RESEARCH

A discourse account of intervention phenomena: An investigation of interrogatives

Nick G. Riches[1] and Maria Garraffa[2]

[1] Newcastle University, UK

[2] Herriot Watt University, UK

Corresponding author: Nick G. Riches (nick.riches@newcastle.ac.uk)

Sentences where like-moves-over-like, e.g. this is the cat that the dog was chasing <the cat>, have occupied language researchers over the past two decades. They are often described as "intervention" sentences as one element intervenes in the movement of another. Such structures are difficult to comprehend by children or adults, and this effect is exacerbated in language-impaired individuals. Dominant theories, e.g. Rizzi's Relativised Minimality (RM), propose that the two NPs interfere with each other by virtue of having overlapping features. However, such sentences are also rarely encountered due to discourse constraints. For example, subject NPs (the dog) tend to be pronominal as they are typically aligned with topic-hood. This paper investigates whether discourse can account for intervention in questions. It employs a mixed methodology. Firstly, corpora were investigated to assess the degree to which discourse impacts on input frequency. Secondly, a behavioural study was conducted to unpack the relationship between frequency and processing in children. It was found that the input frequencies of intervention structures are predominantly influenced by discourse, and that intervention structures are vanishingly rare in the input. However, a link between frequency and processing was not observed, with the findings more supportive of RM. It is suggested that a consideration of discourse as an external phenomenon may yield new insights into intervention structures.

1 Introduction

The term "intervention" is sometimes used to refer to sentences where an argument "crosses over" another argument of a similar type during movement from its initial syntactic position to its surface position, e.g.

(1)   This is **the cat** that **the dog** was chasing $<$**the cat**$>$.

(2)   **Which cat** was **the dog** chasing $<$**which cat**$>$?

Both sentences contain a long-distance dependency linking a Noun Phrase (NP) and the position where it originates prior to movement. This position is often referred to as the "trace", and here is shown using triangular brackets. Inside this dependency lies an NP of a similar structural type (*Det + N*) as the moved element. Such structural configurations have aroused considerable interest in the psycholinguistic literature because they are late to be acquired and difficult to process, especially by language-impaired individuals (Garraffa & Grillo 2008; Friedmann et al. 2009). Difficulties are often explained in terms of interference between two NPs with similar properties, as demonstrated by a range of studies which find that processing is improved when NPs are made dissimilar. These

have manipulated discourse newness (Gordon et al. 2001; Kidd et al. 2007), grammatical gender (Adani et al. 2010), and animacy (Garraffa & Grillo 2008; Gennari et al. 2012).

However, interfering NPs such as those found in (1) and (2) are also highly unlikely for discourse reasons. This paper investigates the role of discourse in intervention phenomena, and whether it provides an alternative means of explaining why such sentences are difficult to process which does not presuppose interference between like NPs. To achieve this, it will employ mixed methods, firstly exploring the input frequency of intervention structures in natural language corpora, and secondly, investigating the processing of intervention and non-intervention structures in children.

One of the first accounts to address like-over-like phenomena was Rizzi's Relativised Minimality (RM) (1990; see also Starke 2001, for a more recent development of this framework). Its basic premise is that movement is outlawed when like moves over like. It was originally proposed to account for ungrammatical sentences such as (3c) where a wh-word moves over another wh-word of the same type (i.e. referring to either an argument or an adjunct).

(3)   a.   **How** did you solve the problem <**how**>?
        b.   I wonder who solved the problem in this way.
        c.   ***How** do you wonder **who** solved the problem <**how**>?

RM has since expanded to account for a wide range of psycholinguistic data. Comprehension difficulties in agrammatic aphasia were addressed by first Grillo (2005), then Garraffa and Grillo (2008) who applied RM to motivate the poor comprehension of object extracted questions and object relatives where both moved and intervening items are animate and "lexically-restricted" (i.e. containing an open-class noun slot). Comprehension in young children (mean age 4;6) was investigated by Friedmann et al. (2009) who observed particular difficulties with sentences such as (1) and (2). According to the authors, young children may be operating with a particularly strict version of RM, such that sentences are outlawed where moved and intervening NPs share even a single feature. By contrast, the adult grammar accepts featural overlap provided that the features of the intervening NP are a subset of the features of the moved NP. Such a subset relationship occurs in the adult grammar as moved NPs are bestowed with extra features. Older children (age > 10;0) with Specific Language Impairment (SLI), a condition characterised by severe unexplained language difficulties, also experience severe difficulties comprehending sentences such as (1) and (2) (Novogrodsky & Friedmann 2006; Friedmann & Novogrodsky 2011). In this sense, they resemble younger language-typical children.

In addition to explaining comprehension and production difficulties across different populations, RM can also account for avoidance phenomena, whereby speakers deliberately avoid producing a structure where a violation of RM occurs. For example, instead of producing an object relative with two lexically-restricted NPs, speakers tend to produce a relativized passive, e.g. *there's **the cat** that **the dog** chased <the cat>* → *there's **the cat** that was chased <**the cat**> by **the dog***. Here the agent *dog* has been moved outside of the syntactic dependency, thus avoiding a violation of RM. This tendency has been robustly demonstrated in elicitation tasks across many languages including Spanish, Serbian, Italian and Mandarin (Gennari et al. 2012; Contemori & Belletti 2013; Hsiao & MacDonald 2016). It has also been observed in sentence repetition by children (Novogrodsky & Friedmann 2006; Riches et al. 2010).

Complementing RM are psycholinguistic models which argue that similarity-based interference takes place in working memory, and this, in turn, affects the retrieval process. For

example, Gordon et al. (2002) found that interference effects are greater for non-canonical sentences. The study involved processing subject and objective relatives while remembering a list of nouns, which were either similar or dissimilar to the NPs in the sentences (dancer-fireman versus Joey-fireman). The authors found that comprehension of object relatives deteriorated sharply when similarity increased, resulting in a structure (subject versus object relative) by similarity interaction. Self-paced reading tasks have also identified stronger similarity/interference effects in non-canonical sentences, leading to longer latencies in the region of the second NP (Gordon et al. 2001, 2002, 2004). There is also clear evidence that avoidance in production is strongly influenced by interference. Passivised relatives, e.g. *there's **the cat** that was chased by **the dog**,* achieve the maximum possible separation between two interfering NPs thereby facilitating processing. When this interference is increased by ensuring that both NPs are animate, rates of avoidance also increase (Gennari et al. 2012).

RM and working memory accounts are closely-related in their focus on similarity-based interference. They differ in the locus of similarity, with working memory accounts open to the possibility that interfering features may be non-syntactic. For example, in a study of avoidance by English and Spanish speakers, Gennari et al. (2012), found that questionnaire-based measures of semantic similarity, e.g. an elf is more similar to a satyr than an astronaut, predicted the degree to which avoidance is observed. However, despite this difference, there are clearly many similarities between RM and working memory accounts. In fact, Rizzi (2013) argues that working memory accounts and RM can be viewed as complementary theories, which operate at different levels of description.

A further recent framework focusing on interference is the PDC (Production-Distribution-Comprehension) account (Gennari et al. 2012; MacDonald 2013, 2015). This argues that interference is predominantly a constraint on language production, not comprehension. Speakers avoid producing sentences where there is strong interference between NPs, e.g. object relatives with lexically-restricted NPs; *there's the boy that the girl pushed*. Consequently, we rarely hear such sentences, and this, in turn, impacts on our ability to comprehend these sentences. The PDC is similar to both RM and working memory accounts in its focus on interference. However, it is more specific regarding the locus of interference effects, which are believed to arise during production, and are then internalised during comprehension.

This study investigates an alternative approach to intervention phenomena based on discourse. The term "discourse" refers to a "continuous stretch of […] language larger than a sentence" (Crystal 2008: 148). Discourse plays an important role in shaping both lexical and syntactic choices. During production, the speaker must ensure that the current utterance is consistent with previous utterances. For example, they must keep track of which referents are new to the discourse in order to select the right forms; pronouns for discourse-old (given) information, and lexically-restricted NPs for discourse-new information. In addition, the discourse status of NPs determines their position in the sentence, with discourse-old NPs often placed in subject position, which is typically used for the topic of the conversation, and discourse-new NPs in a non-topic position, e.g. the object position (DuBois 1987). Furthermore, speakers may select particular syntactic structures according to discourse requirements. For example, object relatives serve to make the head NP (hereafter called NP1) discourse-relevant, a process which Fox and Thompson (1990) describe as "grounding". Finally, the discourse properties of particular slots (subject versus object) may interact with the discourse properties of constructions. For example, the NP2 slot within an object relative clause, e.g. *there's the boy that **she** pushed,* may experience a

double pressure to be pronominal, firstly because it is a subject, and secondly because it exists within a construction with a grounding discourse function.

It is relatively uncontroversial that discourse affects lexical and syntactic choices. However, some researchers have gone further to suggest that discourse may impact directly on language processing (Kidd et al. 2007; Reali & Christiansen 2007). This claim will be referred to as the "discourse account". Discourse may influence processing in two stages. Firstly, it impacts on frequency, defined not purely at the level of syntactic structures (e.g. subject versus object relatives), but also "at the level of abstract cues (e.g. animacy, givenness) and lexical items (i.e. pronouns) that are associated with particular sentence positions" (Ambridge et al. 2015: 3). According to the second stage of the argument, frequency impacts on processing. For example, object relatives with two-lexically restricted NPs are difficult to process because, for discourse reasons, they are rare in the input. An important characteristic of this account is that discourse is not conceptualised as features within a syntactic or cognitive system, but rather is viewed as a primary driver of linguistic usage.

Evidence for the role of input is provided by Reali and Christiansen (2007). While impersonal *it* is relatively rare in the NP2 position of object relatives, with personal pronouns (*he, she*) being much more common, e.g. *there's the boy that **she** pushed,* the two types of pronouns are evenly balanced in the NP2 position of subject relatives, e.g. *there's the girl that pushed **it/him**.* This is because inanimate entities are rarely agents, and are therefore rarely placed in subject position, which is prototypically agentive. Reali and Christiansen tested whether this distributional pattern impacts on processing. They used a self-paced reading task to investigate the processing of subject and object relatives with a range of pronouns in NP2 position. While personal pronouns (*you, them*) facilitated the processing of object relatives in comparison to subject relatives, the reverse pattern was observed for the impersonal pronoun (*it*). Consequently, the processing data were consistent with the frequency data. Importantly, such a pattern cannot readily be explained by interference accounts as all conditions employed a lexically-restricted NP1 which was matched to a pronominal NP2 with identical animacy properties, and therefore interference was identical across all conditions.

On a theoretical level, a number of mechanisms have been proposed to explain how discourse may drive processing. Firstly, discourse pressures lead to frequently occurring chunks. For example, pronominal pressure on the NP2 in object relatives results in the pronoun + verb chunk, e.g. *there's the boy that **she chased**.* This type of chunk, consisting of an open-class plus a closed-class element, may play an important role in processing. Abney (1991) argues that such chunks minimise syntactic ambiguity, while Reali and Christiansen (2007) propose that they facilitate processing as they may be rapidly retrieved. Secondly, case-marking on pronouns provides a cue to subject-hood (Dittmar et al. 2008). In production, intervention structures may be more difficult to produce because again, they cannot be constructed from high frequency chunks. However, this in itself does not explain the strong drive to produce an avoidance structure. One possibility is that speakers avoid placing a discourse new NP in a slot which is prototypically discourse-old (NP2). Passivisation reduces the conflict between the discourse properties of the NP and the discourse properties of the slot.

A key issue which has not been addressed is why these processes are only observed in non-canonical sentences. In comprehension, word order cues facilitate the processing of canonical sentences. By contrast, non-canonical sentences do not provide these cues, and the listener must exploit frequently occurring chunks or case-marked pronouns. When such cues are absent, comprehension is affected. Consequently, intervention effects are

not due to the *presence* of interfering NPs, but the *absence* of facilitative chunks or cues. In production, avoidance does not occur for canonical sentences such as subject relatives, because the NP2 slot is less strongly biased towards pronominal forms. This may reflect the discourse properties of subject relatives which are less likely to be used with a grounding function (Fox & Thompson 1990). In addition, a passivisation strategy for subject relatives cannot be used to add distance between interfering NPs as it will only succeed in placing the two NPs closer to each other (*there's **the girl** that **the boy** was chased by*).

It is important to note that the discourse account does not rule out the possibility of interference. Given the strong experimental evidence for interference (Gordon et al. 2002), and the fact that it explains such a wide variety of intervention phenomena (e.g. lexical restriction, animacy, number and gender), it is not surprising that it has emerged as the dominant account of intervention phenomena. However, there is also evidence that intervention structures have unusual discourse properties, that they are rare in the input, and that rare structures are more difficult to process (Kidd et al. 2007; Reali & Christiansen 2007). Consequently, the discourse account is worthy of further investigation.

Given that the discourse account argues that intervention effects are acquired from the input, one way to test it is via a fine-grained analysis of language corpora. In the words of Hsiao and MacDonald (2016: 103), studies "should […] include corpus analyses to provide broader data about the extent to which sentences that could engender similarity-based interference are truly rarer than sentences in which the relevant words are less similar". A number of studies have investigated the properties of NPs in naturally occurring non-canonical sentences. However, these focused on the properties of only one NP. While Gordon et al. (2004) and Reali and Christiansen (2007) focused on the NP2 properties in English object relatives, Hsiao and MacDonald (2016) focused on the NP1 in Mandarin object relatives. A key test for interference phenomena, which to our knowledge has not been conducted, is whether the properties of both NPs interact. This will be the aim of the first study.

The research will also extend the field of enquiry beyond relative clauses to look at question forms. Relative clauses are constructions with strong discourse properties, and are consequently a paradigm case of how discourse can affect the properties of NPs. By contrast, the discourse properties of object questions have received little attention in the literature on intervention. However, recent studies have found that these questions also exhibit strong intervention effects, particularly in language-impaired individuals (Garraffa & Grillo 2008; Friedmann & Novogrodsky 2011; Bentea et al. 2016). As the discourse function of object questions differs markedly to that of object relatives, they provide a rigorous test of the generalisability of the discourse account.

2 Study 1: Corpus analysis

2.1 Background

This study investigates the input frequency of intervention questions with a view to establishing the role of discourse. Regarding object questions there are two discourse factors which may affect the properties of NP1 (the Question Phrase) and NP2. Firstly, *which* questions, e.g. *which boy was she chasing?* are likely to be rare because they involve complex common ground. In order to ask this question, both speaker and hearer must have knowledge of a particular set of boys, only one of which is being chased. Without this shared common ground, the sentence is pragmatically infelicitous. This kind of situation, where both speaker and hearer have the same set of boys in their discourse model, rarely occurs. The second constraint affects the realisation of NP2, which tends to be pronominal; *which boy was **she** chasing?* As argued above, from a discourse perspective, the topic

of a sentence tends to be placed in subject position (Bates & MacWhinney 1982; DuBois 1987).

According to the above account, the discourse factors regulating NP1 and NP2 are independent of each other. In other words the use of a lexically-restricted NP2 does not make it more or less likely that NP1 will also be restricted. From a statistical viewpoint, if these factors are independent, we would expect the data to best described by an "additive", or main effects only model. In other words, there will be no statistical interaction between the properties of NP1 and NP2. By contrast, according to interference accounts (RM, Working Memory and PDC accounts), speakers actively avoid producing non-canonical structures where NP1 and NP2 are both closely spaced and have similar properties. This will drive down the frequency of intervention structures. From a statistical perspective, this will be manifested as an interaction between the properties of NP1 and NP2 such that the frequency of intervention structures is lower than chance. Here, chance level performance is determined from the main effects of NP1 and NP2 alone. For example, if 20% of NP1s are lexically-restricted, and 20% of NP2s are lexically-restricted, and if we assume that these factors are operating independently, we would predict that intervention structures would occur at a rate of 4% (0.2×0.2). If the frequency is below this, we have evidence for avoidance.

In addition, we can also identify avoidance by seeking out specific structures. If object questions with two lexically restricted NPs are reformulated to ensure greater distance between the NPs, this will result in a passivised object question; *which boy was being pushed by the girl?*

Overall the predictions of the different accounts are as follows:

Interference accounts (RM, working memory accounts and the PDC);

- There will be a significant statistical interaction between the discourse properties of NP1 and the discourse properties of NP2 such that frequencies of intervention questions will be lower than chance.
- There will be evidence for an avoidance strategy characterised by high rates of passivised object questions with two lexically-restricted NPs.

The discourse account;

- There will be no significant statistical interaction between the discourse properties of NP1 and the discourse properties of NP2 such that frequencies of intervention questions will be lower than chance.
- There will be no evidence for an avoidance strategy.

2.2 Procedure

The Corpus of Contemporary American English (COCA: Davies 2008) was searched for question tokens. This consists of 450 million words evenly divided between spoken sources, fiction, popular magazines, newspapers, and academic journals. It has been part-of-speech (POS) tagged using the CLAWS (the Constituent Likelihood Automatic Word-tagging System) parser (Garside 1987). Though the COCA corpus is large, which enhances generalisability, the data are taken from mainly adult sources. It consequently lacks ecological validity if we assume that processing mechanisms are consolidated during language development. To compensate, the decision was made to also analyse the Thomas Corpus (Lieven et al. 2009) from the CHILDES database (MacWhinney 1991). The Thomas corpus consists of 379 transcriptions of one-hour play sessions between a child and his caregiver. The sampling regime between 2;0 and 4;11 was especially dense, consisting

of 5 hours per week. Given this sampling density, the corpus provides a large amount of caregiver speech. Combining corpora with different characteristics may help us to evaluate patterns which are found across both corpora. For example, if the same pattern is found in both corpora we can be relatively sure of the generalisability of the findings, due to the size of the COCA corpus, and relatively sure that this pattern is also a characteristic of child-directed speech.

An important issue to mention is the different varieties of English (American versus British English). We do not know of any motivation for assuming that the basic processes governing question formation differ across these varieties. Both use the same grammatical processes to make questions, e.g. fronting of the Wh-phrase, auxiliary raising, and movement/deletion of the questioned constituent, and there are no obvious reasons for assuming that discourse constraints will differ across the two varieties. However, this issue should be borne in mind when interpreting the data.

The COCA corpus was searched using the grep function from R (R Development Core Team 2014). The search string for object questions was

(4)   (Disc. marker) + Wh-word + (NP) + Aux. verb + NP + (Aux. verb) + Lex. verb + "?"

Brackets show optional elements. The Wh-word excluded *where* and *how* which are only found in adverbial questions. These do not involve strong intervention effects as a non-argument moves over an argument. Questions beginning *which way, which direction* and *what time* were likewise excluded. The string for a lexically restricted NP in an argument position was (Det) + (Adj) + (noun) + noun. Questions with discourse markers, e.g. *so what do you think?* were highly frequent (12% of hits using the above search strings) and were therefore included. However, a separate search string; *Wh-word + Det. + N*, was used to identify exclamative questions, e.g. *what the hell was that?* which comprised 3% of hits.

In addition, subject questions were also identified using the following string;

(5)   (Disc. marker) + Wh-word + (NP) + Aux. verb + (Aux. verb) + Lex. verb + NP + "?"

Though, these are not the focus of the study, a comparison of subject and object questions can elucidate the degree to which the NP2 in object questions experiences pressure to be pronominal.

It was necessary to make a number of ad hoc decisions regarding coding. *What … do?* Questions (n = 3,241) were excluded, as *what* tends to refer to the entire verb phrase, e.g. *what did you do? I had a sandwich,* and is consequently non-argumental.

Questions with verbs which often take clausal complements, e.g. *expect* were kept. These are difficult to categorise because it is often impossible to determine whether the NP1 refers to a clausal or nominal complement, e.g. *Q: What do you expect? A: I expect [CP the boss will get angry]/[NP a lot of trouble]*. Though both constituents are arguments they differ in their syntactic properties which could influence the degree of intervention. However, this issue does not impact on the analysis. This is because if a verb which typically takes a clausal complement is inserted into an intervention question, e.g. *which questions do the students expect (in the exam)?* NP1 automatically takes a nominal reading. Consequently, questions where the NP1 could take a clausal reading, may only occur in non-intervention questions. Therefore, including questions where NP1 refers to a clausal complement will increase rates of non-intervention questions, and in proportional terms, the frequency of

intervention questions will be correspondingly reduced. This will increase the likelihood of identifying a significant interaction resulting in low rates of intervention questions. Consequently, including these items will provide a more stringent assessment of the discourse model, which does not predict an interaction effect.

Finally, subject questions containing the copula, e.g. *what is it?* were removed. This is because though NP1 is in subject position, it refers to the subject complement, e.g. *Q: What is it? A: It's a type of bird.* Consequently, it is not certain whether these can be regarded as genuine subject questions.

After the initial search, frequent question types were investigated. These followed an approximately Zipfian distribution and the top 20 are listed in the appendix. Many of these begin with *what* and exhibit specific pragmatic functions, e.g. *what do you mean?* = CLARIFICATION REQUEST, *what do you think?* = ASKING FOR AN OPINION. For such questions the argument status of *what* is debatable as, due to the specific pragmatic function, one can reply felicitously without mentioning the referent of the *wh*-word. For example, to answer *what do you mean?* one can simply paraphrase one's previous conversational turn. Such formulaic questions problematise the analysis as the loss of argument status would reduce putative intervention effects. However, there is no principled way to determine the degree of formulaicity. To address this issue, a further analysis was conducted focusing on questions where NP1 referred to a human entity, e.g. *who/which man was he following?* Though this lacks the statistical power of the complete question corpus, it exercises control over formulaic questions, given that most begin with *what*. Also, given that NP1s containing *who* cannot refer to subordinate clauses, this potential confound is controlled for.

Turning to the child data, the Thomas corpus was part-of-speech tagged using using the MOR and POST programs in CLAN (MacWhinney 1991). Object questions were extracted by identifying those ending in a verb. The discourse characteristics of NP1 and NP2 were first coded using a regular expression search, and then hand-checked.

2.3 Reliability

To investigate reliability of the COCA search the decision was made to hand check the first 200 questions for each object question type. This was not possible for the Restricted NP1 – Restricted NP2 condition where there were only 79 questions in total. Overall 16% of the entire data set was checked. Only three errors were identified which resulted from the CLAWS tagging algorithm incorrectly classifying a noun as a verb. This yields an error rate of 0.4%.

2.4 Analysis

Frequencies for the COCA are shown in Table 1. These have been plotted on a logarithmic scale in Figure 1. This ensures that visual inferences are consistent with statistical models for count data, which also employ a logarithmic transformation, e.g. loglinear models. Applying a log-transformation allows one to infer interaction effects from the relative gradient of the lines.

The data were analysed using hierarchical loglinear models, a method designed for the analysis of multi-dimensional contingency tables outlined by Howell (2011). A series of loglinear analyses are conducted, including the saturated model, and all possible nested models. If there are two independent variables A and B, then the saturated model will consist of the main effect of A, the main effect of B, and their interaction. In order to test the significance of terms they are dropped from the model. The nested model is then compared to the saturated model, using likelihood ratio tests. If, by dropping a term, there is a significant reduction in model fit, then we can conclude that the dropped term makes

Table 1: Counts of questions by QP and NP type.

Type	Example	COCA	COCA humanª	Thomas	COCA passive Qs
Restricted-Restricted	**Which boy** did **the girl** chase?	79	3	14	23
Restricted-Unrestricted.	**Which boy** did **she** chase?	663	2	253	1
Unrestricted-Restricted	**Who** did **the girl** chase?	678	12	86	10
Unrestricted-Unrestricted	**Who** did **she** chase?	9,717	288	1,139	3
		11,137	305	1,493	37

(a) First NP (Question Phrase) refers to a human.
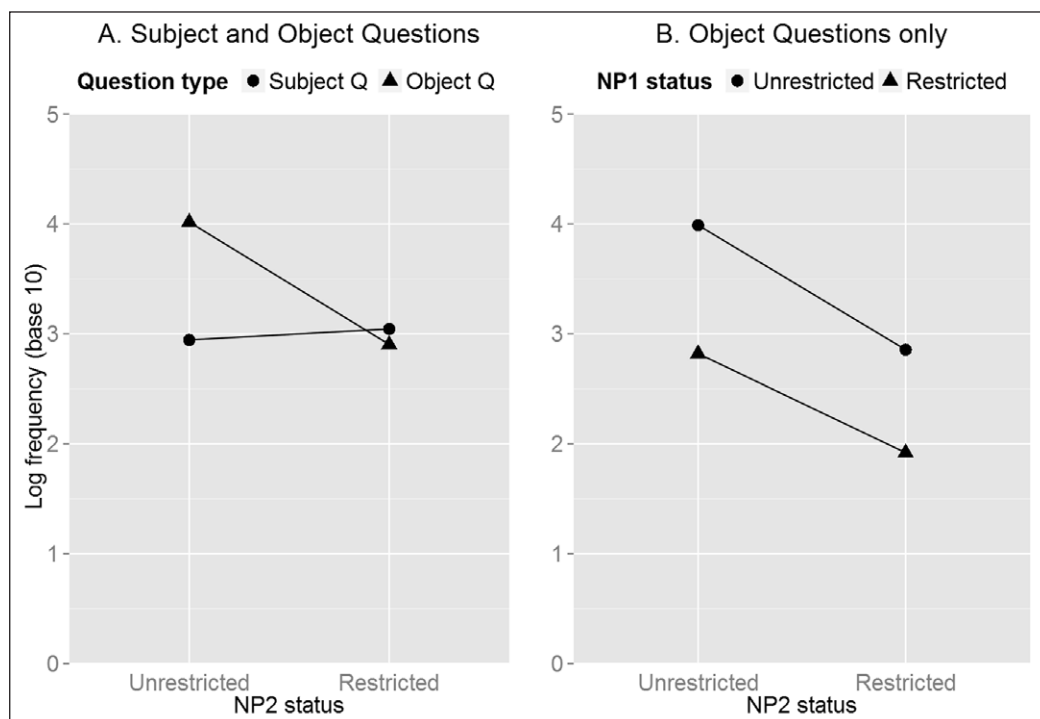
Figure 1: Lexical restriction frequencies from the COCA.

an important contribution to the model. The chi-squared statistic and p-values, which are reported in Table 2, are both derived from the likelihood ratio test. The coefficient from the loglinear model is provided to show the directionality of the effect. The table reports interactions only as they are the theoretical focus of the study, allowing us infer whether frequencies are influenced by interference between NPs. All main effects were significant (all p-values < 0.001).

It can be seen in Table 2 there was a significant interaction in the COCA data between question type (subject versus object), and the lexical restriction of NP2. This reflects a strong tendency for the NP2 in object questions to be pronominal (see Figure 1, Panel A). Subsequent analyses focused on object questions only. For the COCA data there was a significant interaction between the lexical restriction of NP1 and NP2. However, this is not consistent with the predictions of interference accounts, as Figure 1 (Panel B) demonstrates that the driver for this interaction effect is not the low rates of Restricted-Restricted object questions, but the high rates of Unrestricted-Unrestricted object questions e.g. *who did he chase?* For the reduced corpus, where NP1 referred only to human entities, there was a similar pattern, with the interaction again driven by the high

Table 2: Analyses of frequency profiles. OQ = 1 if sentence is an object question, 0 if sentences is a subject question, R1 = 1 if first NP is lexically-Restricted, otherwise 0, R2 = 1 if second NP is lexically-Restricted, otherwise 0.

Effect	Corpus	χ^2 (1 df)	Sig.	Coeff.
OQ × R2	COCA all Qs	2169.2	<0.001	–2.79
R1 × R2	COCA obj. Qs	16.9	<0.001	0.54
	COCA human obj. Qs[a]	12.2	<0.001	3.58
	Thomas obj. Qs	1.17	0.279	0.30

(a) COCA human = Question Phrase refers to a human entity.

rates of Unrestricted-Unrestricted questions. The data from the Thomas corpus were best explained by a main effects only model.

Though there was no evidence of an avoidance strategy based on the analysis of question types, it is possible that a more focused search for avoidance structures may provide support. To search for questions containing the passive, the following string was used

(6)   0 – 2 words + Wh-word + any number of words + "by" + 1 – 3 words + "?"

The search string yielded 608 utterances from the COCA corpus, which were then coded by hand. 37 genuine passive questions were identified with findings shown in Table 1. Most of the passive Qs (62.2%) were of the Restricted-Restricted type, e.g. *Which X was Verbed by the Y?* Overall, while 29% of the Restricted-Restricted object questions were produced in the passive, none of the other question types showed a passive bias. This was confirmed by a chi-square test ($\chi^2(3) = 1579.1$, $p < 0.001$***).

3 Study 2: Behavioural data

3.1 Background

Despite evidence for genuine avoidance effects, the findings of experiment 1 also demonstrated the role of discourse. For example, frequencies of Restricted-Restricted questions were not significantly lower than chance, consistent with the predictions of the discourse account. Experiment 2 tested the second stage in this model; the relationship between input frequency and processing. It focused on children as they are likely to present with large intervention effects, thereby reducing the risk of a ceiling effect. Furthermore, children are of considerable interest from a theoretical perspective given that they are far more sensitive to discourse-based intervention (Friedmann et al. 2009).

Experimentally, we can test the relationship between input frequency and processing by exposing participants to numerous exemplars, and observing whether these impact on performance. Such an approach was conducted by Wells et al. (2009), who found that participants exposed to large numbers of object relatives were indeed faster at processing these structures. However, this method is less applicable to more fine-grained statistical properties of the input, e.g. the types of NPs occurring in particular positions. We could manipulate the NP2 properties of the sentences in both the training, and the experimental task. However, given the powerful effect of NP2 type on processing (Gordon et al. 2001; Kidd et al. 2007), and the relatively low input frequencies achievable under laboratory conditions, such an effect would be hard to detect.

Rather than manipulating input frequency, an alternative approach is to investigate individual variation for items with different input frequencies. If input frequency is a key factor affecting performance, we would expect greater individual variation for low frequency

items. This is because the ability to learn low frequency items is closely linked to language ability. Children with language impairments exhibit strong frequency-dependence, requiring substantially greater input than control children to learn words (Rice et al. 1995) and place novel nouns in argument positions where they have not previously occurred (Skipp et al. 2002). They also need about twice the amount of input to extract the phonotactic probabilities from an artificial language (Evans et al. 2009).

According to this framework, patterns of individual variation may be informative about underlying learning mechanisms, in particular the role of frequency. If intervention structures are difficult due to their low input frequency, then they should be particularly difficult for low-language individuals who are highly frequency dependent. We would also expect there to be relatively large individual variation on these items as the less frequency-dependent individuals would be able to learn them from limited input. For higher-frequency non-intervention structures we would expect individual variation to be reduced. This is because the structures are sufficiently frequent for all individuals to learn, no matter how frequency-dependent they are. As discussed in the literature review, frequency effects may be related to the existence of high-frequency chunks which may be rapidly accessed (e.g. Reali & Christiansen 2007), or the presence of frequently-occurring cues, e.g. case-marking of subject pronouns (e.g. Dittmar et al. 2008).

The assumption that individual differences will be greater for infrequent structures is supported by studies comparing SLI and language-typical children. Between-group differences in comprehension of non-canonical sentences are greatest in conditions where there is interference between lexically-restricted NPs (Friedmann & Novogrodsky 2011; Frizelle & Fletcher 2014). This suggests that low frequency structures lead to greater variation in performance. Though such findings are based on children with extreme language difficulties, it is plausible that this profile could be extrapolated to children with normal range language abilities. Finally, an important proviso is that variation is systematic. In other words, those children with stronger overall language abilities will perform better on the individual structures.

To recap, the experimental hypothesis is that there will be greater systematic variation for intervention structures than non-intervention structures. Such a finding will support the claim that input frequency is a key driver of processing.

3.2 Stimuli

For the question comprehension task four sets of ten questions were created according a two-by-two design manipulating structure (Subject versus Object question), and NP1 type (lexically restricted, e.g. *which dog?* versus non-lexically-restricted, e.g. *who?*). The NP2 was always lexically-restricted. This design is identical to that used by Novogrodsky and Friedmann (2011). The questions were created to be maximally reversible so that semantic cues did not aid comprehension. Each question was matched with a pair of pictures depicting transitive scenes. While the target picture included the agent-patient relationship as depicted by the question, the foil picture reversed the agent-patient relationship. The pictures were separated by a thick black dividing line. An example is shown in the Appendix.

Four different pseudorandomised orders were created, with the proviso that there would be no more than three consecutive stimuli of same question type, and no more than three consecutive stimuli where the target appeared on the same side.

3.3 Participants and procedure

Forty-seven children, mean age 5;2 (s.d. = 3.3 months, min = 4;8, max = 5;8) were recruited from Reception classes in the North East of England. This age range represented a trade-off between the need to recruit young children with a developing language system who may be prone to childhood RM, and a need to recruit children who were suffi-

ciently old to focus on the experimental task. The children were given two comprehension tasks, the bespoke test of question comprehension (described above), and a standardised test of general grammatical comprehension; the Test of Reception of Grammar Second Edition (TROG-2: Bishop 2003). While the former tests a specific construction (subject versus object questions), the TROG-2 tests a wide range of constructions, and thus can be regarded as a more general measure of language abilities. As such, it is often included in language assessment batteries (e.g. Hsu & Bishop 2014).

For the bespoke comprehension task, the children were randomly allocated to one of the four different orders. The experimenter presented the pictures using a booklet. As each picture was presented, the experimenter produced the question, and the child was required to point to the picture corresponding to the question. For example, for the question *which rabbit was the tortoise carrying?* comprehension would be signalled by the child pointing at the rabbit in the picture of the tortoise carrying a rabbit (foil picture = a rabbit carrying a tortoise). To allow for inaccurate pointing, a point to the correct picture, i.e. where the finger fell the right side of the dividing line, was accepted.

Ethical approval was obtained from the Humanities and Social Sciences Faculty Research Ethics Committee at Newcastle University, and research was performed in accordance with the Declaration of Helsinki. Informed consent was obtained from parents and guardians. Child participants were reminded that participant was not obligatory and that they could withdraw from the research at any time.

3.4 Results

The children demonstrated age-appropriate language skills with a TROG-2 mean standard score of 97 (s.d. = 15.4). Table 3 shows the mean items correct and percentage of children performing above chance for each question type. It can be seen that performance was substantially worse for the intervention questions.

A two-way repeated measures ANOVA was conducted with canonicity (subject versus object question) and NP1 type (lexically-restricted versus non-restricted) as within-subjects variables, and number of pictures correctly identified as the dependent variable. Both main effects and interactions were modelled. There was a main effect of canonicity ($F(1,46) = 85.7$, $p < 0.001^{***}$, partial $\eta^2 = 0.40$), a main effect of NP1 type ($F(1,46) = 30.4$, $p < 0.001^{***}$, partial $\eta^2 = 0.65$), and a significant interaction ($F(1,46) = 52.0$, $p < 0.001^{***}$, partial $\eta^2 = 0.53$). This is driven by the low comprehension rates for the intervention questions.

The relationship between the TROG-2 blocks and comprehension scores by question type are show in Figure 2. It can be seen that only performance on the intervention

Table 3: Performance on the comprehension task.

Q type	NP properties	Example	Mean correct (max 10)	St. dev.	Perc. > chance	Correlation with TROG
Obj.	Restricted-Restricted	Which boy was the girl pushing?	5.79	1.73	6.3%	r = 0.01 p = 0.936
	Unrestricted-Restricted	Who was the girl pushing?	8.34	1.39	46.8%	r = 0.37 **p = 0.011****
Subj.	Restricted-Restricted	Which girl was pushing the boy?	8.77	1.29	66.0%	r = 0.41 **p = 0.005****
	Unrestricted-Restricted	Who was pushing the boy?	8.43	1.25	53.2%	r = 0.48 **p < 0.001*****
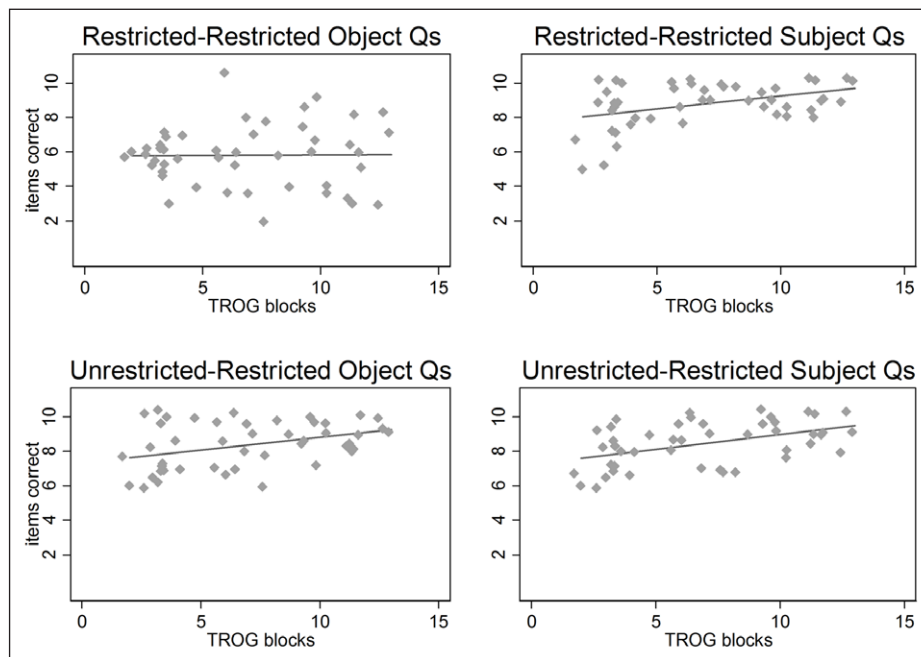
Figure 2: Relationship between the TROG-2 and comprehension scores by question type. Restricted-Restricted = Restricted NP1 + Restricted NP2, e.g. *Which cat was the dog washing?* Unrestricted-Restricted = Unrestricted NP1 + Restricted NP2, e.g. *What was the dog washing?* Spherical noise has been added to aid visual interpretation. This has brought some scores slightly above the maximum (10).

questions failed to demonstrate a strong relationship with the TROG-2 blocks. A series of Pearson product-moment correlations were conducted to investigate this relationship. The results are also shown in Table 3. Again, a significant relationship was observed for all questions except the intervention questions.

Differences between correlation coefficients were subsequently investigated. Firstly, they were converted to z-scores using Fisher's transformation. Then the z-scores were compared using a method proposed by Cohen and Cohen (1983: 54, formula 2.8.5). Analyses compared the coefficient for the intervention structures (Restricted-Restricted Object Qs) with the alternative structures, yielding three comparisons in total. The coefficient for the Restricted-Restricted Object Qs was significantly less than the coefficient for the Unrestricted-Restricted Object Qs ($z = -1.77$, $p = 0.038^*$), the Restricted-Restricted Subject Qs ($z = -2$, $p = 0.023^*$), and the Unrestricted-Restricted Subject Qs ($z = -2.41$, $p = 0.008^{**}$).

4 General discussion

The study evaluated an alternative account of intervention phenomena. This proposed that intervention structures are rare due to discourse constraints, which in turn impacts on our ability to understand and produce them. A mixed methods approach was adopted, combining a corpus study, to investigate input frequency, and a behavioural study to investigate the impact of frequency on comprehension. Findings were equivocal. Though the corpus study provided strong evidence that input frequency was primarily, though not wholly, discourse-driven, the behavioural study did not meet the predictions.

Study 1 found that the frequency of intervention questions did not fall below chance. The frequency of questions in the Thomas corpus was accounted for by an additive model incorporating the main effects of NP1 and NP2. Where interactions arose in the COCA data, they were not driven by low rates of intervention questions but by the high rates

of Unrestricted-Unrestricted questions, e.g. *what do you want?* This is likely to reflect a chunking process whereby high frequency question types become represented via multi-word units, or possibly even a single unit. The possible existence of chunks is demonstrated by the process of phonological reduction, e.g. the assimilation of the question word, auxiliary and the pronoun in *what do you want?* = /wdjʒə wɒnt/ (Bybee, 2010). Such chunking effects have been observed in young children, who depend heavily on both *wh-word + auxiliary* chunks (e.g. *What is…?*), and *Auxiliary + Subject* chunks (*…is it..?*) (Rowland & Pine 2000; Ambridge et al. 2006). However, the current data points to the existence of chunking in adult questions (the COCA is constructed overwhelmingly from adult sources). According to Reali and Christiansen (2007), chunking plays an important role in language processing as high-frequency chunks are easier to retrieve. Evidence for chunk-based language use is consistent with the discourse account, as chunking is one possible mechanism whereby discourse is thought to drive processing (Reali & Christiansen 2007).

It is interesting to note that the chunking effect and the intervention effect were in opposition with each other within the two-by-two design. In other words, an increased chunking effect reduced the chance of observing an intervention effect and vice versa. An implication is that the chunking effect may have actually masked the intervention effect, though it is still reasonable to conclude that the chunking effect was stronger than the intervention effect, as otherwise they would have cancelled each other out and neither would be evident.

Though an analysis of all object questions investigating the interaction between NP1 and NP2 did not find an avoidance effect, a more focused search provided evidence for avoidance. In discourse contexts where both the NPs are lexically restricted, speakers often use passive forms, e.g. **which cat** is being chased by **the dog**? To our knowledge, this is the first time that interference-motivated avoidance has been observed in naturalistic data (though see Hsiao & MacDonald 2016 for a corpus study of avoidance based on the animacy of the head NP alone). It also provides the first evidence that avoidance phenomena which characterise the production of relative clauses also extend to questions. However, this avoidance effect is comparatively weak. Only 21.4% of Restricted-Restricted questions used the passive voice. This contrasts with much higher avoidance rates for object relatives, which in English and Italian-speaking adults can reach as high as 90% (Gennari et al. 2012; Contemori & Belletti 2013). One possibility is that weaker avoidance results from the lower lexical overlap between NP1 and NP2 due to the use of different determiners (*which dog/the cat*) which consequently reduces interference.

An implication of the study is that avoidance rates are much lower in naturalistic contexts than experimental data would suggest. In support of this, Hsiao and MaDonald (2016) found that avoidance in Mandarin object relatives with two animate NPs reached 98% for behavioural data, whereas only 13% of object relatives identified by a corpus study exhibited avoidance. Moreover, the total number of avoidance structures found in the current data is low when viewed as a proportion of object questions as a whole (0.02%). The corpus data suggest that the discourse conditions under which avoidance might occur are rare, and consequently avoidance is best viewed as a strategy for dealing with an exceptional circumstance. They also question whether avoidance structures themselves can be acquired from the input given that avoidance rates in corpus data appear to be nowhere near the rates observed in experimental studies. One possible explanation is that avoidance is not learned by actually hearing avoidance structures, as proposed by the PDC model. Instead it may result from an attempt to reconcile the discourse (and possibly animacy) properties of the message-level representation with available linguistic

templates derived from *high*-frequency sentences. For example, lexically-restricted NPs are moved away from slots which are prototypically pronominal, e.g. NP2. The COCA data strongly support the claim that this slot experiences strong pronominal pressure, as demonstrated by the interaction between question type (subject versus object), and the discourse properties of NP2.

The above discussion should be tempered by the differing nature of the corpora, with the COCA consisting of American English from a variety of mainly adult sources, and the Thomas corpus consisting of data from a single child. Each corpus has different limitations with the COCA focused on adult language use, and the Thomas corpus being relatively small and derived predominantly from a single caregiver. In addition, possible differences between American and British English questions should be taken into consideration.

While study 1 found that discourse factors strongly influenced frequency, study 2 was problematic for the next stage of the model; the claim that the input frequency of different question types drives processing. Here the main prediction, that there would be more systematic variation for the non-intervention questions, was categorically not met. In fact, it was striking that children with good overall comprehension abilities found the intervention questions so difficult. This kind of dissociation between general language abilities and performance on a specific structure is consistent with maturational accounts of language acquisition (Borer & Wexler 1987; Wexler 2003). For example, Wexler (2003) notes that the disappearance of optional infinitives is completely uncorrelated with maternal education, IQ, or vocabulary scores, factors which are likely to be closely associated with general language learning abilities. This dissociation arises, according to Wexler, because the relevant linguistic parameter adheres to a genetically-determined timetable. A similar argument could be put forward to explain the data from Study 2. Intervention effects may be subject to an RM-related maturational constraint and consequently we would expect them to be divorced from domain-general language abilities. By contrast, non-intervention structures which are not governed by such a constraint should show a stronger relationship with overall language abilities. This is precisely the pattern observed in the current data.

Though the data supported a maturational model, there may be alternative explanations for the findings. The intervention questions may have been so infrequent (0.7% of object questions) that even children with good language abilities were unable to learn them. Consequently, the lack of systematic variation may have been the result of a floor effect. Another possibility is that the relationship between input frequency and processing difficulty may not be linear, with a link between the two observable only when input frequency crosses a threshold. This may only have been achieved for the high frequency structures. Clearly, there is a need for a more explicit characterisation of the relationship between frequency and processing, backed up by testable hypotheses.

The current data are consistent with other studies of young children. For example, the children aged 4;6 tested by Friedmann et al. (2009) also struggled substantially on intervention structures. There is some discrepancy, given that the slightly older children in the current study performed even worse (6% above chance versus 32% above chance). However, this may be related to the more child-friendly act out methodology employed by Friedmann et al. There is clearly substantial development throughout childhood as, by the age of 9;6 comprehension of intervention structures is close to ceiling (Friedmann & Novogrodsky 2011). Contemori and Belletti (2013), provide cross-sectional data demonstrating that, during an elicitation task, production of object relatives with two lexically-restricted NPs peaks around age 7, after which there is a strong drive to use avoidance structures. This suggests a protracted developmental trend. The children visited in the current study are clearly in the early stages of this journey.

It is also interesting to compare the findings with the clinical studies. These observe greater systematic differences between language-impaired and control children for intervention structures than non-intervention structures (Friedmann & Novogrodsky 2011; Frizelle & Fletcher 2014). By contrast, the current study, looking at variation in the normal range did not identify systematic differences in the intervention condition. This discrepancy could be explained by the concept of childhood RM. The young language-typical children in the current study (mean age 5;2) may still be at this stage. When older children are investigated (9;6 in Friedmann and Novogrodsky, and 6;11 in Frizelle and Fletcher) they are divided into language-typical children who have progressed beyond childhood RM, and language-impaired children who remain "stranded". However, this stage-like model of intervention effects is at odds with the protracted development identified by Contemori and Belletti (2013). Clearly there is much work to be done on the developmental time-course of intervention effects.

5 Conclusion

Intervention effects in laboratory settings are extremely powerful, impacting on both comprehension and production. There is little doubt that interference accounts provide a parsimonious explanation of these data. In addition, they exhibit strong explanatory adequacy, as they are able to explain a wide range of intervention effects, irrespective of whether features are discourse-related, and moreover provide a convincing explanation of avoidance phenomena.

Nonetheless, it is striking that intervention structures exhibit unusual discourse properties. For example, object relatives with two lexically-restricted NPs are rare given that this construction typically contains at least one NP that is discourse linked. This raises that possibility that intervention effects arise from discourse. In support of this claim, a corpus study found that, despite evidence for avoidance, question frequencies were nonetheless predominantly influenced by discourse factors. In fact, it could be argued that discourse rarely "allows" intervention to happen, when conceptualised in terms of lexical restriction.

However, it is also important to demonstrate a link between input frequency and processing. Study 2, which used individual differences as a means of studying input frequency, did not find evidence for such a link. In fact, the extreme difficulties faced by the children in the intervention condition are consistent with a maturational account of RM (Friedmann et al. 2009).

In order to become a viable model, the discourse account needs to go beyond the analysis of corpus data to empirically demonstrate the link between frequency and processing. Though there are a number of studies demonstrating that highly frequent sentence types are easier to process (Kidd et al. 2007; Reali & Christiansen 2007), there are relatively few providing firm evidence of causality, e.g. via the manipulation of input (Wells et al. 2009). On the other hand, interference accounts need to explain why avoidance effects in naturalistic data are far weaker than those found in laboratory settings. A more integrated explanation of avoidance phenomena addressing how speakers select avoidance structures in naturalistic settings is needed.

Abbreviations

ANOVA = Analysis of Variance, Adj. = adjective, Aux. verb = auxiliary verb, COCA = Corpus of Contemporary American English, CP = Complementiser Phrase, det. = determiner, Lex. verb = lexical verb, NP = Noun Phrase, NP1 = first Noun Phrase in the sentence, NP2 = second Noun Phrase in the sentence, PDC = Production-Distribution-Comprehension account, RM = Relativised Minimality, TROG-2 = Test of Reception of Grammar, 2nd Edition

Additional Files

The additional files for this article can be found as follows:

- **Appendix 1.** Example of picture from Study 2. DOI: https://doi.org/10.5334/gjgl.100.s1
- **Appendix 2.** Occurrences of most frequent questions. DOI: DOI: https://doi.org/10.5334/gjgl.100.s2

Acknowledgements

Competing Interests

The authors have no competing interests to declare.

References

Abney, Steven P. 1991. Parsing by chunks. In *Principle-based parsing,* 257–278. Springer. DOI: https://doi.org/10.1007/978-94-011-3474-3_10

Adani, Flavia, Heather K.J. van der Lely, Matteo Forgiarini & Maria T. Guasti. 2010. Grammatical feature dissimilarities make relative clauses easier: A comprehension study with Italian children. *Lingua* 120(9). 2148–2166. DOI: https://doi.org/10.1016/j.lingua.2010.03.018

Ambridge, Ben, Caroline F. Rowland, Anna L. Theakston & Michael Tomasello. 2006. Comparing different accounts of inversion errors in children's non-subject wh-questions: 'What experimental data can tell us?' *Journal of Child Language* 33(03). 519–557.

Ambridge, Ben, Evan Kidd, Caroline F. Rowland & Anna L. Theakston. 2015. The ubiquity of frequency effects in first language acquisition. *Journal of Child Language* 42(02). 239–273.

Bates, Elizabeth & Brian MacWhinney. 1982. Functionalist approaches to grammar. *Language Acquisition: The State of the Art*. 173–218.

Bentea, Anamaria, Stéphanie Durrleman & Luigi Rizzi. 2016. Refining intervention: The acquisition of featural relations in object A-bar dependencies. *Lingua* 169. 21–41. DOI: https://doi.org/10.1016/j.lingua.2015.10.001

Bishop, Dorothy V.M. 2003. *Test of Reception of Grammar (Trog-2)*. London: Harcourt Assessment.

Borer, Hagit & Kenneth Wexler. 1987. The maturation of syntax. *Parameter Setting*. 123–172.

Contemori, Carla & Adriana Belletti. 2013. Relatives and passive object relatives in Italian-speaking children and adults: Intervention in production and comprehension. *Applied Psycholinguistics*, 1–33.

Crystal, David. 2008. *A dictionary of linguistics and phonetics*. 6th edition; Blackwell. DOI: https://doi.org/10.1002/9781444302776

Davies, Mark. 2008. *The Corpus of Contemporary America English: 450 million words, 1990 – present.*

Dittmar, Miriam, Kirsten Abbot-Smith, Elena Lieven & Michael Tomasello. 2008. German childrens comprehension of word order and case marking in causative sentences. *Child Development* 79(4). 1152–1167. DOI: https://doi.org/10.1111/j.1467-8624.2008.01181.x

DuBois, John W. 1987. The discourse basis of ergativity. *Language*, 805–855.

Evans, Julia L., Jenny R. Saffran & Kathryn Robe-Torres. 2009. Statistical learning in children with specific language impairment. *Journal of Speech, Language, and Hearing Research* 52(2). 321. DOI: https://doi.org/10.1044/1092-4388(2009/07-0189)

Fox, Barbara A. & Sandra A. Thompson. 1990. A discourse explanation of the grammar of relative clauses in English conversation. *Language*, 297–316. DOI: https://doi.org/10.2307/414888

Friedmann, Naama, Adriana Belletti & Luigi Rizzi. 2009. Relativized relatives: Types of intervention in the acquisition of A-bar dependencies. *Lingua* 119(1). 67–88. DOI: https://doi.org/10.1016/j.lingua.2008.09.002

Friedmann, Naama & Rama Novogrodsky. 2011. Which questions are most difficult to understand?: The comprehension of Wh questions in three subtypes of SLI. *Lingua* 121. 367–382. DOI: https://doi.org/10.1016/j.lingua.2010.10.004

Frizelle, Pauline & Paul Fletcher. 2014. Relative clause constructions in children with specific language impairment. *International Journal of Language & Communication Disorders* 49(2). 255–264. DOI: https://doi.org/10.1111/1460-6984.12070

Garraffa, Maria & Nino Grillo. 2008. Canonicity effects as grammatical phenomena. *Journal of Neurolinguistics* 21(2). 177–197. DOI: https://doi.org/10.1016/j.jneuroling.2007.09.001

Garside, Roger. 1987. The CLAWS word-tagging system. In *The computational analysis of English: A corpus-based approach*, 30–41. London: Longman

Gennari, Silvia P., Jelena Mirković & Maryellen C. MacDonald. 2012. Animacy and competition in relative clause production: A cross-linguistic investigation. *Cognitive Psychology* 65(2). 141–176. DOI: https://doi.org/10.1016/j.cogpsych.2012.03.002

Gordon, Peter C., Randall Hendrick & Marcus Johnson. 2001. Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27(6). 1411. DOI: https://doi.org/10.1037/0278-7393.27.6.1411

Gordon, Peter C., Randall Hendrick & Marcus Johnson. 2004. Effects of noun phrase type on sentence complexity. *Journal of Memory and Language* 51(1). 97–114. DOI: https://doi.org/10.1016/j.jml.2004.02.003

Gordon, Peter C., Randall Hendrick & William H. Levine. 2002. Memory-load interference in syntactic processing. *Psychological Science* 13(5). 425–430. DOI: https://doi.org/10.1111/1467-9280.00475

Grillo, Nino. 2005. Minimality effects in agrammatic comprehension. In Sylvia Blaho, Erik Schoorlemmer & Luis Vicente (eds.), *Proceedings of conSOLE XIII*, 106–120.

Howell, David C. 2011. *Statistical methods for psychology*. Cengage Learning.

Hsiao, Yaling & Maryellen C. MacDonald. 2016. Production predicts comprehension: Animacy effects in Mandarin relative clause processing. *Journal of Memory and Language* 89. 87–109. DOI: https://doi.org/10.1016/j.jml.2015.11.006

Hsu, Hsinjen Julie & Dorothy V.M. Bishop. 2014. Sequence-specific procedural learning deficits in children with specific language impairment. *Developmental Science* 17(3). 352–365. DOI: https://doi.org/10.1111/desc.12125

Kidd, Evan, Silke Brandt, Elena Lieven & Michael Tomasello. 2007. Object relatives made easy: A cross-linguistic comparison of the constraints influencing young children's processing of relative clauses. *Language and Cognitive Processes* 22(6). 860–897. DOI: https://doi.org/10.1080/01690960601155284

Lieven, Elena, Dorothé Salomo & Michael Tomasello. 2009. Two-year-old children's production of multiword utterances: A usage-based analysis. *Cognitive Linguistics* 20(3). 481–507. DOI: https://doi.org/10.1515/COGL.2009.022

MacDonald, Maryellen C. 2013. How language production shapes language form and comprehension. *Frontiers in Psychology* 4. DOI: https://doi.org/10.3389/fpsyg.2013.00226

MacDonald, Maryellen C. 2015. The emergence of language comprehension. In *The handbook of language emergence*, 81–99. DOI: https://doi.org/10.1002/9781118346136.ch3

MacWhinney, Brian. 1991. *The Child Language Data Exchange System (CHILDES)*. Laurence Erlbaum Associates.

Novogrodsky, Rama & Naama Friedmann. 2006. The production of relative clauses in syntactic SLI: A window to the nature of the impairment. *International Journal of Speech-Language Pathology* 8(4). 364–375. DOI: https://doi.org/10.1080/14417040600919496

R Development Core Team. 2014. *R: A Language Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Reali, Florencia & Morten H. Christiansen. 2007. Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language* 57(1). 1–23. DOI: https://doi.org/10.1016/j.jml.2006.08.014

Rice, Mabel L., JoAnn Buhr & Janna B. Oetting. 1995. Specific Language Impaired Children's Quick Incidental Learning (QUIL) of words: The effect of a pause. *Journal of Speech and Hearing Disorders* 55. 33–42. DOI: https://doi.org/10.1044/jshd.5501.33

Riches, Nick G., Tom Loucas, Tony Charman, Emily Simonoff & Gillian Baird. 2010. Sentence repetition in adolescents with Specific Language Impairments and autism: An investigation of complex syntax. *International Journal of Language and Communication Disorders* 45(1). 47–60. DOI: https://doi.org/10.3109/13682820802647676

Rizzi, Luigi. 1990. *Relativized minimality*. Cambridge, MA: The MIT Press.

Rizzi, Luigi. 2013. Locality. *Lingua* 130. 169–186. DOI: https://doi.org/10.1016/j.lingua.2012.12.002

Rowland, Caroline F. & Julian M. Pine. 2000. Subject-auxiliary inversion errors and wh-question acquisition: What children do know. *Journal of Child Language* 27(01). 157–181. DOI: https://doi.org/10.1017/S0305000999004055

Skipp, Amy, Kirsten Windfuhr & Gina Conti-Ramsden. 2002. Children's grammatical categories of verb and noun: A comparative look at children with Specific Language Impairment (SLI) and normal language (NL). *International Journal of Language and Communication Disorders* 37(3). 253–271. DOI: https://doi.org/10.1080/13682820110119214

Starke, Michal. 2001. *Move dissolves into Merge: A theory of locality*. Geneva: University of Geneva dissertation.

Wells, Justine B., Morten H. Christiansen, David S. Race, Daniel J. Acheson & Maryellen C. MacDonald. 2009. Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology* 58(2). 250–271. DOI: https://doi.org/10.1016/j.cogpsych.2008.08.002

Wexler, Kenneth. 2003. Lenneberg's dream: Learning, normal language development, and specific language impairment. In Yonata Levy & Jeannette C. Schaeffer (eds.), *Language competence across populations: Toward a definition of specific language impairment*, 11–61. Lawrence Earlbaum Associates.