouwenaars, A., **Hendriks, P.**, & **Ruigendijk, E.** (2018). German Children's Processing of Morphosyntactic Cues in Wh-questions. *Applied Psycholinguistics*, **39**(6), 1279–1318.

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.

Staub, A. (2010). Eye movements and processing difficulty in object relative clauses. *Cognition*, **116**, 71–86.

Stavrakaki, S. (2006). Developmental perspectives on Specific Language Impairment: Evidence from the production of wh-questions by Greek SLI children over time. *Advances in Speech Language Pathology*, **8**(4), 384–396.

Stowe, A. (1989). Parsing wh-constructions: evidence for on-line gap location. *Language and Cognitive Processes*, **1**, 227–45.

Stromswold, K. (1995). The Acquisition of Subject and Object Wh-Questions. Language *Acquisition*, **4**(1/2), 5–48.

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.

Unsworth, S. (2007). Child L2, adult L2, child L1: Differences and similarities. A study on the acquisition of direct object scrambling in Dutch. *Language Acquisition*, **14**(2), 215–217.

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.

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.

Cite this article: Pontikas G., Cunnings I., Marinis T. (2022). Online processing of which-questions in bilingual children: Evidence from eye-tracking. *Journal of Child Language* 1–37, https://doi.org/10.1017/S0305000922000253

https://doi.org/10.1017/S0305000922000253 Published online by Cambridge University Press

Appendix A

Experimental stimuli

Subject questions – Number match – Singular First NP	Object questions – Number match – Singular First NP
1. Which bear is chasing the camel?	1. Which bear is the camel chasing?
2. Which dog is stroking the owl?	2. Which dog is the owl stroking?
3. Which donkey is carrying the zebra?	3. Which donkey is the zebra carrying?
4. Which duck is tickling the chicken?	4. Which duck is the chicken tickling?
5. Which goat is pushing the cow?	5. Which goat is the cow pushing?
6. Which gorilla is kicking the horse?	6. Which gorilla is the horse kicking?
7. Which lion is spraying the elephant?	7. Which lion is the elephant spraying?
8. Which monkey is licking the lamb?	8. Which monkey is the lamb licking?
9. Which rat is kissing the rabbit?	9. Which rat is the rabbit kissing?
10. Which spider is splashing the squirrel?	10. Which spider is the squirrel splashing?
Subject questions – Number mismatch – Singular First NP	**Object questions – Number mismatch – Singular First NP**
1. Which bear is chasing the camels?	1. Which bear are the camels chasing?
2. Which dog is stroking the owls?	2. Which dog are the owls stroking?
3. Which donkey is carrying the zebras?	3. Which donkey are the zebras carrying?
4. Which duck is tickling the chickens?	4. Which duck are the chickens ticking?
5. Which goat is pushing the cows?	5. Which goat are the cows pushing?
6. Which gorilla is kicking the horse?	6. Which gorilla are the horses kicking?
7. Which lion is spraying the elephants?	7. Which lion are the elephants spraying?
8. Which monkey is licking the lambs?	8. Which monkey are the lambs licking?
9. Which rat is kissing the rabbits?	9. Which rat are the rabbits kissing?
10. Which spider is splashing the squirrels?	10. Which spider are the squirrels splashing?
Subject questions – Number match – Plural First NP	**Object questions – Number match – Plural First NP**
1. Which bears are chasing the camels?	1. Which bears are the camels chasing?
2. Which dogs are stroking the owls?	2. Which dogs are the owls stroking?
3. Which donkeys are carrying the zebras?	3. Which donkeys are the zebras carrying?
4. Which ducks are tickling the chickens?	4. Which ducks are the chickens tickling?
5. Which goats are pushing the cows?	5. Which goats are the cows pushing?
6. Which gorillas are kicking the horses?	6. Which gorillas are the horses kicking?
7. Which lions are spraying the elephants?	7. Which lions are the elephants spraying?
8. Which monkeys are licking he lambs?	8. Which monkeys are the lambs licking?

https://doi.org/10.1017/S0305000922000253 Published online by Cambridge University Press

(Continued)

Subject questions – Number match – Plural First NP	Object questions – Number match – Plural First NP
9. Which rats are kissing the rabbits?	9. Which rats are the rabbits kissing?
10. Which spiders are splashing the squirrels?	10. Which spiders are the squirrels splashing?
Subject questions – Number mismatch – Plural First NP	**Object questions – Number mismatch – Plural First NP**
1. Which bears are chasing the camel?	1. Which bears is the camel chasing?
2. Which dogs are stroking the owl?	2. Which dogs is the owl stroking?
3. Which donkeys are carrying the zebra?	3. Which donkeys is the zebra carrying?
4. Which ducks are tickling the chicken	4. Which ducks is the chicken ticking?
5. Which goats are pushing the cow?	5. Which goats is the cow pushing?
6. Which gorillas are kicking the horse?	6. Which gorillas is the horse kicking?
7. Which lions are spraying the elephant?	7. Which lions is the elephant spraying?
8. Which monkeys are licking the lamb?	8. Which monkeys is the lamb licking?
9. Which rats are kissing the rabbit?	9. Which rats is the rabbit kissing?
10. Which spiders are splashing the squirrel?	10. Which spiders is the squirrel splashing?

https://doi.org/10.1017/S0305000922000253 Published online by Cambridge University Press

https://doi.org/10.1017/S0305000922000253 Published online by Cambridge University Press

Appendix B

Overview of all baseline measures

Task	Mean score for monolingual children (SD)	Mean score for bilingual children (SD)	Test statistic	Effect size	Significance
	Receptive language				
C&F-raw	50.11 (3.19)	48.03 (4.23)	U = 392.000	z = –2.103	p = 0.035*
C&F-scaled	11.83 (1.69)	10.77 (2.14)	U = 3 86.000	z = –2.197	p = 0.028*
TROG-2 raw (n blocks passed)	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

Task	Mean score for monolingual children (SD)	Mean score for bilingual children (SD)	Test statistic	Effect size	Significance
	Phonological Short Term Memory (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
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

C&F: Concepts & following directions; TROG-2: Test of Reception of Grammar; WS: Word Structure; FS: Formulated Sentences; RTWF: Renfrew Test of Word Finding; CNRep; Children's Test of Nonword Repetition; RS: Recalling Sentences (previous tests from CELF-4); CLS: Core Language Score; RLI: Receptive Language Score; ELI: Expressive Language Score; LMI: Language Memory Index (composite scores from CELF-4); ANOVAs were used for normally distributed data – Mann Whitney tests when they were not.

https://doi.org/10.1017/S0305000922000253 Published online by Cambridge University Press

Appendix C

Full models fitted for accuracy and reaction time data

Table 8. Secondary model for accuracy motivated by structure by number and structure by first NP interactions: fixed effects for subject/object questions separately.

	Subject questions				Object Questions			
Variable	β	SE	z value	p	β	SE	t value	p
	4.13	0.24	17.49	<0.0001	2.26	0.16	13.93	<0.0001
Group	0.12	0.18	0.68	0.495	0.06	0.15	0.40	0.688
Number	−0.156	0.15	−1.07	0.286	0.21	0.07	3.02	0.003
First NP	−0.36	0.15	−2.49	0.013	−0.03	0.07	−0.46	0.649
Age	0.49	0.18	2.80	0.005	0.49	0.14	3.41	0.001
CLS	0.325	0.15	2.27	0.023	0.49	0.14	3.55	0.001
Group x Number	−0.01	0.15	−0.06	0.95	−0.02	0.07	−0.23	0.821
Group x First NP	0.19	0.15	1.36	0.175	0.03	0.07	0.45	0.651
Number x First NP	−0.05	0.15	−0.34	0.737	<0.01	0.07	<0.01	0.99
Group x Number x First NP	0.09	0.15	0.67	0.506	0.01	0.07	0.15	0.879

Table 9. Secondary models for reaction times motivated by the group by number by First NP interactions: fixed effects for monolingual and bilingual children separately

	Monolingual children				Bilingual children			
Variable	β	SE	t value	p	β	SE	t value	p
	2416.88	128.96	18.74	<0.001	2801.75	312.23	8.97	<0.001
Structure	275.61	41.86	6.59	<0.001	417.95	94.94	4.40	<0.001
Number	−38.17	68.71	−0.56	0.593	−125.78	73.14	−1.72	0.121
First NP	−19.31	68.40	−0.28	0.784	48.21	78.18	0.62	0.552
Structure x Number	5.94	57.21	0.10	0.920	65.33	84.79	0.77	0.460
Structure x First NP	2.27	34.92	0.07	0.948	36.31	81.72	0.44	0.667
Number x First NP	42.57	62.44	0.68	0.512	−123.51	75.11	−1.64	0.114
Structure x Number x First NP	−31.72	34.92	−0.91	0.364	−14.03	58.94	−0.24	0.812

https://doi.org/10.1017/S0305000922000253 Published online by Cambridge University Press

https://doi.org/10.1017/S0305000922000253 Published online by Cambridge University Press

Table 10. Full fixed effects for gaze data

Variable	β	SE	t	p	Variable	β	SE	t	p
Group	0.03	0.04	0.78	0.438					
Structure	−0.19	0.02	−9.25	<0.0001	Group x Structure x Time_3	0.00	0.02	−0.20	0.842
Number	0.02	0.02	1.20	0.231	Group x Number x Time_1	−0.06	0.02	−2.89	0.004
First NP	0.01	0.02	0.26	0.795	Group x Number x Time_2	−0.03	0.04	−0.86	0.389
Time_1	0.30	0.02	14.14	<0.0001	Group x Number x Time_3	−0.02	0.02	−1.23	0.221
Time_2	0.47	0.04	13.07	<0.0001	Structure x Number x Time_1	−0.02	0.02	−0.97	0.331
Time_3	0.37	0.02	23.81	<0.0001	Structure x Number x Time_2	−0.09	0.04	−2.57	0.010
Age	0.05	0.03	1.45	0.153	Structure x Number x Time_3	0.00	0.02	−0.07	0.942
CLS	0.01	0.03	0.32	0.753	Group x First NP x Time_1	−0.03	0.02	−1.50	0.132
Group x Structure	0.00	0.02	−0.05	0.962	Group x First NP x Time_2	−0.05	0.04	−1.29	0.197
Group x Number	−0.01	0.02	−0.53	0.599	Group x First NP x Time_3	0.00	0.02	−0.26	0.792
Structure x Number	0.07	0.02	4.17	<0.0001	Structure x First NP x Time_1	−0.04	0.02	−1.77	0.077
Group x First NP	0.03	0.02	1.53	0.128	Structure x First NP x Time_2	0.04	0.04	1.11	0.269
Structure x First NP	−0.01	0.02	−0.40	0.691	Structure x First NP x Time_3	0.02	0.02	1.11	0.267
Number x First NP	0.02	0.02	1.28	0.200	Number x First NP x Time_1	0.04	0.02	1.93	0.054
Group x Time_1	−0.09	0.02	−4.21	<0.0001	Number x First NP x Time_2	−0.04	0.04	−1.02	0.307
Group x Time_2	−0.16	0.04	−4.49	<0.0001	Number x First NP x Time_3	0.01	0.02	0.63	0.528
Group x Time_3	−0.08	0.02	−5.28	<0.0001	Group x Structure x Number x First NP	−0.01	0.02	−0.65	0.515
Structure x Time_1	0.09	0.02	4.05	<0.0001	Group x Structure x Number x Time_1	−0.02	0.02	−0.85	0.398
Structure x Time_2	0.04	0.04	1.21	0.225	Group x Structure x Number x Time_2	0.02	0.04	0.64	0.523
Structure x Time_3	0.11	0.02	6.93	0.000	Group x Structure x Number x Time_3	−0.03	0.02	−1.64	0.101

https://doi.org/10.1017/S0305000922000253 Published online by Cambridge University Press

Table 10. *(Continued)*

Variable	β	SE	t	p	Variable	β	SE	t	p
Number x Time_1	0.03	0.02	1.27	0.204	Group x Structure x First NP x Time_1	0.01	0.02	0.59	0.554
Number x Time_2	−0.05	0.04	−1.49	0.135	Group x Structure x First NP x Time_2	0.03	0.04	0.83	0.408
Number x Time_3	−0.03	0.02	−1.81	0.070	Group x Structure x First NP x Time_3	0.03	0.02	1.82	0.069
First NP x Time_1	0.02	0.02	0.91	0.363	Group x Number x First NP x Time_1	0.07	0.02	3.10	0.002
First NP x Time_2	−0.02	0.04	−0.66	0.506	Group x Number x First NP x Time_2	0.00	0.04	0.10	0.919
First NP x Time_3	0.02	0.02	1.46	0.143	Group x Number x First NP x Time_3	0.01	0.02	0.83	0.405
Group x Structure x Number	−0.02	0.02	−1.37	0.170	Structure x Number x First NP x Time_1	0.01	0.02	0.39	0.696
Group x Structure x First NP	−0.01	0.02	−0.80	0.426	Structure x Number x First NP x Time_2	0.10	0.04	2.95	0.003
Group x Number x First NP	−0.01	0.02	−0.51	0.612	Structure x Number x First NP x Time_3	0.04	0.02	2.52	0.012
Structure x Number x First NP	−0.07	0.02	−4.31	0.000	Group x Structure x Number x Time_1	0.02	0.02	0.75	0.455
Group x Structure x Time_1	0.00	0.02	−0.16	0.870	Group x Structure x Number x First NP x Time_2	0.04	0.04	0.98	0.325
Group x Structure x Time_2	0.03	0.04	0.75	0.456	Group x Structure x Number x First NP x Time_3	0.01	0.02	0.50	0.619
For bilingual children only	β	SE	z value	p					
LoE	−0.08	0.05	−1.46	0.160					
Dominance	0.05	0.06	0.83	0.417					
Exposure	−0.01	0.07	−0.14	0.892					

https://doi.org/10.1017/S0305000922000253 Published online by Cambridge University Press

Table 11. Secondary model for gaze data motivated by structure by number: fixed effects for subject/object questions separately.

	Subject questions				Object Questions			
Variable	β	SE	z value	p	β	SE	t value	p
Group	0.03	0.04	0.86	0.394	0.01	0.04	0.23	0.820
Number	−0.04	0.02	−1.69	0.092	0.08	0.03	3.24	0.001
First NP	0.01	0.02	0.25	0.801	−0.01	0.03	−0.34	0.731
Time_1	0.22	0.03	8.24	<0.0001	0.41	0.03	12.30	<0.0001
Time_2	0.41	0.05	8.99	<0.0001	0.50	0.05	9.35	<0.0001
Time_3	0.25	0.02	12.98	<0.0001	0.49	0.02	20.53	<0.0001
Age	0.06	0.03	1.90	0.063	0.04	0.04	1.17	0.247
CLS	0.04	0.04	1.10	0.275	−0.02	0.04	−0.44	0.659
Group x Number	0.03	0.02	1.22	0.221	−0.03	0.03	−1.20	0.229
Group x First NP	0.04	0.02	1.60	0.112	0.02	0.03	0.58	0.563
Number x First NP	0.10	0.02	4.19	<0.0001	−0.06	0.03	−2.27	0.023
Group x Time_1	−0.09	0.03	−3.45	<0.0001	−0.09	0.03	−2.64	0.008
Group x Time_2	−0.19	0.05	−4.15	<0.0001	−0.12	0.05	−2.30	0.022
Group x Time_3	−0.08	0.02	−4.06	<0.0001	−0.08	0.02	−3.41	0.001
Number x Time_1	0.04	0.03	1.63	0.104	0.01	0.03	0.25	0.801
Number x Time_2	0.03	0.05	0.74	0.462	−0.15	0.05	−2.84	0.004
Number x Time_3	−0.03	0.02	−1.51	0.132	−0.02	0.02	−0.92	0.356
First NP x Time_1	0.06	0.03	2.03	0.043	−0.01	0.03	−0.26	0.796
First NP x Time_2	−0.06	0.05	−1.31	0.190	0.02	0.05	0.31	0.755

https://doi.org/10.1017/S0305000922000253 Published online by Cambridge University Press

Table 11. *(Continued)*

Variable	Subject questions				Object Questions			
	β	SE	z value	p	β	SE	t value	p
First NP x Time_3	0.00	0.02	0.19	0.852	0.04	0.02	1.77	0.077
Group x Number x First NP	0.01	0.02	0.24	0.814	−0.02	0.03	−0.85	0.396
Group x Number x Time_1	−0.05	0.03	−1.70	0.090	−0.07	0.03	−2.05	0.041
Group x Number x Time_2	−0.05	0.05	−1.20	0.232	0.00	0.05	0.01	0.991
Group x Number x Time_3	0.01	0.02	0.31	0.755	−0.03	0.02	−1.40	0.161
Group x First NP x Time_1	−0.04	0.03	−1.63	0.104	−0.03	0.03	−0.90	0.367
Group x First NP x Time_2	−0.08	0.05	−1.71	0.088	−0.04	0.05	−0.65	0.514
Group x First NP x Time_3	−0.03	0.02	−1.70	0.089	0.02	0.02	0.80	0.426
Number x First NP x Time_1	0.04	0.03	1.32	0.188	0.05	0.03	1.46	0.144
Number x First NP x Time_2	−0.14	0.05	−3.09	0.002	0.07	0.05	1.25	0.211
Number x First NP x Time_3	−0.03	0.02	−1.77	0.077	0.05	0.02	1.99	0.046
Group x Number x First NP x Time_1	0.05	0.03	1.90	0.058	0.09	0.03	2.58	0.010
Group x Number x First NP x Time_2	−0.03	0.05	−0.64	0.519	0.04	0.05	0.73	0.469
Group x Number x First NP x Time_3	0.01	0.02	0.34	0.732	0.01	0.02	0.44	0.660