Understanding semantic competition in complex phrase comprehension and production

Shihui Wu

PhD

University of York

Psychology

September 2019

Abstract

This thesis investigated the relationship between complex phrase comprehension and production. The work aimed to identify whether parallel competition effects exist between the two tasks; and importantly, the extent to which these effects drive from common processes and knowledge bases. In three studies, participants viewed pictures of various entities doing different actions. For each picture, they comprehended a recorded description (e.g. the teddy bear/man that the girl is hugging) about the entity being acted upon, and were asked to describe it in a production task. They also completed a number of cognitive assessments measuring vocabulary, inhibition, etc. Study 1 found that phrases containing highly similar and reversible nouns are more difficult to comprehend and produce in adults. Importantly, this difficulty varied as a function of individual inhibition skills over above vocabulary in both tasks, and production additionally recruited task-specific motor inhibition processes. Study 2 replicated the reversibility-based effects with children and adolescents. But young children differed from older participants as they experienced greater production interference and are less skilled in using certain production options to alleviate interference. Unlike adults, their language performance was predicted by variance on working memory capacity. Study 3 used eye-tracking to examine the time course of production competition. The results showed reversibility-based competition manifest at verb position, and is particularly relevant to individual's semantic inhibition skill. This parallels previous comprehension findings, thus suggests shared competition resolution processes across tasks. Together, these findings suggest common reversibility-based competition processes underlying comprehension and production, and across development. Current models arguing for shared prediction processes in adults can potentially incorporate common inhibition mechanisms; however, our data imply that non-shared processes should also be considered. On the other hand, unlike adults, our children's data supports the capacity-constraint account in language processing, thus suggesting a discontinuity of cognitive functioning in language development.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I would like to thank the following people, without whom I would not have been able to complete this thesis.

First, and foremost, I would like to thank my supervisors, Silvia Gennari and Lisa Henderson. Their support, guidance and insights helped me all the time, in determining my research directions, devoting research efforts, and writing this thesis. I could not have imagined having better supervisors for my PhD study.

I would also like to thank the members of my research committee: Beth Jefferies, Emma Hayiou-Thomas and Jelena Mirkovic, for their encouragement, invaluable comments and hard questions.

I wanted to also thank all the individuals who participated in my studies, particularly all the children and their families, my research would not have been possible without their help.

Finally, I would like to thank my parents, for their continuous love and support throughout my life, including allowing me the freedom to study what I choose and supporting my every academic decision. Their support has undoubtedly helped me to finish this thesis.

Author's Declaration

I declare that this thesis is a presentation of original work and has not previously been submitted for a degree at this, or any other, University. Parts of the research described in this thesis have been presented at conferences and a paper based on Chapter 2 has been submitted to *Cognition*.

Conference Posters:

Wu, S., Gennari, S. P., & Henderson, L. (2017). *Individual difference predicts performance in language production and comprehension.* Poster presentation at the AMLaP Conference, Lancaster, UK.

Wu, S., Gennari, S. P., & Henderson, L. (2017). *Understanding semantic competition in sentence comprehension and production.* Poster presentation at the EPS Conference, Reading, UK.

Paper submitted:

Wu, S., Gennari, S. P., & Henderson, L. (2019). *Common and distinct inhibition skills underpinning sentence production and comprehension.* Submitted to *Cognition*

Chapter 1

Literature Review

1.1 Introduction

Understanding the cognitive mechanisms underpinning language comprehension and production has been a central goal of psycholinguistics. To date, most researchers have addressed this issue by investigating comprehension and production separately or at only word-level. Little is known about the relationship between the two regarding phrase or full sentence composition. Also, until recently, there has been little direct testing of how the findings in adult processing extend to younger and less experienced populations. Answers to these questions should improve our understanding of language more generally, and also form an important part of any language models. Thus, the purpose of this thesis is to investigate the cognitive basis of sentence/phrase comprehension and production across a wide age range (i.e. from childhood to adulthood).

The first aim of the thesis is to examine whether comprehension and production in adults recruit common or distinct cognitive processes. Secondly, because children and adolescents have underdeveloped cognitive skills (e.g. working memory, inhibition skills) as compared to adults (Bedard et al., 2002; Crone, Wendelken, Donohue, van Leijenhorst, & Bunge, 2006), this thesis attempts to investigate whether age-related improvements in language performance is mediated by developmental increases in cognitive abilities, or whether different cognitive abilities are important at different phases of development. This is because sentence/phrase processing may rely on certain cognitive abilities in childhood but not in adulthood once people have mastery over these structures. For example, it has been reported that younger and less skilled readers' difficulty in comprehension may be attributed to their limitations in bottom-up skills (e.g. decoding) to a greater extend as compared to older and more skilled readers (Catts, Hogan, & Adlof, 2005). Thus, the findings from this work should have implications for extending current psycholinguistic models to developmental changes, and also for education settings in developing trainings targeting cognitive abilities to improve children's language skills.

1.2 Comprehension and production models

Traditionally in the psycholinguistic literature, comprehension and production have been studied separately, and as a result, different models have been proposed to explain the underlying mechanisms in each task. This section will start with summarizing dominant comprehension and production models regarding full sentence composition, then describe more recent psycholinguistic models which have incorporated commonalities between comprehension and production.

1.2.1 Comprehension models

The majority of comprehension models assume some degree of interactivity in online comprehension, where comprehenders access and integrate various types of information to structure the input, including both linguistic (e.g. phonology, syntax and semantic) and non-linguistic (e.g. conceptual knowledge) information. However, there is little agreement on the time course with which different types of information being assessed. Two competing and largely incompatible models dominate the literature.

The first class of models assume that the analysis of the input proceeds serially, in two stages (Ferreira & Clifton, 1986; Frazier, 1987; Rayner, Carlson, & Frazier, 1983). According to this account, an initial interpretation of a sentence is guided by syntactic information only, and once an interpretation is chosen, other information is used to evaluate its plausibility at the second stage, revising the initial interpretation if necessary (e.g. *the garden-path model* developed by Frazier, 1987). In support of this model, Ferreira and Clifton (1986) monitored participants' eye-movement while reading complex phrases as in examples (1)-(4). The sentences in example (3) and (4) are unambiguous as compared to (1) and (2) due to the presence of syntactic marker ("that was", which encourages an object-extracted interpretation), whereas (1) and (3) are less ambiguous than (2) and (4) based on the semantic plausibility of the head noun (the inanimate noun "evidence" is unlikely to act as the agent of the event, thus it should be the patient. This encourages an object-extracted interpretation). It was found that participants read sentences like (3) and (4) significantly faster, indicating that they were sensitive to the disambiguating syntactic information to guide initial analysis. However, the animacy configuration of the head noun did not influence the first-pass reading time, suggesting readers did not use semantic information to form their initial syntactic analysis.

(1) The evidence examined by the lawyer turned out to be unreliable
(2) The defendant examined by the lawyer turned out to be unreliable
(3) The evidence that was examined by the lawyer turned out to be unreliable
(4) The defendant that was examined by the lawyer turned out to be unreliable

On the other hand, constraint-satisfaction models propose a more incremental analysis during comprehension, in which all information (including both syntactic and non-syntactic information) is activated in parallel to form multiple interpretations (MacDonald, Pearlmutter, & Seidenberg, 1994; John C Trueswell, Tanenhaus, & Garnsey, 1994). Based on this account, comprehenders may entertain competing interpretations simultaneously at the point of the ambiguity, with the level of activation determined by probabilistic constraints derived from language experience. If this account is correct, the processing of example (1) should be easier than the processing of example (2), because the initial interpretation should also be guided by the semantic information of the head noun which constrains the thematic role it can take in the event. That is what has been found in Trueswell, Tanenhaus and Garnsey (1994)'s study after replicating Ferreira and Clifton (1986)'s experiments with revised stimuli (this latter study has been criticized for using inappropriate stimuli, e.g. many of the unambiguous sentences could have more than one syntactic interpretation).

Taken together, the contrast between the two-stage and constraint-satisfaction models usually focus on the extent to which the analysis of syntactic structure and lexical/contextual information are separable. Also, the two accounts differ in the assumptions made about the sources of comprehension difficulties. The two-stage models suggest processing difficulties only arise at the reanalysis stage, when the initial interpretation is incompatible with later information. The constraint-satisfaction models, on the other hand, would suggest competition is one source of processing difficulty, which is particularly strong when alternative interpretations have near equal activation.

1.2.2 Production models

Language production models (Bock & Levelt, 1994) suggest that there are different levels of processing to produce a sentence. To start with, speakers generate a conceptual representation of the message they wish to convey, followed by linguistic encoding, which is subdivided into two processes: a structural assignment, which establishes the grammatical relationships between concepts and determines an appropriate word order; and a phonological encoding which guides articulation. However, the model does not necessarily imply that planning of the entire sentence is completed before articulation. Rather, it is generally accepted that production functions incrementally as speech unfolds, meaning that the speech is initiated once minimal chunks of sentence is planned (see Brown-Schmidt & Konopka, 2008, for evidence that the size of planning unit can be as small as a single word). Thus, the planning of partial speech, articulation and subsequent planning are interleaved.

Many studies have investigated the degree of planning before articulation. Some studies have suggested that elements of speech are planned in a word-by-word or concept-by-concept fashion as a consequence of an accessibility-based approach, with the most accessible element encoded first and produced early in the utterance (termed 'linear incrementality'; e.g. Gleitman, January, Nappa, & Trueswell, 2007; Levelt, 1982). Alternatively, others have argued that production is guided by a more complex higher-level *plan* concerning the structural relationship between concepts, in which the planning of the earlier portion of speech also encode its relational information with the later portion of speech (termed 'hierarchical incrementality'; e.g. Griffin & Bock, 2000).

Evidence supporting the linear incrementality approach mainly comes from studies which have found visual-attentional factors (such as perceptual salience) influence word order choice in production (Gleitman et al., 2007; Myachykov, Thompson, Scheepers, & Garrod, 2011; Tomlin, 1997; Tanaka, Branigan, & Pickering, 2010). For example, in an eye-tracking study, participants were asked to describe events presented in pictures (Gleitman et al., 2007). The authors manipulated participants' attention towards a particular character, by asking them to fixate at a cued location before viewing each picture. It was found that participants were more likely to describe this character first as the sentential subject, even when this subject assignment increased the likelihood of using the less favoured verb (e.g., "the dog *flees* from the man" rather than "the man *chases* the dog") or sentence structure (e.g., "the boy is being kicked by the girl" rather than "the girl is kicking the boy", i.e., the less frequent passive structure is preferred over the active structure). This finding suggests that an early endogenous shift in attention influences the order of planning operation. Namely, the character fixated first is planned first and assigned to the subject position without extensive consideration of its relationship with other characters. The rest of the speech is then built to accommodate the early produced subject, by shifting attention to the second character or concept and adding it to the utterance.

On the other hand, the hierarchical incrementality approach suggests that the relational structure of the event to be described initiates, rather than follows the planning of any increment. For example, Griffin and Bock (2000) instructed participants to describe pictures depicting simple agent-patient events (e.g. *a woman is shooting a man*, or *a man is being shot by a woman*). It was found that the initial self-generated fixation to one character was not attributable to description order or making this character the subject of upcoming sentence. It also did not predict which sentence structure was produced (active or passive). Rather, fixation of the two characters did not differ or diverge rapidly during the first 400 ms after picture onset. Thus, this result suggests a rapid encoding of *who did what to whom* within the initial gaze shifting between characters, allowing participants to select the sentence subject based on the comprehended event rather than its perceptual salience. This approach allows for a top-down control of sentence structure, with the remaining concepts becoming easier and being guided by an early encoded conceptual framework.

More recently, Konopka and Meyer (2014) have suggested a more flexible conception of incrementality. In two eye-tracking experiments, participants received different types of priming before completing a picture description task. In a lexical priming condition, participants were presented with words which are semantically or associatively related to a character in the picture (e.g. *pony/milk* before seeing a picture of a horse kicking a cow), and this type of priming should facilitate the encoding of target character. On the other hand, in a structure priming condition, participants heard lexically unrelated descriptions with either active or passive structures and this type of priming should facilitate generation of sentence structure. It was found that when primed lexically, speakers are more likely to engage in linear incremental strategy and prioritize naming of a single character during initial planning; whereas structural priming encouraged the hierarchical incremental strategy and increased the likelihood of encoding the relational information between characters. Thus, this result implies a flexible approach, where the likelihood of adopting a particular planning strategy varies between situations, and is guided by the availability of context-specific information. The planning strategy which exploits easy processes in early stage is preferred, in order to minimize the cognitive load as the sentence unfolds.

1.2.3 Models concerning the connection between comprehension and production

To summarize, traditional psycholinguistic models have viewed sentence comprehension and production as recruiting separate processes. Comprehension is accomplished through the rapid weighing of different constraints (either serially or in parallel) to generate possible interpretations of a sentence. Sentence production is guided by the development of an accessibility or hierarchical-based utterance plan which is presumably unambiguous. Indeed, in certain aspects, there is little doubt that comprehension and production must engage some distinct processes. Production is generally more difficult than comprehension, and production skills lag behind comprehension during development and in second-language acquisition (Bates, Bretherton, & Snyder, 1991; Fenson et al., 1994). Also, processes such as retrieval of word meaning in comprehension and motor sequence planning in production are often suggested to be task-specific.

However, despite these asymmetries, it is unlikely that comprehension and production operate independently, there are certainly commonalities between the two tasks. Lexical and semantic knowledge must be shared across comprehension and production (e.g. Levelt, Roelofs, & Meyer, 1999: lemmas are shared across modalities). Also, it was found that syntactic priming (i.e. repetition of syntactic structures from recent experience) occurs across modalities, such that comprehending a sentence with a particular structure increased the likelihood of using the same structure during the subsequent production task and vice versa, and the size of this effect is comparable to priming within a modality, e.g. *from production to production* (Bock, Dell, Chang, & Onishi, 2007; Branigan, Pickering, & Cleland, 2000; Branigan, Pickering, & McLean, 2005). This then suggests syntactic information is also shared across comprehension and production.

Based on these observations, recent psycholinguistic models begun to incorporate the connections between the two, specifically focusing on a shared prediction processes within an encompassing language system. Pickering and Garrod (2013), for example, propose a model in which prediction in comprehension and production is primarily implemented by the production system, which acts as an internal simulator generating forward models of upcoming linguistic information (Pickering & Garrod, 2007, 2013). This production simulator generates predictions of upcoming utterance units (the speaker's intended action or the predicted action of others), which are then checked against the actually produced or comprehended linguistic unit. The production system is thus a component part of the comprehension system. The evidence in support of this model includes behavioural results indicating that comprehenders predict upcoming stimulus words/phrases or that speakers repeat previously heard structures in production and align with interlocutors (Pickering & Garrod, 2013). It also includes imaging evidence showing vocal motor-related activity in speech sound processing (D'Ausilio et al., 2009; Schomers, Kirilina, Weigand, Bajbouj, & Pulvermüller, 2015), or very rapid responses to one's speech (Tian & Poeppel, 2013). Much of this evidence can be explained by associations at different levels of linguistic representations (e.g., speech sound-motor articulation associations) or by inferences from common ground and context without intervention of production mechanisms.

Other models also argue that production amounts to prediction but instead propose a dual-path recurrent network architecture with interconnected sequencing and meaning pathways (Bock, Dell, Chang, & Onishi, 2007; Chang, Dell, & Bock, 2006; Dell & Chang, 2014, see figure 1). The network learns from linguistic experience and is able to explain experimental results from a variety of acquisition, production and comprehension studies (Dell & Chang, 2014, Bock et al, 2007). Its learning algorithm suggests that prediction sometimes leads to prediction error, and which is then used to improve the connection weights in the network and minimize future error (Rumelhart & McClelland, 1986). Prediction error thus drives implicit learning during language acquisition, and the acquired representations drive comprehension *and* production predictions in mature language processing.

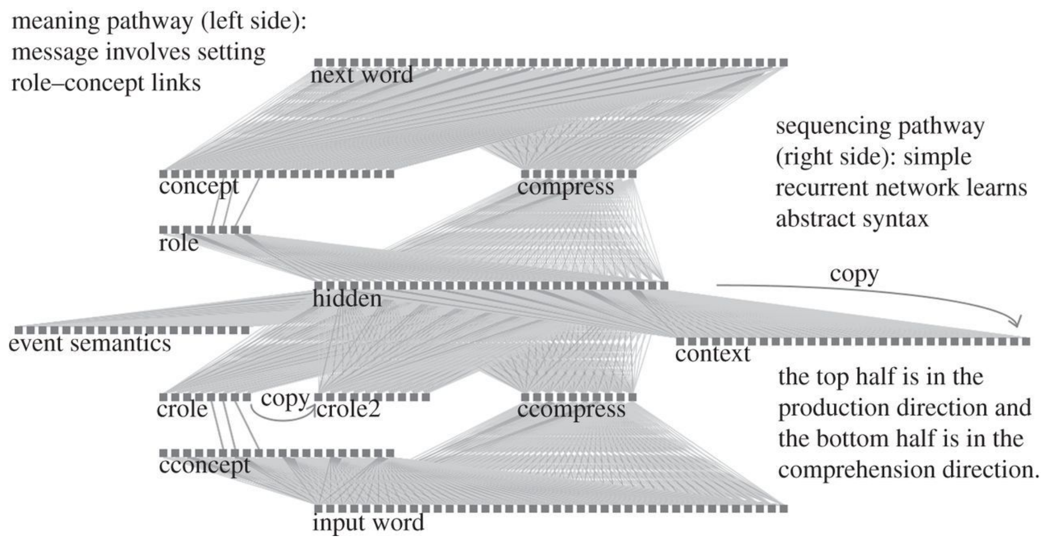


Figure 1 The dual-path model (figure taken from Dell & Chang, 2014)

The above models intended to introduce unified frameworks by focusing on a specific aspects of language comprehension and production (e.g. prediction), which is often applied to prediction in highly probably cases that are easy to processes, e.g., when words or sentences are primed. But it can hardly be said they have fully elucidated the relationship between these two tasks. What remains to be investigated is the way comprehension and production are related in when processing is difficult, e.g., in cases of sentential ambiguity such as those discussed in (1)-(4). For example, whether competition resolution mechanisms or re-analysis processes are shared across sentence production and comprehension. To address this issue, one approach is to make direct comparisons of comprehension and production tasks targeting sentence-level processes. Sentence level investigations are important as some similarities or differences between comprehension and production are likely to be evident only when factors such as syntactic planning/interpretation come into play.

Given that very few studies have directly contrasted comprehension and production performance at sentence level, the following sections will review comprehension and production findings separately, focusing on complex phrase processing. Complex phrases such as relative clauses merit detailed attention, mainly because it is one of the most common complex structures in English, and it has received much attention in many areas of psycholinguistic literature, including language processing and acquisition, and investigations with language or reading impairments such as SLI, dyslexia (Diessel & Tomasello, 2005; Friedmann, Yachini, & Szterman, 2015). Also, it has been suggested that both adults and children's comprehension and production of relative clauses require coordination of multiple cognitive processes (Gordon & Lowder, 2012; MacDonald, 2013), but what remains unclear is whether the two tasks engage common or distinct processes and how these processes might differ across development.

1.3 Relative clause comprehension and production in adults

1.3.1 Comprehension in adults

Relative clauses are noun modifiers that include a verb (e.g. the man being punched by the woman). In the comprehension literature, many studies have focused on the contrast between subject relative clause (SRC) and object relative clause (ORC). In SRC, as in example 5, the head noun "the girl" is modified by the bracketed relative clause and serves as the subject of the verb "hug"; whereas in ORC, the head noun serves as the object of the relative clause verb (as in example 6).

(5) Subject relative clause (SRC): The girl (who hugged the woman) was dripping wet.

(6) Object relative clause (ORC): The girl (who the woman hugged) was dripping wet.

It has been consistently reported that ORCs are associated with greater processing cost than SRCs, and this asymmetry has been found consistently with different methodologies including self-paced reading (King & Just, 1991), eye tracking (Traxler, Morris, & Seely, 2002), and functional magnetic resonance imaging (Marcel Adam Just, Carpenter, Keller, Eddy, & Thulborn, 1996). Possible explanations for this processing asymmetry are divided between a syntactic complexity/memory-based account, and an experience-based account.

Syntactic complexity/memory-based account

Briefly, this account suggests that ORCs are inherently harder than SRCs, because they impose greater memory demands which are largely driven by the complexity of syntactic structure. For example, it has been argued that ORCs are difficult because they involve longer distance between dependent elements (Grodner & Gibson, 2005). As in example (5) and (6), comprehenders need to hold the head noun “the girl” in working memory (WM) until they encounter the verb “hug” to interpret the thematic role (i.e. agent/patient role) of the head noun. The duration for which the head noun must be maintained in WM is much longer for ORCs, thus requires greater WM load during processing. This supports the general assumption about processing capacity limitation, as illustrated in Alan Baddeley’s working memory (WM) model (Baddeley & Hitch, 1974). According to this model, comprehenders have a limited supply of neural resources to support cognitive operations in sentence comprehension and information is lost from WM when exceeding available resources.

Other explanations suggest that the meanings of SRCs are processed in a more straightforward way as compared to ORCs, because they follow English canonical subject-verb-object order, whereas ORCs start with describing the object of the event. Thus, when comprehending ORCs, people often misinterpret the head noun as the subject of RC at an early stage, and need to engage reanalysis at a later stage; whereas for SRCs, the initial interpretation is already correct (Frazier & Clifton, 1989; Traxler et al., 2002). Another possibility is that comprehenders tend to follow the agent’s point of view as the sentence unfolds. Comprehension of ORCs may involve switching perspectives from the initial head noun (the object of the action) to the embedded noun (the subject of the action), whereas comprehension of SRCs entails no shift (MacWhinney, 1977; MacWhinney & Pleh, 1988).

Taken together, the above accounts emphasize inherent processing differences between different relative clause types, which are largely driven by higher memory demands and higher syntactic complexity of certain structures. Specifically, some explanations concern the WM burden of maintaining unintegrated noun phrases in memory, whereas others suggest that the greater processing demand in ORC is imposed by engaging additional analysis or process (reanalysis or perspective shift) to resolve comprehension ambiguity.

Experience-based account

Unlike the Syntactic complexity/memory-based approach, experience-based accounts do not claim that ORCs are inherently more complex than SRC. It argues that comprehension of complex sentences depends on past linguistic experience of similar structures. Corpus analyses report that SRCs appear more frequently in speech and written materials than ORCs. Thus, the greater difficulty in interpreting ORC may arise from comprehenders' comparative lack of experience with this structure, which leads to misinterpretation (Reali & Christiansen, 2007).

Moreover, this theory suggests that by given extensive exposure to similar sentence structures, even young children would be able to comprehend syntactically complex relative clause. In support of this idea, Wells, Christiansen, Race, Acheson, and MacDonald (2009) found that participants who received extensive exposure to ORCs over several training sessions read this structure significantly faster than another group of participants who received the same amount of training on other types of complex sentence structures.

Lexical-syntactic features affecting comprehension difficulty

Whereas the above accounts have focused on the structural contrast between SRC and ORC, other studies have identified a number of lexical-syntactic features that greatly affect the processing difficulty of ORCs, such as noun types (pronouns, proper names, descriptive nouns, etc), noun-verb pairing (King & Just, 1991; Gordon, Hendrick, & Johnson, 2001; Ferreira & Dell, 2000).

One widely investigated lexico-syntactic feature is the animacy configuration of the nouns. A number of studies have reported that ORCs with inanimate head nouns (as in example 8) are easier to comprehend as compared to animate head nouns (as in example 7; Gennari, Mirkovic, & MacDonald, 2012; Gennari & MacDonald, 2008; Mak, Vonk, & Schriefers, 2002) .

(7) The girl that the woman hugged was dripping wet

(8) The toy that the woman hugged was dripping wet

There are several explanations for this animacy effect. For example, it has been suggested that the reanalysis of inanimate head is easier than the reanalysis of the animate head. This is because inanimate nouns (e.g. the toy) are unlikely to act upon animate nouns (e.g. the girl); thus it is easier to consider inanimate nouns as the patient of the event at the reanalysis stage (Traxler et al., 2002; Traxler, Williams, Blozis, & Morris, 2005)

Another explanation concerns similarity-based competition in comprehension. Humphreys, Mirković, and Gennari (2016) asked participants to rate conceptual similarity of the two nouns focusing on both physical and semantic aspects (e.g. similarity in function). It was found that this rating positively correlated with the response time taken to comprehend animate-head phrases. Participants spent longer to understand phrases describing similar nouns (two animates e.g. *the woman*, *the girl*) as compared to less similar nouns (one animate and one inanimate e.g. *the toy*, *the girl*). This is because similar nouns are more likely to compete for the allocation of syntactic roles (i.e. agent, patient roles), and causes more interference in processing. And this relationship remains significant even in cases where all nouns involved are animate entities, e.g. *the woman, the girl*, vs. *the dog, the girl*. Thus, this finding provides evidence for the existence of similarity-based competition in comprehension, in which the degree of competition is modulated by the specific semantic features of noun phrases, rather than only categorical animacy.

1.3.2 Production in adults

An aspect often examined in production studies is the producer's structure choice, because there are many different ways to convey a message, varying in word order, sentence structure and lexical choice. For example, when asking which girl is wearing red in a given picture, one may describe it as *the girl being hugged by the women* or *the girl who the woman is hugging,* and both are plausible answers to this question. However, what most researchers are interested in is how producers converge on a single form over other alternatives, which provides insight into the underlying mechanisms of language production. Many production studies used a paradigm of sentence completion or picture description to investigate elicited production responses. In a sentence completion task, participants are usually presented with several words describing the entities and the action of a to-be-described event (e.g. movie, director, pleased) and they need to produce a meaningful referential phrase with the words provided (e.g. Gennari & MacDonald, 2009). In a picture description task, participants are often asked to describe a specific character or event in a presented picture (e.g. Gennari et al., 2012).

Factors influencing production choices

One area of interest is to investigate the factors/bias that shape utterance choice in phrase or sentence production. MacDonald (2013) identified several production biases speakers often adopt (consciously or unconsciously) when developing utterance plans. The first factor is called *Easy First*. Because utterance plan is maintained in WM before articulation, to avoid the memory burden of maintaining a large plan and to maximize fluency, people tend to utter more easily planed (or more assessable) word or concept first and place it at more prominent syntactic position (e.g. as the sentence subject), in order to leave more planning time for less accessible elements. The more easily planed elements are characterized as having higher frequency, shorter word or phrase length, higher conceptual salience (e.g. animate entities) or previously mentioned in the discourse (Arnold, Losongco, Wasow, & Ginstrom, 2000; Bock, 1982; Levelt, 1982). By uttering some elements at the beginning, it often constrains the structure for the rest of the speech; therefore, syntactic structure of an utterance may not be a deliberate decision but as a consequence of accessibility-based approach of utterance plans. For example, it was found that in both English and Spanish, speakers tend to locate the more assessable entity at the initial position of a sentence when asked to describe a transitive event with two entities. This resulted in more active descriptions being produced when the agent was made more accessible through priming, e.g. *the truck hit the boy*; and more passives being produced when the patient was made more accessible, e.g. *the boy was hit by the truck* (Bock, 1982, 1987; Prat-Sala & Branigan, 2000).

Secondly, speakers tend to choose easier plan (more frequent, more practiced or recently used ones) over more complex plans (a factor named *plan reuse*). This is evident in studies which found that the syntactic structure of the priming sentence significantly affected the structure of produced sentence following it (Bock, 1986; Branigan, Pickering, McLean, & Stewart, 2006; Pickering & Branigan, 1998). It is suggested that this is not only a temporal activation of recently used plan but entails retrieval of favoured structures from long-term linguistic knowledge (Bock et al., 2007).

The final factor is named *Reduce interference*: speakers tend to select the syntactic structure which helps to reduce interference in planning. Since the utterance plan is thought to be maintained in WM before articulation, elements of the plan can interfere with each other especially when they are semantically (e.g. couch and sofa) or phonologically (e.g. boy and ball) similar. MacDonald suggested that this interference can be reduced by strategically choosing certain sentence structure (i.e. passive object relative clause) to place the two interfering elements further apart or placing one of the elements in grammatically less prominent position.

For example, in Humphreys et al., (2016)'s production task, participants were instructed to complete a sentence starting with the patient of the event (either animate or inanimate entity), and in such a way that they were forced to produce an ORC with either active or passive structure (see examples below).

(9) Animate Head, Passive: The man that is being punched by the woman

(10) Animate Head, Active: The man that the woman is punching

(11) Inanimate Head, Passive: The punch bag that is being punched by the woman

(12) Inanimate Head, Active: The punch bag that the woman is punching

In the active structure (as in the examples 10 and 12), the head noun and embedded noun are positioned quite close to each other, and both have prominent grammatical roles. Whereas in the case of passive structure (as in examples 9 and 11), the head noun and embedded noun are separated by the verb.

It was found that participants tend to use more passives in description when both nouns were animate entities (as in example 9), whereas they were equally likely to produce passive or active structures when describing an inanimate entity being acted on by an animate entity. This is because when the head noun is inanimate, the topicalization of head noun promotes a passive structure, whereas the tendency to maintain animate agent in prominent grammatical position encourages an active structure. These conflicting forces result in a combination of active and passive structures. The reason that passive structures are often produced to highlight the head noun is because by using passives, the embedded noun was positioned further away from the head noun and is demoted to a by-phrase which can be eliminated entirely (e.g. the man that is being punched). On the other hand, in active structures (as in example 10 and 12), both the head and embedded nouns take prominent grammatical roles. When the head noun is animate entity, again the topicalization of head noun encourages a passive structure. However, there is tendency to make either noun as the grammatical subject of the sentence, as both nouns are equally prominent (they are both animate nouns). Thus there is no push toward an active structure, which results in a greater proportion of passive utterance when describing animate head noun (Branigan, Pickering, & Tanaka, 2008; Gennari & MacDonald, 2009).

Similar to comprehension literature, this animacy effect can also be explained by a similarity-based competition between agent and patient nouns (Gennari, Mirkovic, et al., 2012; Humphreys et al., 2016). It was found that as the semantic similarity between the nouns increase, the preference for producing passives and omit agents in passives also increase. This suggests speakers experience greater interference or competition during planning, and thus need to differentiate the two nouns by using passives to position them further apart or completely omit the patient (e.g. the man that is being punched).This competition arises from the confusion of assigning the thematic roles of the nouns, as for highly similar nouns, both can act as reasonable agents in a sentence (e.g. a man can punch a woman, and a woman can punch a man).

The above studies have shown that variation in speaker's utterance choices (word order variation, passive/active forms) has functional importance, and is moderated by different constraints. The word order variation may be shaped by speakers' tendency to place easily retrieved elements at the early position of utterance in order to maximize production incrementality. The choice of passive/active forms may not only reflect the speaker's attempt (consciously or unconsciously) to convey a particular message, but also due to a strategy to reduce production difficulty. Passive is selected over active to emphasize the conceptually more prominent entity, or it is chosen to reduce similarity-based competition between highly similar entities, suggesting multiple factors can contribute to the variation in utterance plan.

1.3.3 Summary of comprehension and production in adults

Despite the fact that the majority of studies have investigated relative clause comprehension and production separately, the results seem to indicate some commonalities between the two tasks. At first, difficulties in comprehending and planning these structures can be attributed to multiple influential factors, including memory-related (e.g. memory-based account in comprehension, accessibility-based production bias) and experience-related (e.g. experience-based account in comprehension, tendency to reuse prior plan in production) constraints in processing. Also, certain types of relative clauses (e.g. animate-head ORCs) are difficult to comprehend and also rare in production, owing to similarity-based competition, which occurs in both tasks when semantically similar representations share agent/patient associations and compete for syntactic role allocation. Thus, it seems that comprehension and production are sensitive to the same linguistic constraints (e.g. animacy), and also show similar competition-related effects. However, these findings cannot unambiguously indicate whether the source of semantic competition in each task results from common processes, or distinct processes which happen to elicit a parallel behavioural manifestation.

Moreover, any effective language models or theories need to account not only for cognitive processes, but also for development. Relative clauses take time to learn and are to some extent challenging, thus it is possible that young children who have less experience with these structures and also underdeveloped cognitive skills, might show systematically different behavioural patterns than adults. The following section will review studies examining children's comprehension and production of relative clauses.

1.4 Relative clause comprehension and production in children

1.4.1 Comprehension in children

Similar to adults, young children also find ORCs more problematic than SRCs. Even though both structures are equally frequent in child-directed speech (Diessel, 2004), children under age of six can comprehend SRCs, but not ORCs, which are interpreted correctly only 50% of the time and mastered later in development (Arnon, 2010; Corrêa, 1995; Diessel & Tomasello, 2000; Goodluck & Tavakolian, 1982; Kidd & Bavin, 2002). And just like adults, possible explanations fall into two broad categories: the syntactic complexity/memory-based account and the experience-based account.

Syntactic complexity/memory-based account

This account suggests that ORC is inherently more complex than SRC due to its syntactical features, and the ability to interpret these features mature late in normal development (Durrleman, Marinis, & Franck, 2016). For example, Friedmann and Novogrodsky (2004) reported that children with syntactic SLI and 4-year-old controls experienced more difficulty with interpreting ORCs (SLI children as old as 7;3 and even 11;2 don't understand ORCs), whereas their comprehension of SRCs and simple sentences are relatively good. Given their good performance on other sentence structures, their processing difficulty with ORCs cannot be attributed to a deficit in comprehending lexical items, but an impaired comprehension of the syntactic representation of ORCs. Specifically, the difficulty is associated with processing relative clauses with non-canonical structure, and which is more cognitively demanding as correct interpretation of ORCs involves the process of reanalysis and thus causes greater WM load. One option for children is to stick with the first analysis, which leads to comprehension failure. A second option is to search for the correct interpretation by maintaining the head noun in memory and looking for a new agent. It is clear that reanalysis in ORCs require previous information to be stored and retrieved in order to achieve a correct reanalysis. So, it is not surprising that only children with enough memory resources can interpret ORCs correctly.

In support of this idea, Arosio, Guasti and Stucchi (2011) adopted a self-paced listening paradigm and found that children with low memory capacity (as measured by digit-span) made more errors and slowed down in listening to ORCs, regardless of whether their answer to the comprehension question was correct or not. In contrast, comprehension of SRCs is always at ceiling for both children with high or low digit span. This suggests that children with limited memory resources typically activate an SRC interpretation initially, and their accuracy in ORC comprehension depends on whether they are able to abandon their first interpretation and engage in reanalysis. Moreover, the authors investigated whether the animacy configuration can modulate comprehenders' processing strategies. The results suggested two possibilities: when hearing ORCs with inanimate head, adults and children with higher memory resources may either not process an SRC analysis immediately (because inanimate nouns make bad agents), and wait for the next piece of information to make the decision; or they do engage, but their reanalysis is less costly and never fails as compared to children with limited memory resources. In contrast, both adults and children, regardless of their memory capacity, always engage in an SRC analysis when encountering ORCs with animate heads (because animate nouns are equally plausible to be agents or patients).

Weighall and Altmann (2011), however, reported slightly different results. They investigated 6 to 8-year-olds' comprehension of spoken relative clauses. It was found that children with high and low memory-span (as measured by a listening span task) performed similarly in accuracy, suggesting that higher WM capacity is not associated with more accurate comprehension of relative clauses. However, the authors did find that children with a higher memory span demonstrated positive effect of context. They made less error when provided with additional contextual information about the actions involved, which was not evident in low-span children. This suggests that WM capacity is not associated with processing difference between different types of relative clauses, instead it influences children's ability to integrate contextual information to support comprehension.

Experience-based account

On the other hand, the experience-based account suggests that comprehension of syntactically complex sentence depends on children's past experience with similar sentence structures. According to this account, the SRC/ORC asymmetry observed in many studies is because children were tested with ORCs they rarely hear or produce in daily life. In other words, ORCs used in previous studies fail to satisfy the distributional frequency of ORCs in child and child-directed speech. It was found that when children were tested on ORCs which they often hear and say (contains an inanimate patient and a pronominal agent, such as "Can you give me the sweater that he bought?"), this processing asymmetry disappeared (Brandt, Kidd, Lieven, & Tomasello, 2009). Furthermore, children with extra experience and feedback on processing relative clauses (including both SRC and ORC) improved their comprehension performance significantly compared with a group of children who were trained with processing other structures (Roth, 1984). Taken together, this suggests that young children are sensitive to the distributional frequencies of complex structures and make use of this statistical information in the acquisition process.

Lexical-syntactic features affecting children's comprehension performance

Another area of research focuses on whether children are sensitive to the same constraints on relative clause processing as adult comprehenders. Several studies have demonstrated that children are affected by at least two lexical constraints: animacy of noun phrases and the type of lexical NP. In both act-out and self-paced listening studies, children comprehended ORCs with inanimate patients better than those with animate patients (e.g. the baker watches the *mouse/ball of yarn* that the cats are chasing) (Arosio et al., 2011; Corrêa, 1995). Children also showed improved comprehension when tested on ORCs with first person nouns as compared to those with full lexical NPs (e.g. the nurse that *I/the girl* is drawing, Arnon, 2010).

Similar to adults, multiple pressures drive the preference for certain lexical combinations in children, consistent with either memory or experience-based explanations. For example, the preference for processing first person pronoun can be explained as it presents the perspective of the speaker and is phonologically shorter than full lexical NP, which makes it more accessible from long-term linguistic knowledge and imposes less WM load. Also, ORCs containing first person nouns appear more frequently in child-directed speech, thus children's improved comprehension can be traced to their greater linguistic experience with familiar ORC types. However, at this stage, it is difficult to determine the degree of influences from different sources of constraints, as it would be difficult to conclude whether a certain ORCs imposes greater level of memory constraints because they are infrequent, or the other way around.

1.4.2 Production in children

Very few studies have investigated relative clause production in children. In general, children showed a preference for certain types of relative clauses in a way that matched the difficulty that was found in comprehension research: SRCs are preferred over ORCs, and the production of SRCs is mastered earlier than ORCs (e.g. McDaniel, McKee, & Bernstein, 1998). Also, their very early production of ORCs (around 2 years old) owes to certain types of ORCs, such as describing an isolated head noun, e.g. *another picture I made*, or attached to the predicate nominal of a copular clause, e.g. *that is the sugar that goes in there* (Diessel & Tomasello, 2000). It was suggested that starting from these simple structures at early age, children gradually learn to produce more complex relative clauses (such as two propositions expressed by main and relative clauses), and children at early age are not capable of producing structures which are tested in many studies (e.g. the dog that jumps over the pig bumps into the lion). Thus, until children have completely acquired relative clauses, experiments which use unfamiliar discourse context and multiple propositions of ORCs are not ecologically valid in testing children's production competence.

Factors influences production choices in children

As young children are capable to produce certain types of ORCs, many studies have investigated whether children are like adults, also attend to multiple source of information in choosing between syntactic variants, and if they do, whether their production choice is driven by the same constraints identified in adults' literature.

Firstly, Friedmann et al., (2009) found that Hebrew-speaking children aged 3;7 to 5;0 are sensitive to lexical NP restriction and they tend to avoid producing ORCs where both entities are lexically restricted (i.e. descriptive nouns such as "the man"). Thus, they tend to produce ORCs including one non-lexically restricted noun phrase (the entity is bare wh-word and/or pronoun), such as "who that the man is feeding" or "The tiger that someone is feeding", instead of "the Tiger the man is feeding". The authors argued that this is because (1) non-lexically restricted noun often refer to default referents and are more accessible compared to lexically restricted noun; or (2) lexical-restricted and non-restricted nouns belong to different lexical categories and interfere less with each other during planning.

Furthermore, Kidd, Brandt, Lieven, & Tomasello (2007) instructed three to four year olds to repeat SRCs and ORCs with animate and inanimate heads, as well as ORCs with pronoun and full NPs. The logic for using a sentence repetition task is because this method has been widely used to assess production skill as well as grammatical knowledge of syntactically complex sentence. It has been argued that when the sentence length exceeds children's short-term memory span, repetition will require a reliance on long-term linguistic knowledge. Thus, children need to apply a semantic analysis of the sentences they are asked to repeat, and when repeating the sentences back, they are using the same production mechanism they use when producing regular speech. The results from this study suggested that ORCs were easier to repeat with inanimate head nouns and pronominal embedded subjects (e.g. this is *the dog* that *you* saw), and this lexical combination is common in child-directed speech, suggesting that children's early relative clause behaviour closely tracks their linguistic experience with relative clause types.

To further investigate how children develop sensitivity to certain linguistic feature (i.e. the animacy of the nouns), and whether this is acquired from the input they receive, a recent study has made a direct comparison between children's elicited production of ORCs and their actual input of this sentence types, and this study has analysed linguistic input from two different sources: child-directed speech and child-directed text (J L Montag & MacDonald, 2015). In a picture description task similar to the ones used in adults' literature (Gennari, Mirkovic, et al., 2012; Humphreys et al., 2016), it was found that older children (12 year olds) produced more passive ORCs and their production contains fewer pronouns as compared to younger children (8 year olds). More importantly, children's implicit production choices are related to their linguistic input from written materials, as a corpus analysis has revealed that child-directed text contains a higher proportion of passive ORCs and fewer pronouns as compared to spoken language. Also, individual difference in a measure of text exposure significantly predicts the passive ORCs usage in children's elicited production over and above chronological age. This result highlights the importance of linguistic experience in shaping early production behaviour. Specifically, an influence across modalities in which literacy can affect not only reading comprehension, but also spoken language skills. Furthermore, this finding is consistent with the *plan reuse* bias identified in MacDonald (2013)'s production model, in which speaker's production choices often reflect their long-term implicit learning of syntactic structures.

1.4.3 Comprehension-production asymmetry in children

It is generally agreed in theories of language acquisition that comprehension always precedes production (Clark & Berman, 1987). However, in relative clause literature, the opposite pattern (i.e., production precedes comprehension) has been observed (Diessel & Tomasello, 2000; Goodluck & Tavakolian, 1982). It has been reported that children start to produce relative clauses as early as 3-years-old and they are able to produce both SRCs and ORCs, but they do not comprehend them before the age of 5. This has been demonstrated across different languages, including English (Sheldon, 1974), Swedish (Håkansson & Hansson, 2000), Hebrew (Friedmann & Novogrodsky, 2004).

For example, it was suggested that young children possess little knowledge of the recursive properties of language and use inappropriate strategies to interpret complex sentences. In tasks where they were asked to act out sentences, children before the age of 5 interpreted relative clauses as conjoined sentences instead of noun modifiers: "the pig bumps into the horse (that jumps over the giraffe)" is misinterpreted as "the pig bumps into the horse and jumps over the giraffe" (Sheldon, 1974; Tavakolian, 1981). However, these studies have been criticized for violating the pragmatic assumptions of relative clauses, and thus underestimates children's performance. It is generally agreed that relative clauses are often produced to distinguish between different referents (e.g. the above example would imply several horses in the context and the bracketed phrase indicates which horse the sentence refers to). Many studies only provided one toy for children to act out. It was found that four-year-olds' performance improved significantly and achieved 92% correct when they were tested within a pragmatically appropriate context, such as providing more toys of the to-be restricted referent (Hamburger & Crain, 1982). This suggests that children have an emerging knowledge of relative clause constructions at early age. They showed good comprehension when tested with sentences in appropriate contexts, with no lag behind production. Thus, the puzzling result of superior production performance found in many studies could be due to problems in the assessment materials used.

1.4.4 Summary of comprehension and production in children

To summarise, it seems that just like adults, young children are sensitive to multiple factors which significantly constrain their comprehension performance and production choices, and possible explanations can be attributed to either memory-based or experienced-based accounts. However, at this stage, it is difficult to conclude the degree of influences from different sources of constraints: how age, experience-related factors, and other knowledge and skills (working memory, vocabulary) contribute to comprehension and production abilities, owing to the general shortage of children studies in the literature (especially production studies).

Moreover, unlike in adults, whose comprehension and production of relative clauses reveal parallel behavioural effects (e.g. similarity-based competition) and pointed to shared underlying mechanisms, an asymmetry between comprehension and production is often observed in language acquisition. This observation appears to challenge the theoretical argument of a unified language architecture for comprehension and production as proposed in many psycholinguistic models (e.g. dual-path model). Many researchers attempted to explain this asymmetry as reflecting methodological issues: different comprehension and production tasks imposed different task demands. However, given that there is scarcity of studies that have directly contrasted comprehension and production performance using the same stimuli and paradigm, further work is needed to more accurately investigate whether relative clause processing in children also shows parallel behavioural effects and developmental trajectories.

1.5 Research objectives

Taken together, in traditional psycholinguistic literature, comprehension and production have been studied separately, and different models have been proposed to account for the underlying mechanisms recruited in each task. However, recent research with relative clauses may point to some similarities between the two tasks. For example, it has been consistently reported that certain types of relative clauses (e.g. active ORCs with animate heads) are difficult to comprehend, and also rare in production (speakers prefer to use passive ORCs to describe animate entities). Also, difficulties in comprehending and planning these structures can be attributed to the same explanation concerning either memory-based (e.g. similarity-based competition) or experience-based constraints in processing.

However, the nature of the common processes recruited by comprehension and production remains to be established, as investigation of sentence or phrase production is scarce, and very few studies have directly contrasted the two tasks and across development. To address this issue, we adopted an individual differences approach to examine the comprehension and production of ORCs (referred as *complex phrases* in the following chapters) that are known to elicit semantic-syntactic competition in both tasks. This approach not only allows us to directly contrast processing difficult across the two task but also allows us to determine whether common individual differences underpin both production and comprehension. If the same individual skills, e.g., vocabulary or inhibition skills, predicts performance in both production and comprehension, it provides some grounds to suggest that the processes embodied by these individual measures, e.g., inhibition of context-irrelevant meanings, operate in both tasks. The goal of the present work is thus to examine the extent to which comprehension and production may recruit common and separate processes to resolve competition in production and comprehension, as revealed by parallels across tasks and the individual measures underpinning task performance.

1.5.1. Thesis Outline

The following chapters of this thesis describe a series of behavioural studies that were designed to investigate the nature of competition resolution processes in complex phrase comprehension and production.

Chapter 2 presents a behavioural experiment where a picture-based paradigm was used to investigate whether comprehension and production recruit the same mechanisms to resolve semantic-syntactic competition, and whether these point to the retrieval of shared linguistic knowledge, or other processes recruited beyond the shared knowledge base. The results show common cognitive processes operate in both comprehension and production: a measure on semantic inhibition predicts performance over above vocabulary knowledge; as well as distinctive processes only underpinning production, i.e. motor inhibition. Chapter 3 then use the same paradigm to examine the development of competition resolution mechanisms, by comparing performance between children and adolescents. The results suggest that unlike older participants, young children's comprehension and production performance is predicted by working memory measures. This then reflects that the relative importance of different cognitive skills underpinning language processes tend to change with development. Chapter 4 describes the results of an eye-tracking production study with adults, where the time-course of semantic competition was investigated. It was found that competition resolution in planning complex phrases manifests at verb positions and is particularly related to individuals' semantic inhibition skill. This then parallels previous comprehension findings, and points to shared competition resolution processes across tasks. Finally, chapter 5 presents a general discussion of the thesis and future directions for research.

Chapter 2

Common and distinct inhibition skills underpinning sentence production and comprehension

2.1 Introduction

This chapter examines the extent to which complex phrase comprehension and production share common or distinct processes to resolve semantic competition. In psycholinguistic literature, production and comprehension of spoken language have been traditionally studied as distinct processes entailing disparate cognitive architectures. Sentence comprehension, for example, may involve the incremental mapping of the input into probabilistic alternative meanings (Levy, 2008; MacDonald et al., 1994), whereas sentence production might involve mapping meaning into articulation across several encoding stages (Bock, Loebell, & Morey, 1992; Griffin & Bock, 2000). Although there is little doubt that the lexical and conceptual knowledge feeding into these mapping processes must be shared across tasks, other aspects of processing such as self-initiated word retrieval and motor sequence planning appear task-specific, rather than shared. What cognitive processes then are common to language production and comprehension? This issue is critical to elucidate the architecture of the language system and the nature of the cognitive mechanisms operating in each task.

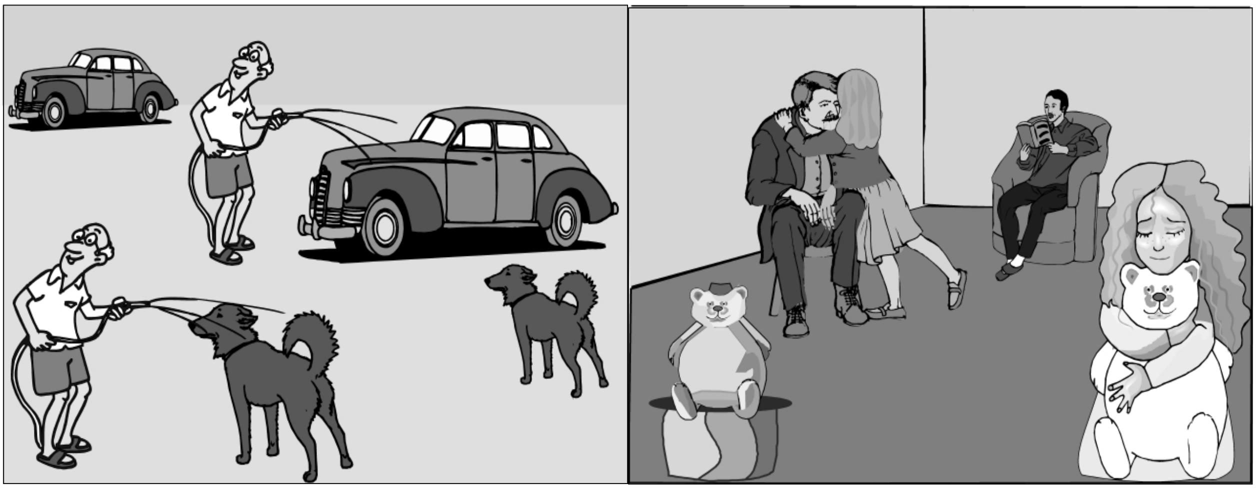
One recent approach to this issue has been arguing for a common language architecture on the basis of existing similarities across separate production or comprehension studies. Unified models of comprehension, production and language acquisition have begun to be developed, and most of which focused on shared prediction processes (e.g. Pickering & Garrod, 2007, 2013; Dell & Chang, 2014). Prediction in these models is often understood as the pre-activation of highly probably linguistic elements, such as those resulting from word and structure priming. However, most of language is not made of highly predictable elements, and in many cases, speakers and listeners experience difficulty. The focus on the present research is on structures that have been argued to be difficult due to competition between alternative interpretations or plans, and thus, not easily predictable. This research thus has the potential to shed some light on current prediction-based models of production and comprehension.

Other approaches to the relationship between production and comprehension have instead focused on documenting broad relationships between the two (MacDonald, 2013) or similar processes and common brain regions involved in both sentence production and comprehension (Humphreys & Gennari, 2014; Humphreys et al., 2016; Menenti, Gierhan, Segaert, & Hagoort, 2011). Humphreys et al (2016), for example, examined similarity-based competition in planning and comprehending complex referential expressions such as *the man that the girl is hugging* in the context of a visual scenes, see also (Gennari, Mirkovic, et al., 2012). They varied the degree of conceptual similarity between the agent and the patient nouns (e.g., *man/girl* vs *teddy-bear/girl*), where high-similarity leads to more conceptual interference or planning competition in syntactic role assignment (i.e. deciding the agent/patient role for each noun). They found that as similarity increased, comprehension and production difficulty increased, as indexed by comprehension times, production choices and eye-movements. These results suggested parallel semantic competition mechanisms in both sentence production and comprehension.

However, the nature of common production and comprehension mechanisms remains controversial, in part due to lack of compelling evidence. Common connectionist architectures for language production and comprehension have not been sufficiently developed to account for cases in which production and comprehension are both cognitively demanding due to competing alternative semantic-syntactic predictions such as those involved in structurally complex phrases (Gennari & MacDonald, 2009; Gibson, 1998; Gordon, Hendrick, & Johnson, 2001b; Staub, 2010; Traxler et al., 2002), despite some progress in comprehension (Fitz, Chang, & Christiansen, 2011). On the other hand, behavioural evidence for example, can point to similar processes in production and comprehension such as prediction or similarity-based competition, but cannot unambiguously indicate whether such processes result from shared lexical knowledge or whether they recruit other mechanisms outside this shared knowledge such as executive control processes. For example, individuals may struggle to process or produce the above complex referential phrases efficiently for multiple different reasons. First, their lexical representations of the nouns (predicted by their lexical knowledge) are not sufficiently well represented. Retrieval of poor-quality lexical representations is less efficient and more vulnerable to semantic interference, resulting in greater difficulties in relying on these representations to interpret or plan the syntactic relationships of the noun concepts. Secondly, they lack the executive control processes to rapidly select the appropriate interpretation or lexical/syntactic production decision and inhibit the alternatives. Finally, it may reflect a failure to maintain and manipulate lexical items or conceptual representations in WM to suit the requirement of task goal, e.g. keep relevant information in WM and assessing this information to assign syntactic roles. The effect of WM is particularly evident in children than adults, generally because the role of storage has been deemphasized in adult literature given that adults are assumed to possess sufficient WM capacity to maintain relevant information (the role of WM will be elaborated in Chapter 3). Taken together, it remains unclear what sort of common mechanism, if any, operates in complex sentence comprehension and production.

2.1.1 The present study

The present work addresses this issue and specifically asks whether difficult production and comprehension may share executive mechanisms such as inhibition (Friedman & Miyake, 2004; Hsu & Novick, 2016), which cannot solely be explained by shared lexical knowledge. To this end, we examined the comprehension and production of complex phrases that are well-known to elicit processing difficulty as a function of noun animacy. Phrases such as *the **man** that the girl is hugging* in Figure 2.1 are well established to be more difficult to comprehend than phrases such as *the **teddy bear** that the girl is hugging* (Gennari & MacDonald, 2008; Mak et al., 2002; Traxler et al., 2002, 2005). These phrases have the same structure but differ in whether an animate or inanimate entity is the target of the description. Comprehension difficulty also increases when the two animate nouns in the phrase are semantically similar, as in *the **man** that the girl is hugging,* compared to less similar nouns as in *the **dog** that the man is spraying*, suggesting that similarity-based competition contributes to comprehension difficulty (Gordon et al., 2001b; Gordon, Hendrick, Johnson, & Lee, 2006; Humphreys et al., 2016). Comprehension difficulty in animate-target descriptions is thought to occur because as the phrase unfolds, comprehenders experience competition between alternative semantic features in working memory or between alternative semantic-syntactic roles of the animate nouns (e.g., who is acting on whom), all of which requires the inhibition of one alternative in favour of the other. Similar animate entities share semantic features and are often potentially reversible within the event so that they are likely to be equally good agents or subjects and compete for this role. In contrast, inanimate nouns are generally poor agents, and thus do not elicit such competition (Trueswell, Tanenhaus, & Garnsey, 1994).



HEARD	**Animate-target:** *The dog that the man is spraying* **Inanimate-target:** *The car that the man is spraying*	**Animate-target:** *The man that the girl is hugging* **Inanimate-target:** *The teddy bear that the girl is hugging*
SPOKEN	**Animate-target:** *The dog being sprayed (by the man)* **Inanimate-target:** *The car being sprayed (by the man)* **or** *The car the man is spraying*	**Animate-target:** *The man being hugged (by the girl)* **Inanimate-target:** *The teddy bear being hugged (by the girl)* **or** *The teddy bear the girl is hugging*

Figure 2. 1 Examples of picture stimuli used in the production and comprehension tasks. Below the images, the text illustrates the heard stimuli in the comprehension task and the most typical spoken answers provided in the production task.

Production data on these referential phrases is less abundant, but it is well documented that in several languages, speakers tend to avoid active structures with animate-targets and highly-similar nouns, as shown in Figure 2.1, and instead tend to produce passives such as *the man being hugged (by the girl)* (Gennari, Mirkovic, et al., 2012; Hsiao, Gao, & MacDonald, 2014; Hsiao & MacDonald, 2016; Perera & Srivastava, 2015). When actives are produced due to explicit task instruction, they have been shown to be more difficult than simpler structures such as *the girl that is hugging the man* (Scontras, Badecker, Shank, Lim, & Fedorenko, 2015). For unconstrained descriptions, it has been argued that speakers often opt for passive structures because they experience interference or competition during planning between highly-similar entities (e.g., man and girl), resulting in the inhibition of one of the nouns or concepts (e.g., girl). This inhibition is subsequently manifested in the noun's demotion to the end of the structure or its omission (by-phrase omission). The high similarity between animate nouns not only elicit competition at lexical retrieval of the first noun, but also in determining its syntactic/semantic roles because the nouns are equally good agents and subjects of the verb. Consistent with this view, an eye-tracking production study indicated that semantic similarity modulates fixations before speech and at verb planning, thus suggesting competition in lexical selection and in determining the subject/agent of the structure (Humphreys et al., 2016).

The role of inhibition skills in semantic competition

There is a widespread consensus that *inhibition* may not be a unified structure, it can be referred to a range of attentional control processes that can occur at a behavioural or cognitive level (e.g. Miyake et al., 2000; Fisk & Sharp, 2004; Nigg 2000). In this thesis, we are specifically defining inhibition as the ability to actively suppress or ignore previously activated semantic representation or information that is already in WM, in order to suit the requirement of current task goal (i.e. *resistance to proactive interference* in Friedman & Miyake, 2004), as opposed to other types of inhibition mechanisms such as inhibiting habitual responses or ignoring task-irrelevant distractors. As it appears clear that this form of inhibition is particularly important in complex phrase comprehension to resolve the conflict when a strongly preferred syntactic role interpretation become inconsistent with the upcoming input. Similarly, successful production requires one to actively suppress activation of potentially misleading lexical and/or structure planning in favour of a less ambiguous planning strategy.

Neuroimaging and neuropsychological studies have provided evidence that this form of inhibition relies on partially distinct neural mechanisms to those involved in domain-general control. Executive control over semantic information engages a strongly left-lateralized network, with the left inferior frontal gyrus (LIFG) being the most reliably activated region across participants and tasks, when processing high-ambiguity sentence or words is contrasted with automatic semantic retrieval processes (e.g. Rodd, Davis & Johnsrude, 2005; Acheson & Hagoort, 2013). The "conflict resolution account" suggested that this region does not support retrieval of stored lexical-semantic representations, but is associated with the cognitive control processes that operate on these representations, particularly in circumstances when conceptual, lexical, semantic or syntactic representations compete for a response and creates high demands for conflict resolution (Novick, Kan, Trueswell & Thompson-Schill, 2009; Novick, Trueswell & Thompson-Schill, 2005; Thompson-Schill, et al., 1997). Thus, this view predicts that individual's abilities to comprehend or produce complex phrases will be closely related to their performance on executive control tasks that load heavily on semantic inhibition/selection.

Many standardized inhibition tasks have used perceptual stimuli devoid of semantic meaning (e.g. GO/NO-GO task, STOP-IT), or word stimuli inducing conflicts between its semantic and perceptual features (e.g. stroop task, naming the ink colour of a colour word while ignoring its meaning). In the current study, we sought to assess more directly the ability to resolve the conflict between incompatible semantic representations. To this end we created semantic inhibition tasks targeting ambiguous words where participants are able to establish lexical-semantic representations during processing (see method section for more details), based on large literature on lexical ambiguity resolution, e.g., (Gottlob, Goldinger, Stone, & Van Orden, 1999; Kawamoto, 1993; Simpson, 1994), and previous studies using similar measures (Gernsbacher & Faust, 1991; Hala, Pexman, & Glenwright, 2007). Ambiguous word with multiple meanings differs from colour word in that the stimulus input maps onto distinct semantic representations, and the dominant meaning need to be suppressed in the selection of context-appropriate subordinate meaning. Thus, even the context primes the meaning of river, presentation of the word *'bank'* activates its financial institution meaning which is later inhibited (e.g. Simpson, 1994). We propose that the inhibition mechanism invoked in suppressing a dominant semantic representation could be the same mechanism invoked when a preferred syntactic interpretation/planning is inconsistent with the task goal and need to be inhibited, thus is closely related to the efficiency and success of complex phrase comprehension and production.

2.1.2 Study hypotheses

Based on this prior research, we adopted an individual difference approach to examine whether common executive skills underpin both phrase production and comprehension. We reasoned that if shared executive mechanisms underpin both production and comprehension when dealing with competition, an individual's executive abilities should explain his/her performance in both tasks. Specifically, if production and comprehension share competitive inhibition processes, an individual's ability to inhibit irrelevant information should explain production and comprehension difficulty. Sentence comprehension difficulty has already been shown to correlate with executive skills involving inhibition, although less is known about sentence production. In particular, poor executive skills correlate with more processing difficulty (Gernsbacher, 1993; Gernsbacher & Faust, 1991; Hsu & Novick, 2016; Novick, Hussey, Teubner-Rhodes, Harbison, & Bunting, 2014; Trude & Nozari, 2017; Vuong & Martin, 2011; Woodard, Pozzan, & Trueswell, 2016). However, an appropriate test of the role of inhibition skills on processing requires controlling for variables that are already known to influence language processing such as vocabulary knowledge

(Braze, Tabor, Shankweiler, & Mencl, 2007; Tunmer & Chapman, 2012; Van Dyke, Johns, & Kukona, 2014). Moreover, it is possible that vocabulary measures account for difficulty even in cases of competition, because vocabulary tests of crystallised knowledge provide an index of the quality of an individual's lexical knowledge, which is not an "all or nothing" factor (words are either known or unknown), but represents variations in lexical knowledge even for highly familiar words and which directly influences language performance: the quality of lexical representations entertained influence the ability to keep them distinct in memory, making them more or less susceptible to interference (Van Dyke et al., 2014, Perfetti & Hart, 2002). Because vocabulary measures likely reflect shared lexical knowledge operating in both production and comprehension, we specifically test the possibility that inhibition skills account for unique variance in performance over and above the influence of vocabulary measures. Therefore, if inhibition skills underpin both production and comprehension performance over and above any influence of shared vocabulary, individuals with poor inhibition should experience more difficulty in dealing with highly competitive phrases than those with better inhibition.

2.2 Experimental Methods

2.2.1 Participants

83 native English speakers (65 females, mean age=20.91, *SD*=2.77; 18 males, mean age=22, *SD*=4.13) from the University of York completed the two main experimental tasks and six cognitive tasks but for technical reasons, the production data from 12 of these participants was not usable. Thus, for the analysis of production data, there were 71 participants (56 females, mean age=20.89, *SD* =2.86; 15 males, mean age=21.80, *SD*=3.69). For similar reasons, data from the homonym and homograph tasks was not collected for two participants (see details of these tasks below).

2.2.2 Materials

For the main production and comprehension tasks, 84 grey-scale pictures were adopted from previous studies (Humphreys et al, 2016, Humphreys and Gennari, 2014), comprising 42 experimental pictures and 42 filler pictures, see Appendix A. Each experimental picture contained two events to be described with the same transitive verb (cf. Figure 2.1). The events contained either an animate or inanimate target character that was acted on by another character. Because relatively few verbs can occur with animate and inanimate nouns in the object or patient position, some verbs were repeated across pictures. Of a total of 25 distinct verbs, 13 appeared in more than one picture.

For the comprehension tasks, phrases referring to picture characters were used as stimuli. These phrases were in an active form (see Figure 2.1) and included nouns and verbs that participants in a previous production study have more often used to refer to the characters and action depicted (Humphreys et al., 2016). For both experimental and filler trials, each stimulus phrase was recorded by a female native British English speaker in a sound-proof booth using Cool Edit software. The sound files were normalized to 68 dB SPL to minimise intensity differences throughout the recording session.

***Pre-test of relative agent-role likelihood*.** To obtain measures of agent-role likelihood for the animate nouns in a phrase, we created two online questionnaires. Following previous studies (Ferretti, McRae, & Hatherell, 2001; McRae, Ferretti, & Amyote, 1997), for each stimulus phrase (e.g., *the man the girl is hugging*), we asked participants to indicate how likely it was for each noun to be the agent of the corresponding action (e.g., *how likely is it for a girl to hug a man?* Or *how likely is it for a man to hug a girl?*). Each question was assigned to a different stimulus list (Latin Square design) so that the same participant did not assess the same question with reversed characters. 21 and 19 participants completed lists 1 and 2 respectively and provided a rating using a 1-7 scale. From the average agent-role likelihood for each noun on a phrase, we computed a difference score by subtracting the agent-role likelihood of the patient entity in the phrase (*man*) from that of the agent (*girl*). Small differences in agent-likelihood indicate that the patient entity (*man*) is equally likely to be an agent as the agent entity (*girl*), i.e., the event is highly reversible, whereas large differences in likelihood indicate no or less reversibility. As expected, these *reversibility scores* (range= -0.84-5.31 mean= 1.85 and SD = 1.90) were significantly correlated with the similarity ratings previously collected for these same items ($r(42)=-.543$, $p<.001$) (Humphreys et al., 2016), indicating that as similarity increased, the difference score decreased (more reversible nouns). From these scores, we defined a high, medium and low-reversibility grouping for picture items by dividing the scores into thirds (high-reversibility mean difference score=-0.19, *SD*=0.38; medium-reversibility difference score: 1.57, SD=-0.92; low-reversibility mean difference score=4.17, *SD*=-0.77).

2.2.3 Design and Procedure

2.2.3.1 Phrase comprehension and production task

For each of the 42 picture items, two different characters—either an animate or an inanimate one—could be targeted to elicit a production or comprehension response, and thus could be accompanied by an animate- or inanimate-target description (animacy condition). In Figure 2.1, for example, the targets for the right-side picture were either the teddy bear or the man being hugged. Animate- or inanimate-targets were allocated to two different lists (Latin-square design) so that a picture was seen in a given task only once. In both tasks, participants received 21 trials targeting animate entities and 21 trials targeting inanimate entities plus 42 filler items. The filler pictures were similar to the experimental pictures in that at least two entities interacted, but filler trials differ for each task (see below). Participants completed the comprehension task before the production task but were exposed to different lists in each task. If they had heard the animate-target reference in comprehension for a given picture (e.g., the man being hugged), they would then be prompted to describe the inanimate target in the production task and vice versa. This arrangement allowed participants to get familiar with the complex phrase structures in the comprehension task but did not contain repetitions of the same targets across both production and comprehension. These tasks were conducted on E-Prime.

In the comprehension task, participants were instructed to indicate whether the description they heard over the headphones was an accurate description of the character highlighted with a red square in the picture. They pressed one of two buttons on a button box to indicate their response (i.e., *yes* or *no*). The experimental trials in a list required a *yes* response, but the filler trials elicited *no* responses, thus balancing the number of yes/no responses throughout the task. The order of trial presentation was random. In each trial, a picture first appeared on the screen for 3 seconds, a character in the picture was then highlighted with a red box and an auditory description was presented (e.g., *the man that the girl is hugging*). The picture and the red box stayed on the screen until a button response was made, otherwise, the trial ended after 10s. Filler trials in this task contained a variety of phrase structures (e.g. *the boy playing with a ball; the dog being washed by the woman*). Participants' reaction times (RTs) for each correct trial were computed by subtracting the audio length of the recorded sentence from the total duration of the response computed from the presentation of the red box. Comprehension accuracy was at ceiling for animate and inanimate conditions (98.6% and 98.8% respectively).

In the production task, participants saw the same 42 pictures as in the comprehension task but in a different animacy condition (i.e., a different character was highlighted). The additional filler items elicited a variety of structures (e.g. *the tree on the playground, the dog burying the bone in the sand*). The order of trial presentation ensured that a filler always occurred between experimental items, but experimental items followed a random order. Participants in this task were instructed to verbally describe the highlighted character and their responses were recorded using a microphone. Practice trials and instructions indicated to participants that they should give descriptions uniquely identifying the character and using the actions being performed, rather than location or shape characteristics (e.g., *the man on the left, the short girl*). The trial structure was similar to that of comprehension. The picture was shown for 3sec for visual inspection, and then a red box highlighted a character on the screen. Participants produced a verbal response and then pressed a key to move onto the next trial (the red box stayed on the screen until the end of the trial). After the main experimental tasks, participants completed several vocabulary and inhibition tests described in detail below.

2.2.3.2 Individual differences measures

Based on our hypotheses, our measures of individual skills mainly consisted of two main groups—knowledge-based measures (vocabulary and reading experience) and inhibition-related skills. Vocabulary measures included tests of crystallised knowledge measuring the breadth and depth of vocabulary: PPTV (requires lexical recognition) and WASI vocabulary subtest (assess depth of knowledge of each word meaning). Measures of inhibition skills included a measure of motor inhibition (the STOP-IT task) and two measures of semantic and/or phonological inhibition in word production and comprehension. The STOP-IT task has been widely used to examine response inhibition in adults (Verbruggen & Logan, 2008). The word inhibition tasks were developed for the present study based on a large literature on lexical ambiguity resolution, e.g., (Gottlob, Goldinger, Stone, & Van Orden, 1999; Kawamoto, 1993; Simpson, 1994), and previous studies using similar measures (Gernsbacher & Faust, 1991; Hala, Pexman, & Glenwright, 2007). See below for details.

Vocabulary Measures. We used the Vocabulary subtest from the Wechsler Abbreviated Scale of Intelligence 2nd Edition (Wechsler & Hsiao-pin, 2011), which is a standardised measure of vocabulary depth. In this test, participants were instructed to verbally define a list of words that gradually increased in difficulty. Each definition is scored from 0 to 3 based on accuracy and completeness, and testing ceased when participants reached ceiling performance (3 consecutive scores of 0) or the last item. We also used the Peabody Picture Vocabulary Test 4th Edition (Dunn & Dunn, 2007), which is a measure of vocabulary breadth. In this test, participants were presented with a spoken word and required to choose one of four pictures which best described this word.

This test comprised of 228 words, grouped into 19 sets of 12-items arranged in increasing difficulty. Participants started from age appropriate word set, and testing ceased when ceiling performance (8 or more errors in any given set) or the last word set reached. Raw scores from both vocabulary measures were converted to age-normed standardized scores based on the scoring manuals.

Text Exposure. A measure of text exposure—the author recognition test— previously shown to correlate with relative clause processing (Acheson, MacDonald, & Wells, 2008; J L Montag & MacDonald, 2015) was also included in our study. Participants were presented with a list of author names and foils and were asked to identify which ones are names of authors. This measure correlated with expressive and receptive vocabulary measures, as shown in Table 2.2.

Motor Inhibition. The STOP-IT task was taken from Verbruggen and Logan (2008) as a measure of inhibition, and was conducted using the dedicated programme (https://ore.exeter.ac.uk/repository/handle/10871/13860). In this task, participants' primary goal is to respond with two different keys to two different visual cues (square and circle) as fast as possible. Participants are instructed to withhold their response if they hear an auditory stop signal, which randomly occurs in one fourth of the trials. In an adaptive staircase tracking procedure, the time between the primary stimulus and the stop signal is increased or decreased, depending on whether inhibition was successful, resulting in a 0.5 probability of responding. This allows to estimate the time it takes to covertly stop a response—the stop-signal reaction time (SSRT).

Semantic and phonological inhibition in homograph and homonym tasks. A set of ambiguous words (24 homographs e.g., *wind*, and 24 homonyms e.g., *bank*) were selected from existing studies and databases (Henderson, Snowling, & Clarke, 2013; Twilley, Dixon, Taylor, & Clark, 1994), along with unambiguous filler words (24 and 15 in each task respectively). See Appendix C for a full stimulus list and trial structure. An ambiguous word was presented twice in two different contexts and blocks. The first block presented ambiguous words in contexts priming their dominant interpretation, whereas the second block presented these words in contexts priming their subordinate interpretation. The contexts were words preceding the target word, which were semantically related to the dominant or subordinate meaning of the ambiguous words (e.g., *money* or *river* preceded *bank* in the homonym task and *blow* or *turn* preceded *wind* in the homographs task). In the homograph task, participants were instructed to read the words out loud into a microphone, whereas in the homonym task, they indicated by pressing one of two keys whether or not the second word was semantically related to the previous one. Sound files from the homograph tasks were saved, and speech onset times (SOTs) were obtained by visually identifying the first sound of the word. The difference in response times or SOTs between the two presentations of the same word (subordinate meaning – dominant meaning) was taken as an index of inhibition difficulty (henceforth referred to as *inhibition scores*), if the response was correct. Since the first word presentation strengthens the already dominant meaning of the word (e.g., *wind*), the second presentation in the context of the subordinate meaning (e.g., *turn*) requires the inhibition of the dominant prepotent meaning and/or pronunciation. In the homograph task, this inhibition likely requires both semantic and phonological inhibition, whereas in the homonym task, only semantic inhibition is likely to occur.

2.2.4 Data coding and analyses

In the comprehension task, only responses to the experimental trials were coded for accuracy and analysed. One participant's data was removed for below-chance performance (<50% correct). The production data were first coded for accuracy. If a description was skipped or did not include the targeted structure but was correct, e.g., *the girl sitting down, the man looking scared*, *the apple on the pole*, it was removed from the analyses (38 responses per condition were excluded from a total of 1491 responses in each condition). In the remaining responses, the accuracy of the descriptions was generally high (animate target, M=100%, SD=0.9; inanimate target, M= 97%, SD =6), but there was a significant difference across conditions (Wilcoxon test: z=4.5, p<.001). The errors in the inanimate-target condition were due to speakers often describing the wrong target, for example, the animate entity was described rather the inanimate one, e.g., *the girl hugging the* *teddy bear* for the inanimate target description in Figure 2.1. This might be due to a tendency to focus on the human character, rather than the objects. Finally, valid responses were coded as active or passive phrases, and passives were further coded for agent omissions (*by*-phrase omission).

From the audio files recorded in production, the total duration of the spoken phrase was automatically obtained in Praat and manually checked. This duration was divided by the total number of characters in the phrase, which was taken as a proxy for the total number of sounds in the phrase (total utterance duration /number of sounds). This provided a measure of the proportion of time spent per sound, so that a larger ratio indicates more pauses or intermediate hesitations during production. This speech fluency measure has been extensively used in clinical research, e.g., (Buchanan, Laures-Gore, & Duff, 2014; Gernsbacher, Sauer, Geye, Schweigert, & Hill Goldsmith, 2008; Marchina, Wang, Wan, Norton, & Schlaug, 2013), and provides a global assessment of utterance difficulty that is not linked to the retrieval of the first word, which is typically reflected in speech onset times (SOTs). Note that because inanimate nouns are less frequent and less conceptually salient, and moreover were generally longer than animate nouns, we would expect SOTs in our study to be longer for these nouns (J K Bock & Warren, 1985; Z M Griffin, 2001; Z M Griffin & Bock, 2000; McDonald, Bock, & Kelly, 1993). Therefore, SOTs were not a good measure to examine production difficulty beyond the first noun, even if they also include other non-lexical processes. Nevertheless, we compared SOTs for animate nouns across reversibility conditions, and the results of these comparisons are reported in the Appendix D.

Analyses were conducted using linear mixed-effects models (LMEMs) in R (version 3.4.1, bobyqa optimizer and maximum iterations set at 100,000 for dichotomous dependent variables (Kuznetsova, Brockhoff, & Christensen, 2017; R Team Core, 2017). All the initial mixed-effects models included the maximal random-effects structures allowed by the design (by-subject and by-item intercepts, by-subject and by-item random slopes for the animacy condition; and only by-subject slope for the reversibility conditions). In the animacy-based analyses, the animate target condition was coded as 1 and inanimate target condition was coded as -1. In the analyses with only animate targets, the high-reversibility condition was the reference category. Thus, a positive coefficient of the condition main effect would represent a higher value in the more difficult condition (i.e. animate or high-reversibility condition).

Continuous dependent variables (DVs) and cognitive predictors were mean-centred prior to analysis (Iacobucci, Schneider, Popovich, & Bakamitsos, 2016). For analysis of dichotomous DVs (i.e. passive/active sentence structure), logistic regression models were used, and predictors were z-scored to achieve convergence. To minimize the influence of outliers, we removed extreme values in all DVs (e.g., RTs longer than 5s) and values falling above 3.5 *SDs* from an individual's mean per animacy condition (i.e. for animate and inanimate conditions separately). This procedure was applied to comprehension RTs and production SOTs and fluency scores. In all cases, these exclusions comprised less than 2.2% of the data.

To examine the unique contribution of different cognitive skills, we entered individual difference measures and their interactions with conditions in a priori selected order. Vocabulary measures (i.e., expressive and receptive vocabulary) and reading experience (author recognition scores) were entered first to account for the role of lexical knowledge and linguistic experience. Inhibition measures (including stop-it SSRT, homograph and homonym inhibition scores) were entered secondly to investigate whether inhibition ability explains additional variance after accounting for the effects of vocabulary or experience. In all analyses, individual cognitive skills with non-significant main effects or interactions were pruned to identify the simplest most explanatory model. Thus, only the significant effects are reported. As shown in Table 2.1, except for the vocabulary measures and reading experience measure, none of the inhibition measures were correlated with each other or with vocabulary measures, suggesting that these measures tap on different underlying skills. The correlation matrix between predictors is shown in Table 2.2.

Table 2. 1 Descriptive statistics for individual differences variables

Measures	*N*	Range	Mean	*SD*
Expressive vocabulary (WAIS-II)	83	46 - 80	57.94	7.94
Receptive vocabulary (PPTV-IV)	83	87 - 139	107.49	10.29
Author Recognition Score	83	1 - 44	12.85	7.48
STOP-IT (SSRT)	78	147 - 396	264.50	42.08
Homonym Inhibition Score	81	-273 - 449	119.64	159.54
Homograph Inhibition Score	81	-140 - 660	161.23	161.37

Table 2. 2 Pearson's correlations between individual differences measures

	Expr. Vocab.	Recep. Vocab.	Author Rec.	STOP-IT	Homonym inhibition	Homograph inhibition
Expr. Vocab.		0.58**	0.38**	-0.00	-0.08	0.11
Recep. Vocab.	0.58**		0.54**	-0.00	0.08	0.04
Author Rec.	.38**	.54**		0.13	0.14	-0.02
STOP-IT	0.00	0.00	0.13		-0.03	0.08
Homonym Inhibition	-0.08	0.08	0.14	-0.03		0.01
Homograph Inhibition	0.11	0.04	-0.02	0.08	0.01	

**Correlation is significant at the 0.01 level (2-tailed).
Note: Expr. Vocab. stands for Expressive vocabulary (WAIS-II), Recep. Vocab. stands for Receptive vocabulary (PPTV-IV), Author Rec. stands for Author Recognition Score

2.3 Results

2.3.1 Comprehension task

Animacy effects. Based on previous results, we expected that RTs would be longer for animate-target than inanimate-target phrases. Importantly, we expected that individuals with poorer inhibition skills would experience more difficulty in the animate-target condition after vocabulary measures were accounted for in the model (interaction between inhibition skills and animacy). Table 2.3 summarizes the results. There was a significant main effect of animacy condition such that animate-target phrases took longer to comprehend than the inanimate-target phrases. Moreover, there was no main effect of Expressive Vocabulary, but there was a significant interaction between Expressive Vocabulary and Animacy: Participants with poorer vocabulary experienced more difficulty in the animate than the inanimate condition, compared to those with better vocabulary, who showed smaller differences between conditions (Figure 2.2). Importantly, as expected, there were a significant main effect of Homograph Inhibition and a Homograph Inhibition*Animacy interaction. Individuals with poorer inhibition scores were generally slower that those with better scores and showed more difficulty in the animate-target than the inanimate-target condition (Figure 2.2).

Table 2. 3 Results of LMEMs predicting comprehension RTs from head-noun animacy and individual difference measures

	Coefficient	**SE**	**t-score**	**p-value**
Intercept	441.90	40.65	10.86	<0.01*
Animacy	141.30	31.52	4.48	<0.01*
Expressive Vocabulary	-5.86	4.58	-1.28	0.21
Homograph Inhibition Score	1.11	0.23	4.88	<0.01*
Expr. Vocab.*Animacy	-7.95	2.17	-3.66	<0.01*
Homograph Inhibition*Animacy	0.21	0.11	1.93	0.05

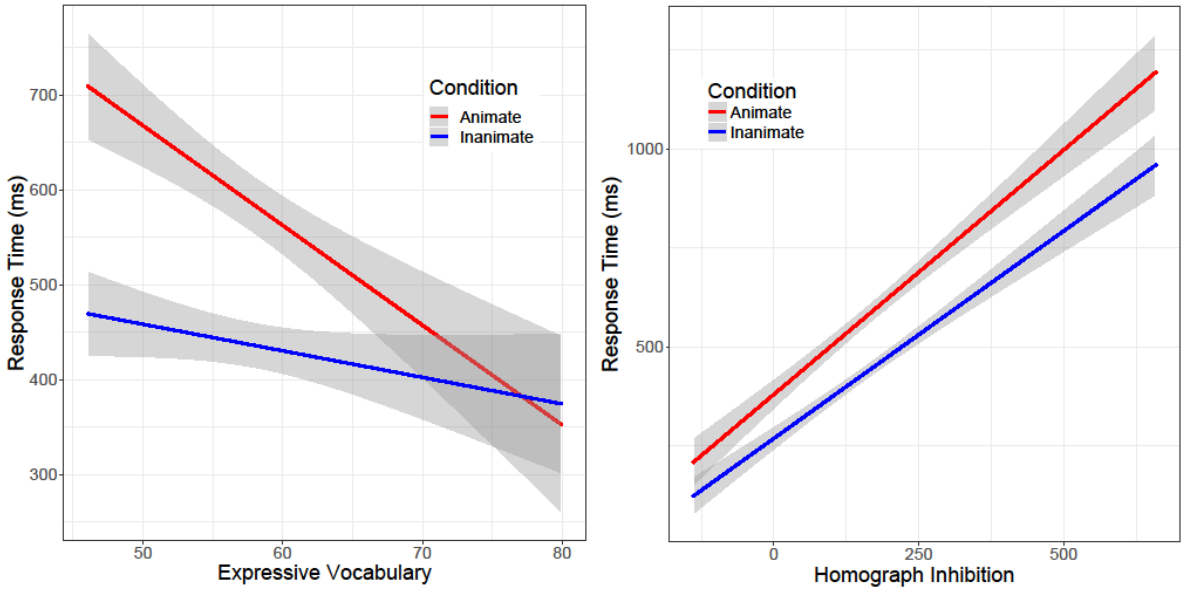


Figure 2. 2 Interaction between target animacy and vocabulary (left panel) and between Animacy and Homograph Inhibition Scores in predicting comprehension RTs. Shading indicates standard errors. Inhibition scores represent the RT difference in naming subordinate vs dominant pronunciations of ambiguous words so that smaller differences indicate better inhibition.

Reversibility effects. We expected that within animate-target phrases, high-reversibility phrases (e.g., *the man that the girl is hugging*) should be more difficult than low-reversibility phrases (e.g., *the dog that the girl is hugging)*. In addition to any influence of vocabulary knowledge, participants with poorer inhibition scores are expected to experience more difficulty in the high-reversibility condition. The results are shown in Table 2.4. There was a significant main effect of Reversibility: high-reversible phrases took longer to process than low-reversibility ones. There were also main effects of Expressive Vocabulary and Homograph Inhibition, indicating that better inhibition and vocabulary skills were associated with faster RTs. Importantly, there were interactions between Reversibility and Expressive Vocabulary and Homograph Inhibition (see Figure 2.3). Participants with lower vocabulary scores found the High- and Medium-reversibility conditions more difficult than the Low-reversibility condition, whereas participants with poorer inhibition skills showed more difficulty in the High-Reversibility condition. In contrast, participants with better vocabulary and inhibition scores were less affected by Reversibility, and thus showed smaller differences between conditions.

Table 2. 4 Main effects and interactions for reversibility conditions and individual difference measures in predicting comprehension RTs with animate-head phrases

	Coefficient	SE	*t*-score	*p*
Intercept	654.51	60.85	10.76	<0.001*
High v Low Reversibility	-170.53	60.19	-2.83	0.007*
Expr. Vocab	-19.55	6.29	-3.11	0.003*
Inhibition Scores	1.60	0.31	5.13	<.0001*
Expr. Vocab* High v Low Reversibility	10.162	4.33	2.35	0.02*
Inhibition Scores*High v Medium Reversibility	-0.4266	0.19	-2.136	0.03*
Inhibition Scores*High v Low Reversibility	-0.4424	0.217	-2.039	0.04*

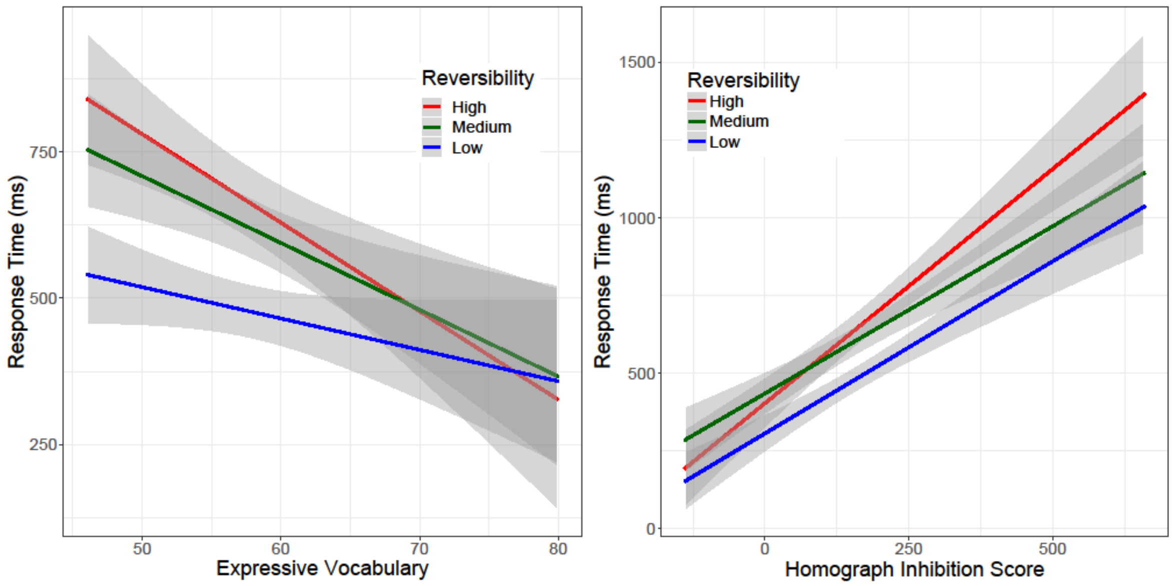


Figure 2. 3 Interactions between Reversibility and a vocabulary measure (left panel) and Reversibility and Homograph Inhibition (right panel) in predicting comprehension performance for animate-target pictures. Shading indicates standard error.

Taken together, the Animacy and Reversibility results indicate a role for vocabulary knowledge and inhibition skills in resolving competition in phrase comprehension: participants with better vocabulary and inhibition skills were faster at processing the more difficult animate-target phrases than those with poorer vocabulary and inhibition skills, hence their interaction with conditions.

2.3.2 Production task

2.3.2.1 Active vs. passive phrase structure choices

Animacy effects. Following previous studies, we expected that animate-targets would tend to be described in passive forms (e.g., *the man being hugged by the girl* for Fig. 2.1) more often than inanimate-targets, although passives are the most frequent overall strategy because they are also used for inanimate targets about 50% of the time. This strategy is likely to be more frequent in the present task, because unlike previous studies using more naturalistic question-answer paradigms, participants have a single goal throughout the task, namely, identify the highlighted character, and could thus redeploy a previously used strategy. We therefore reasoned that since resourcing to passives rather than actives is the most frequent production strategy, the production of active structures should correlate with a better ability to resolve competition in a way that is as cost-effective as producing passives. Thus, participants with better inhibition skills might be able to produce more active phrases (fewer passives) than those with poorer inhibition in both conditions, and specifically, in the animate-target condition.

Results are shown in Table 2.5. There was a main effect of Animacy indicating that participants produced more passives for animate-targets (mean= 80%, SD=22%, median=89%), than inanimate-targets (mean= 64%; SD=31%, median=65%). There were also two interactions with individual differences measures: An Animacy*Recep. Vocab. interaction indicated that participants with poorer receptive vocabulary showed larger differences between the two conditions than those with better vocabulary, and participants with poorer vocabulary produced fewer passives (more actives) for inanimate-targets. In contrast, the Animacy*Homograph Inhibition interaction showed the opposite pattern: Participants with poorer inhibition (larger scores) tended to produce passives in the two animacy conditions (there was a significant relationship in each separate condition, $p<0.02$) whereas participants with better inhibition generally produced fewer passives but showed larger differences between conditions, hence the interaction. Vocabulary and inhibition skills thus exerted different influences on structure choices but both measures were better predictors of inanimate-target than animate-target production, in part because there was more room to observe differences (more variability) in the inanimate-target distribution.

Table 2. 5 Model results predicting passive choices from noun animacy and individual difference measures

	Coefficient	**SE**	***z*-score**	***p*-value**
Intercept	1.1079	0.2816	3.934	<0.01*
Animacy	1.3301	0.1780	7.474	<0.01*
Receptive Vocabulary (PPTV)	0.4321	0.2587	1.670	0.09
Homograph Inhibition	0.7115	0.2720	2.616	0.01*
Receptive Vocabulary*Animacy	-0.2789	0.1235	-2.259	0.02*
Homograph Inhibition*Animacy	-0.2755	0.1432	-1.924	0.05*

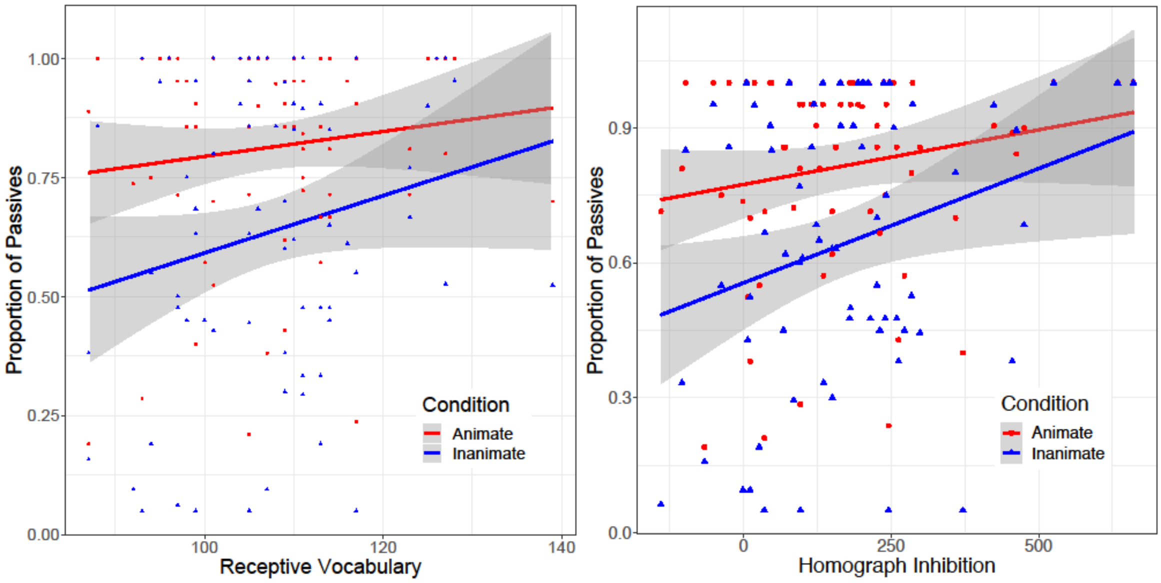


Figure 2. 4 Interaction between Animacy and the receptive vocabulary measure (left panel) and Homograph Inhibition scores (right panel) in predicting the proportion passives produced. Shading indicates standard error.

The fact that participants with poorer vocabulary tended to produce more actives (fewer passives) might stem from more general linguistic experience. Recall that our vocabulary measures were positively correlated with our measure of reading experience (see section 2.3.2 and Table 2.2). Less exposure to authored texts, which tend to contain more passives than oral language, has been shown to correlate with production of fewer passives (J L Montag & MacDonald, 2015). Participants with poorer vocabulary may thus not entertain passives as a viable alternative structure, unless they find the planning difficult (hence the animacy difference). This possibility is consistent with alternative statistical models we have conducted in which Recep. Vocab. is replaced by scores from the author recognition test. Although these scores do not explain unique variance over and above the Recep. Vocab. scores, they yield almost identical results to those of Table 2.5. These observations therefore suggest that the role of vocabulary in passive structure choice may not be due to vocabulary knowledge per se but to other aspects of linguistic knowledge correlated with vocabulary.

The inhibition results on the other hand, were generally consistent with the expectation that participants with good inhibition skills should be better able to entertain actives as a viable alternative structure than those with poorer inhibition. Nevertheless, the role of inhibition not only in animate-target phrases but in inanimate-target ones additionally suggests that there may be some level of competition in these phrases too. Taken together, the present results are consistent with those of comprehension in that both vocabulary and homograph inhibition scores underpin production choices.

Reversibility effects. Following previous findings, we expected more passive descriptions for high-reversibility than low-reversibility findings. Despite the small variability among animate-target passives, there was a main effect of reversibility in explaining passive use. In particular, the high-reversibility condition was associated with more passives than the low-reversibility condition ($z=1.97$, $p=0.05$), suggesting that overall, participants resorted to passive structures more often when the nouns were reversible. Unlike the animacy model, however, reversibility did not interact with individual difference measures.

2.3.2.2 Animacy and reversibility in agent omissions

Previous production results have shown that speakers who produce passives tend to omit the agent *by*-phrase more often in animate-target-descriptions (Gennari, Mirkovic, et al., 2012; Hsiao et al., 2014; Hsiao & MacDonald, 2016; Humphreys et al., 2016; Perera & Srivastava, 2015). The rate of omissions however is significantly reduced by contextual manipulations promoting the use of active structures (see Experiment 1 in Humphreys et al., 2016). Compared to previous studies, our study revealed relatively fewer omissions (mean animate condition: 8%, mean inanimate condition: 4%), perhaps because participants believed they had to fully describe the event in which the target character took part. Although it is unclear how agent omissions should relate to individual differences, one possibility is that speakers tend to omit the agent *by*-phrase because the agent concept was initially inhibited during selection of the first patient noun (e.g., *man* in *the man being hugged by the girl*), and thus, the omission reflects the lesser accessibility of this noun by the end of the structure (Gennari et al, 2012; Hsiao et al, 2014). This possibility predicts that when passives are produced, participants with better inhibition would inhibit the agent relatively more strongly than those with poorer inhibition, and thus, would tend to omit the agent more often in the more competitive conditions (animate-target and high-reversibility conditions).

Animacy results indicated a main effect of Animacy ($z=2.19$, $p=0.02$) but no relationship or interactions with individual differences measures. Interestingly, Reversibility results indicated a main effect of condition and interactions with individual differences. There were more agent omissions in the High-reversibility group (mean =11%, SD=19%) than the Low-reversibility group (mean =5%, SD=18%). Reversibility interacted with Homograph Inhibition scores and marginally with Expressive Vocabulary (Table 2.6, Figure 2.5). Vocabulary and inhibition scores showed opposite trends, as reported above for passives. The marginal interaction with vocabulary suggests that participants with better vocabulary omitted agents equally often across conditions, whereas those with poorer vocabulary omitted agents more often in the High-reversibility than the Low-reversibility condition. This suggests that the less accessible concept/noun, which had been inhibited at the beginning of the phrase, was harder to retrieve for participants with relatively poorer vocabulary. In contrast, participants with better inhibition scores (smaller values), omitted the agent more often in the High-reversibility condition than the Low-reversibility condition. As expected, participants with better inhibition, strongly inhibited the agent concept during first noun selection, making it less accessible later on in the structure. Overall, as in phrase structure choices, vocabulary and Homograph Inhibition underpin *by*-phrase omissions.

Table 2. 6 Model results predicting agent omission from reversibility conditions and individual difference measures

	Coef.	SE	*z*-score	*p*-value
Intercept	-3.60	0.60	-6.04	<0.01*
Reversibility High v Low	-3.99	1.75	-2.27	<0.02*
Expressive Vocabulary	-0.21	0.35	-0.60	0.54
Homograph Inhibition	-0.33	0.35	-0.92	0.35
Expressive Vocabulary* Reversibility High v Low	1.20	0.63	1.90	0.06
Homograph Inhibition*Reversibility High v Low	0.62	0.37	2.07	0.04*

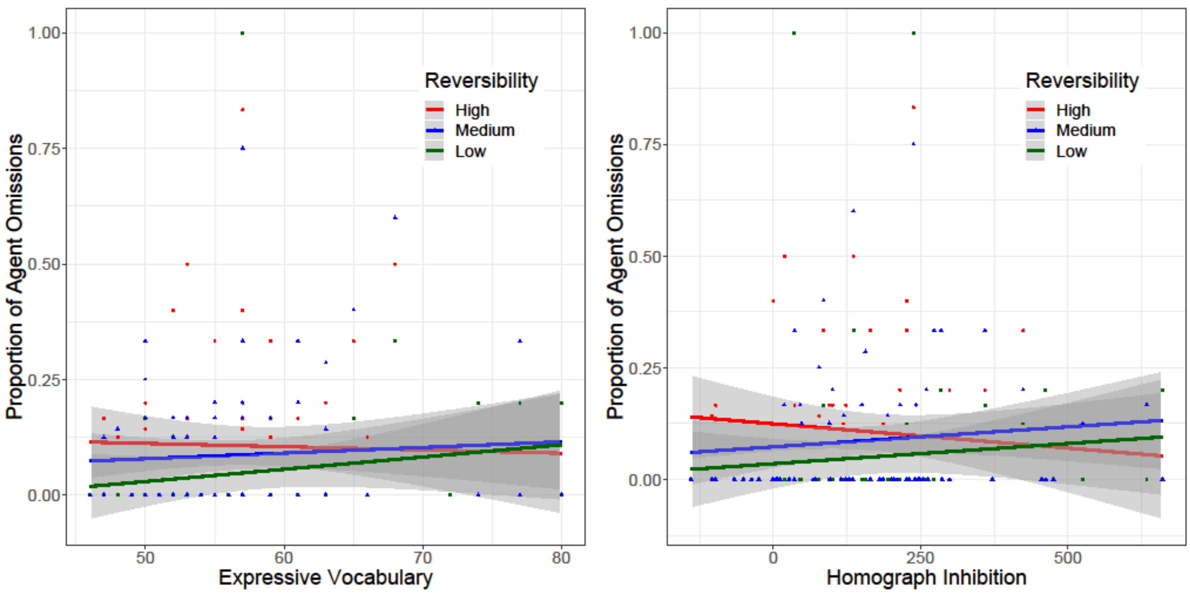


Figure 2. 5 Interaction between reversibility conditions and Expressive vocabulary (left panel) and Homograph inhibition scores (right panel) in predicting the proportion of agent omissions. Shading indicates standard error.

2.3.2.3 Verbal fluency

The proportion of time spent per phonemes in the utterance provides a global measure of difficulty independently of speech onset times (total utterance duration/number of sounds), with larger values representing more time spent per sound and thus, slower sequential planning and phrase-internal disfluencies. In line with our hypotheses, we expected that if inhibition played any role during production of animate and inanimate-target phrases, animacy should interact with inhibition measures. To test this prediction, we analysed only passive structures so as not to confound results with structure choice.

The results are summarized in Table 2.7. There was a significant main effect of Homograph Inhibition such that participants with better inhibition produced more fluent descriptions. Importantly, there was a significant interaction between target animacy and STOP-IT performance (Figure 2.6). Specifically, as the SSRT (time to stop a response) increased, disfluency increased more in the animate-target than the inanimate-target condition. That is, people with better motor inhibition produced more fluent descriptions for animate-target pictures than inanimate-target pictures, whereas this pattern reverses for people with poorer inhibition. These results suggest that the speed with which participants inhibit a motor response plays a role in production.

Table 2. 7 Results of models predicting production fluency from noun animacy or Reversibility and individual difference measures

Model		**Coefficient**	**SE**	***t*-score**	***p*-value**
Animacy	Intercept	77.04	1.84	41.83	<0.01*
	Animacy	-0.08	1.04	-0.08	0.93
	Stop it	0.031	0.04	0.79	0.43
	Homograph Inhibition	0.03	0.00	2.62	0.01*
	Stop it * Animacy	0.05	0.02	2.23	0.03*
Reversibility	Intercept	78.47	2.46	31.81	<0.001*
	High v Low	-5.15	2.50	-2.05	.05*
	Homograph Inhibition	0.02	.009	2.06	.04*

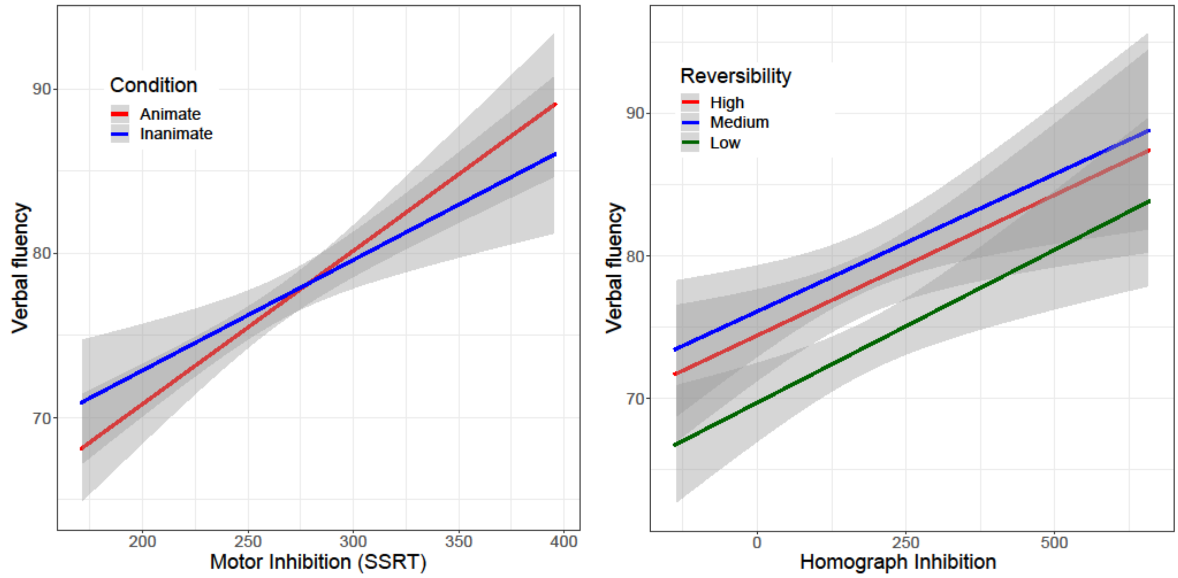


Figure 2. 6 The interaction between Animacy and Motor Inhibition (left panel) in predicting verbal fluency (utterance duration/phonemes), and the relationship between verbal fluency and Homograph Inhibition across reversibility conditions (right panel). Shaded area represents standard errors.

We also examined the possibility that verbal fluency would vary with reversibility conditions, as we expected potentially more disfluent utterance in the high-reversibility condition, even though there was no difference across animacy conditions. The results indicated that there was a main effect of reversibility, with low-reversibility items being more fluent than high-reversibility items (Table 2.7). As in the animacy model, Homograph Inhibition significantly predicted verbal fluency beyond the reversibility effect, suggesting that speakers with better inhibition produced more fluent descriptions. See Figure 2.6, right panel. Overall, the fluency results suggest that like in comprehension, Homograph Inhibition was a good predictor of production in all conditions. However, unlike comprehension, a different measure of inhibition—motor inhibition—also contributed to production processes. Thus, although there were some commonalities across production and comprehension, the present results also highlight different skills underpinning each task.

2.4 Discussion

2.4.1 Summary of results

The present work aimed to establish whether inhibition skills explained production and comprehension over and above any influence of vocabulary skills. Using an individual difference approach, we aimed to establish whether the same inhibition-related cognitive skills underpin both production and comprehension of complex referential phrases varying in animacy and reversibility. Based on previous results, we expected that target-animacy and noun-reversibility would modulate production and comprehension difficulty and moreover, that this modulation would vary as a function of individual inhibition skills, with better performers exhibiting less difficulty than poorer performers in the most difficult conditions.

2.4.1.1 Animacy and reversibility results

The present results are consistent with these expectations. Animacy and reversibility modulated both production and comprehension, consistent with previous reports. Speakers produced more passives, more often omitted agents and comprehension responses were longer for animate-target and high-reversibility phrases, compared to inanimate-target and low-reversibility phrases respectively, suggesting that animacy and high-reversibility elicited more difficulty. Moreover, production fluency was poorer in high-reversibility than low-reversibility phrases, suggesting that high-reversibility increased production difficulty. These results are consistent with previous claims that highly similar and reversible nouns compete during processing (Gennari et al., 2012; Humphreys et al., 2016).

Table 2. 8 Summary of results

Task	DV	Fixed factor	Fixed Effect	Predictors	Interaction
Comprehension	RT	Animacy	An > In		Expr. Vocab.
				Hmg Inhibition	Hmg Inhibition
		Reversibility	High > Low	Expr. Vocab	Expr. Vocab
				Hmg Inhibition	Hmg Inhibition
Production	Passives	Animacy	An > In	Hmg Inhibition	Recep. Vocab.
					Hmg Inhibition
		Reversibility	High > Low		
	Ag Om	Animacy	An > In		
		Reversibility	High > Low		Expr. Vocab
					Hmg Inhibition
	Fluency	Animacy	-	Hmg Inhibition	Stop-it inhibition
		Reversibility	High > Low	Hmg Inhibition	

Note: *Hmg* stands for *homograph, Ag Om* stands for *agent omissions, An* and *In* stands for *Animate* and *Inanimate* respectively.

2.4.1.2 Individual differences results

Common skills

Vocabulary knowledge. We found that vocabulary knowledge interacted with animacy and/or reversibility condition in predicting comprehension times and utterance choice. This confirms Perfetti's lexical quality hypothesis and Van Dyke, Johns and Kukona (2014)'s findings, in suggesting comprehenders with better vocabulary are less affected by an interference. The comprehension of animate-head and high-reversibility phrases is more costly for participants with poorer vocabulary not necessarily because of their vocabulary knowledge per se, as they experienced more difficulties with animate words that are surely well known (animate nouns such as "*the man/woman*" should be more frequent and accessible than inanimate nouns such as "*the gong/trophy*"). Rather, it might be that the lexical representations they are likely to retrieve to encode the noun concepts are poor in quality. These representations may omit important features (e.g. information regarding to possible thematic roles) to discriminate between similar concepts when they need to be processed close in time in actives, thus resulting in greater competition. Similarly, in production, the influence of vocabulary knowledge can be attributed to the robustness of representations, but with syntactic structures rather than noun concepts. In production, speakers are not restricted to plan two similar nouns in a sequence, and individuals with better vocabulary (and possibly more exposure to authored text) produced more passive structures which appear more frequently in written text than oral language, suggesting that they are more likely to consider passives as a viable alternative in addition to the primed actives, and production behaviour is shaped by individual's familiarity with different structures.

Homograph inhibition. More importantly, among three measures of inhibition, only homograph inhibition scores—the difference between the subordinate and dominant word meanings and pronunciations—predicted performance in both production and comprehension, over and above the influence of vocabulary knowledge (see Table 2.8). Homograph inhibition scores predicted comprehension times, verbal fluency and passivisation rates for the two animacy conditions. Individuals with better inhibition scores were faster at comprehending, more fluent and produce fewer passives than individual with poorer scores. These associations suggest that participants with better inhibition scores were not only faster in processing but also more able to plan active structures with two nouns in sequence—a configuration that most speakers avoid and comprehenders find difficult, particularly with animate nouns (Gennari & MacDonald, 2009; Roland, Dick, & Elman, 2007).

For some comprehension and production measures, inhibition scores more strongly predicted the more difficult conditions, namely, animate-targets and high-reversibility phrases (see Table 2.8), although this was not the case for passive proportions. Nevertheless, individuals with better inhibition tended to use fewer passives and omit agents more often in the animate-target or the high-reversibility condition, suggesting stronger inhibition and subsequent less accessibility of the agent concept/noun at the end of the structure.

An indication of what specific semantic inhibition processes are involved can be inferred from the nature of the homograph inhibition task. Participants do not only need to inhibit the recently primed and most accessible meaning of an ambiguous form, but also the pronunciation that corresponds to the most accessible meaning. Thus, the inhibition component captured by this task specifically targets the mapping from meaning to word phonology. It is this link that is critical to both listening and speaking tasks and distinguishes this task from a purely motor inhibition task (the STOP-IT task) and a non-phonological semantic inhibition task (the homonym task).

Taken together, these results suggest that common skills underspin both production and comprehension. That vocabulary knowledge may matter for both tasks is not surprising, but that the homograph task, which capture the ease with which participant can inhibit context-inappropriate meanings, is significant. Given that this inhibition process plays a role in explaining production and comprehension performance in cases involving competition, the results suggests the possibility that a similar type of inhibition operate in both tasks. This is consistent with the influential "conflict resolution account" which argues the LIFG is involved in both production and comprehension of high-ambiguity sentences, and is associated with semantic control and inhibition processes to resolve competition between activated representations (e.g. Thompson-Schill et al, 2005; Novick et al., 2009; Novick, Trueswell & Thompson-Schill, 2005; Thompson-Schill, et al., 1997).

Distinct skills

Motor inhibition. Unlike comprehension, a measure of motor inhibition from the Stop-it task was a better predictor of verbal fluency in animate-target than inanimate targets. This finding is consistent with the fact that only production engages an overt motor response and suggests that the mechanisms engaged in the STOP-IT task may have been more strongly recruited in the animate-target condition. Note that the STOP-IT task not only requires motor inhibition but also monitoring and flexible adjustments to competing task demands, namely, responding as fast as possible and stopping the response when instructed. Indeed, performance in this task correlates with interference control more generally (Friedman & Miyake, 2004; Verbruggen & Logan, 2008). Thus, the present results may indicate not only motor inhibition processes, but also more general executive functions, which may operate at different levels of linguistic representations (phonology, syntax and semantics).

Our results also show that comprehension and production differ in terms of task-difficulty. This is evident from the finding that in some analyses, inhibition measures were more strongly associated with producing inanimate-head phrases as compared to animate-head phrases; whereas the opposite relationship was observed with comprehension times. For example, inhibition measure more strongly predicted the propensity to produce passives in the inanimate-target condition as compared to the animate condition. Besides differences in variability across conditions as previously discussed, this suggests that unlike comprehension, production recruit inhibition processes even in cases where assigning the syntactic roles is assumed to be non-interfering (i.e. inanimate condition). Thus, production of inanimate-target phrases can be viewed as more cognitively demanding than comprehension of the same events. This is because in describing inanimate-targets of transitive events, agent-initial descriptions (e.g. the girl hugging the teddy bear), which are the most frequent in the language (Roland, Dick, & Elman, 2007), and animate characters, which are visually and conceptually more salient (Bock, 1987; Bock & Levelt, 1994), must be inhibited. Indeed, we found more errors for inanimate-target than animate-target phrases in which speakers described the animate agent first, thus naming the person rather than the highlighted object. Therefore, unlike comprehension where semantic competition mostly depend on the degree of reversibility-based effects, production competition occurs at different levels of linguistic representations and is not necessarily restricted to the semantic reversibility between noun concepts. In describing nouns which are not semantically similar and reversible, competition occurs as speakers must access less accessible concepts and forms, while avoiding naming the more accessible characters first and alternative agent-first plans.

2.4.2 Conclusion

Overall, the relationship between individual cognitive skills and performance indicate common as well as distinct cognitive skills underpinning phrase production and comprehension. Our results are the first to provide initial evidence suggesting that inhibitory skills may be shared across sentence production and comprehension over and above the influence of vocabulary knowledge. These results imply that common processes linked to competition resolution and inhibition of competing alternatives may operate in both tasks. Thus, any theory of comprehension, production and the relationship between the two should not only account for knowledge-based commonalities, but also for the presence of competitive processes in both tasks. We will come back to the implications of these results for production-comprehension theories in the general discussion.

Chapter 3

Children's comprehension and production of complex phrases: the role of age and individual differences

3.1 Introduction

Study 1 suggests that healthy adults experience similar syntactic-semantic competition in both comprehension and production of complex phrases. The to-be-comprehended or to-be-planned nouns elicited competition relating to the noun semantic features and the syntactic role assignment (i.e. highly similar nouns competed for the agent role). For comprehension, this manifested as a processing cost in comprehension response time. For production, competition occurs as different points and was reflected by increased SOTs, an increased tendency to use agentless passives to ameliorate planning burden and reduced verbal fluency throughout the utterance. Crucially, we found that common cognitive skills predicted individual's sensitivity to semantic competition across comprehension and production: variations in semantic inhibition contributes to competition resolution over above vocabulary knowledge; as well as distinctive processes underpinning production (i.e. action planning). These findings have important implications for current language models, which need to be developed further to account for a common verbal inhibition across comprehension and production, as well as distinctive production mechanisms (see General Discussion).

However, effective models of language comprehension and production need to account not only for individual differences, but also for development. Cognitive skills such as vocabulary and inhibition are not static but vary substantially within individuals and across the lifespan. For example, there is considerable growth within individuals in vocabulary size and depth over the school years into adolescence (Bornstein, Hahn, & Haynes, 2004; Verhoeven, van Leeuwe, & Vermeer, 2011). Across the same developmental period, the prefrontal cortex, which is assumed to be closely related with executive function (including inhibitory control), undergoes neural structural development (e.g., changes in grey matter volume, white matter volume, cortical thickness (Gogtay et al., 2004; Shaw et al., 2008). This prefrontal immaturity is associated with diminished performance on inhibition tasks such as GO/NO-go, stroop tasks (Bedard et al., 2002; Rubia et al., 2006; Tamm, Menon, & Reiss, 2002). These findings raise the question of whether and how the relationship between cognitive skills and language ability changes over this developmental period. Do cognitive skills such as inhibition and vocabulary remain strong predictors of comprehension and production over development? Or, are there different cognitive skills important operating at different phases of development?

3.1.1 Relationship between language skills and cognitive variables in children

A number of previous studies have identified correlations between aspects of sentence comprehension and several cognitive skills, whereas investigation of sentence production has received scant attention. First of all, vocabulary is one of the best known predictors of language comprehension and production, generally because it provides a knowledge base for children to start understanding or making references to concepts encountered in daily life, and it is also one of the most commonly used measures of language competence across the lifespan (Dickinson, Golinkoff, & Hirsh-Pasek, 2010). A number of studies have reported that the breath, depth of vocabulary, or the speed of access of individual words is related to children's literacy acquisition and educational achievement (Lee, 2011; Ouellette, 2006). Moreover, evidence from children with poor reading comprehension suggested that when apparent difficulties with WM tasks and inhibition arise (i.e., inhibiting the contextually irrelevant meaning of a homonym), they are likely to be a consequence of poor vocabulary knowledge, rather than poor memory capacity or inhibition *per se* (Henderson et al., 2013; Nation, Adams, Bowyer-Crane, & Snowling, 1999). This suggests that vocabulary is central to comprehension, which limits poor comprehenders' performance on verbally mediated WM and inhibition tasks.

Closely related to vocabulary knowledge, linguistic experience has also been argued to be a key predictor of children's language development. For example, several studies have reported that parental interactions with children in spoken (e.g. child-directed speech from caregivers) or written language (e.g. shared book reading) significantly predicts early vocabulary growth (Farrant & Zubrick, 2013; Weisleder & Fernald, 2013), even after controlling for socioeconomic status (Rowe, 2012). Greater linguistic experience does not only broaden the range of words that children might encounter, but also enriches their exposure to different syntactic structures, and thus could potentially contribute to children's grammatical development. The role of linguistic experience in sentence comprehension skills has been well attested in children (Diessel & Tomasello, 2005; Kidd et al., 2007), including a training study (Roth, 1984) which found more exposure to relative clauses significantly improves children's comprehension of the same structures. More recently, a production study reported a correlation between increased reading experience and increased tendency for children to produce adult-preferred structures which are more commonly seen in written context (Jessica L Montag & MacDonald, 2015). Notably however, vocabulary knowledge was not assessed in these studies, and thus it is not possible to tease apart the effects of linguistic experience from vocabulary knowledge.

As in adults, children's language performance could also be associated with non-linguistic cognitive skills. Working memory (WM) capacity has long been regarded as an important constraint of sentence comprehension in children (Marcel A Just & Carpenter, 1992; Swanson, 1996), but there is little research in production. In comprehension, WM serves to temporarily store and manipulate linguistic input, the capacity and processing resource of this system varies between individuals and directly constrains comprehension depending on resource availability. For example, during comprehending complex phrases such as *the man that the girl is hugging*, the comprehender must hold the two noun phrases (NP) in WM until the verb is processed, at which point the NP1 must be reactivated from memory and integrated into the developing structure (after the verb), with the correct agent/patient role assigned to each NP (Gordon & Lowder, 2012; Van Dyke & McElree. 2011). Thus, both the storage and reactivation of NP1 are involved in complex phrase comprehension. Similarly, production demands WM resources as speakers need to temporarily maintain information in WM before it can be grammatically produced. Previous research with adults has found that WM is involved in lexical accessibility (Belke, 2008) and agreement aspects of grammatical encoding (Slevc & Martin, 2016), such that when under a verbal WM load, speakers are less likely to produce assessible information early (Slevc, 2011); and low-span speakers make more agreement errors (Hartsuiker & Barkhuysen, 2006). These data show that increased cognitive load introduced by a secondary task slow down production process and low-span speakers are more affected, but it is less clear whether WM resources constraint production performance when cognitive load is introduced by difficulties in completing the syntactic planning of animate nouns.

Also, to date, the relationship between WM and complex sentence processing remains unclear, as different WM models make fundamentally different assumptions about the nature of WM effect and the appropriate WM task should be used to assess such effect. For example, there is a broad distinction between approaches that conceptualize WM as a unitary, domain-general system that controls the focus of attention; or approaches suggesting more domain-specific role of WM in processing verbal information. Unitary views hold that the domain-general attention mechanism in WM is central to cognitive performance, it serves to control attention to encode information from the input and then retain focus on specific information for the purpose of immediate cognitive goal (e.g. Cowan et al., 2005; McElree, 2001), or it allocates domain-general cognitive resources to a capacity-limited subsystem that stores phonological information (e.g. a *central executive* directs attentional resources to the *phonological loop*; Baddeley & Hitch, 1974). These approaches assume that the domain-general central executive is particularly important for the association between WM and cognitive performance, thus performance on language tasks could be predicted by measures on both verbal and non-verbal WM tasks.

Other approaches have suggested a more domain-specific system such as verbal WM. Some of these argued for a single-resource system with shared cognitive resources devoted to storage and processing of verbal information (Just & Carpenter, 1992; King & Just, 1991). Others suggested that the verbal WM could be divided into verbal WM for non-linguistic verbally mediated information (e.g. digits, nonword) and those for linguistic processing (Martin et al., 1999), or even further divided between specific types of linguistic processing, such as separate resources devoted to online vs. offline sentence comprehension (Caplan & Waters, 1999; Waters & Caplan 1997, 2004). Based on these approaches, complex span tasks targeting both storage and processing components of verbal WM (e.g. listening/reading span), would be a better predictor of language performance than simple span tasks (e.g. serial recall of a word list) which only measure short-term memory ——— the storage component of verbal WM.

In contrast to the above accounts in which some relation between WM resource capacity and language performance is assumed, a third perspective suggests that such relation is in fact mediated by individuals' long-term linguistic knowledge and experience with different components of language (MacDonald & Christiansen, 2002; Acheson & MacDonald, 2009; Wells et al., 2009), given that most conventional verbal WM tasks used (e.g. reading/listening span) have an inherent sentence processing component. Thus, instead of posing a causal role for WM capacity, this account would suggest that increase in children's linguistic experience and knowledge (e.g. vocabulary, reading experience) predicts improvement in WM task measures as well as improvement in comprehension performance (Cowan, Rouder, Blume, & Saults, 2012; Ericsson & Kintsch, 1995). This is also consistent with the claim that weak vocabulary in poor comprehenders restricts their ability to represent and manipulate verbal information in WM, and in turn affects verbally mediated WM task performance (Nation et al., 1999). Thus, the extent to which WM and vocabulary skills make independent contributions to reading comprehension still needs further investigation.

Finally, another key cognitive skill is inhibitory control. Since it is acknowledged that comprehension and production often involves competition between linguistic representations (Bedny, McGill, & Thompson-Schill, 2008; Gennari, Mirković, & MacDonald, 2012; Gordon, Hendrick, & Johnson, 2001a; Humphreys et al., 2016), it should not be surprising that inhibitory control underpins language processing. Again, the role of inhibitory control in children's language performance is mainly based on comprehension studies, investigation of production is scarce. For example, inhibitory control plays an important role in children's sentence comprehension, when the initial interpretation of a sentence needs to be inhibited in favour of a later alternative interpretation (i.e. garden-path sentence). Studies have found that, as compared to adults and adolescents, children experience greater difficulties in recovering from an initial misinterpretation (Lorsbach, Katz, & Cupak, 1998; Lorsbach & Reimer, 1997), and children's garden-path recovery ability is predicted by their performance on cognitive control tasks (Woodard et al., 2016). Moreover, counter to the claim that inhibition difficulties in children with low levels of comprehension can be accounted for by weaknesses in vocabulary (Henderson et al., 2013), it has also been claimed that inhibition deficits are central to comprehension difficulties (Borella, Carretti, & Pelegrina, 2010; Cain, 2006), and may mediate the relationship between WM and comprehension. A meta-analysis of the literature suggested that the evidence of WM and comprehension difficulties in poor comprehenders can be partly attributed to inefficiencies in inhibiting irrelevant linguistic information which has been activated and occupies WM resources (Carretti, Borella, Cornoldi, & De Beni, 2009).

In sum, previous literature has emphasized the importance of several linguistic and non-linguistic cognitive skills in sentence comprehension and production. Vocabulary knowledge very likely underpins both comprehension and production abilities; since it predicts both lexical and grammatical development in language acquisition (Fernald, Perfors, & Marchman, 2006; Hoff, Quinn, & Giguere, 2018). However, it is not clear whether linguistic experience would explain additional variance over above vocabulary knowledge, since the effects of these two skills are interrelated and many studies reporting significant influences of linguistic experience have not accounted for vocabulary effects. Similarly, the influences of other non-linguistic cognitive skills are also interrelated with vocabulary knowledge. The relationship between inhibition-related processes, WM and vocabulary in comprehension and production are far from clear. Is vocabulary all that matters such that it mediates the relationship between non-linguistic cognitive skills and language performance? Or do other skills contribute over above vocabulary knowledge? One possibility is that since children and adolescents are characterized by underdeveloped executive control compared to adults, they need to rely on other skills (e.g. vocabulary knowledge) to comprehend or plan cognitively demanding sentences. Thus, it is plausible that variations in children and adolescents' language performance with complex phrases may be less strongly associated with differences in inhibition control until such skills reach adult-like levels.

3.1.2 The present study

To shed light on these issues, we adopted the same methodology utilized in the adult study to examine the developmental changes in the relationship between cognitive skills and comprehension and production of complex phrases. Prior work indicates that large changes in executive functions such as inhibitory control are likely to take place in late childhood and mature levels are generally reached in adolescents (Diamond, 2013; Luna, Padmanabhan, & O'Hearn, 2010). Thus, in the present study, we contrasted children and adolescents' language performance from three different age groups to capture development changes of those processes: 8-10 years old (young children), 11-13 years old (old children) and 14-16 years old (adolescents). Briefly, our participants were tested with complex phrases which are known to induce semantic competition or not (animate vs. inanimate condition, e.g. the man/teddy bear that the girl is hugging) in adults

(Gennari & MacDonald, 2009; Gennari, Mirković, et al., 2012; Humphreys et al., 2016). They were also assessed on cognitive measures used in Study 1: two vocabulary measures (i.e., vocabulary depth and breadth), a children's author recognition test, and three inhibition measures: a measure of motor inhibition (STOP-IT task) and two measures of semantic inhibition (homonym and homograph tasks), see pages 48-50 for descriptions of these assessments. Given that WM is often suggested to play a significant role in children's comprehension performance, in the current study, we additionally included a measure of backward digit span to assess children's verbal WM capacity. In this test, participants were asked to remember lists of spoken digits and recall them in reverse order. This test was chosen over a forward digit span as it imposes demands on both storage and processing of verbal information, thus measures verbal WM more specifically rather than a phonological short-term memory (which only concerns the capacity component). The storage component of this test is to retain the digits, the processing component involves sorting lists of digits in reverse order. This test is also chosen over the conventional reading/listening span or word repetition task because it does not entail an inherent linguistic processing component, thus this measure reflects more specifically constraints in WM resources rather than variations in long-term linguistic knowledge and experience. And a significant relationship between this measure and language performance would imply cognitive resources in verbal WM is shared between processing linguistic and non-linguistic verbal information, as suggested by the single-resource verbal WM account (Just & Carpenter, 1992).

The current study had two primary aims. First, to investigate age-related changes (from childhood to adolescence) in comprehension and production of complex phrases, as extensive research has been conducted with adults, we know far less about the nature of different types of complex phrases in children and adolescents' language. Generally, we would expect to see age-related improvements in all types of complex phrases (animate or inanimate headed descriptions) for both comprehension and production tasks. We are also interested in whether there is age-related difference in the degree of semantic interference individuals may experience, which will be shown as an interaction effect between animacy condition and age group comparison.

Secondly, we aim to investigate whether any cognitive skill predicts children or adolescents' susceptibility to comprehension or production interference. Specifically, we are interested to see whether children or/and adolescents recruit the same cognitive skills as adults (i.e. vocabulary knowledge and semantic-phonological inhibition from the homograph task) to resolve semantic competition in comprehension and production. These findings could potentially have implications for informing theories of language development.

3.2 Experimental Methods

3.2.1 Participants

31 young children aged from 8-10 years old (24 females, mean age=9.34, *SD*=1.00; 7 males, mean age=9.19, *SD*=0.83), 49 old children aged from 11-13 years old (25 females, mean age=12.39, *SD*=0.85; 24 males, mean age=12.49, *SD*=0.80) and 32 adolescents aged from 14-16 years old (20 females, mean age=15.15, *SD*=0.87; 12 males, mean age=15.57, *SD*=0.66) were recruited from the wider community in the city of York. One children's comprehension performance and one children's production performance (both from the 11-13 age group) was not recorded due to program malfunctioning.

3.2.2 Materials and procedure

The tasks and procedures were identical to those described in the adult study, except that we reduced the number of trials, and used simpler picture and word stimuli in several tasks to shorten the study length and complexity (see below for more details). For the experiment, all participants completed both comprehension and production tasks, followed by 6 cognitive tasks, all administered individually in one hour session.

Comprehension and production tasks. We selected 20 experimental pictures and 20 fillers (see Appendix A) from the adult study which described scenes or actions more familiar to young children (e.g. carry books/ a baby). For both comprehension and production tasks, each participant was tested with 10 trials of animate descriptions and 10 trials of inanimate descriptions, as well as 3 practice trials before the main experimental block.

Backward digit span. This task is adopted from the Working Memory Test Battery for Children (WMTB-C, Pickering & Gathercole, 2001). On each trial, participants were required to recall a sequence of spoken digits in its reverse order, e.g. 6 1 3 9 5 2. Test trials begin with 3 digits and increase by one digit in each level (the levels ranged from 3 to 7 digits, with 6 trials in each level), and the task ends until the participant is unable to recall 4 correct trials at any given level.

Semantic inhibition tasks. We selected 11 homograph and 11 homonyms (out of 24 homographs and 24 homonyms used in the adult study) which have both meanings well known to young children (we selected words which have been used in existing children studies: Henderson et al., 2013; Hala, et al., 2007; Norbury, 2005). These words were paired with 6 unambiguous filler words for the homograph task, and 11 unambiguous words for the homonym task (see Appendix C).Participants also received 3 practice trials before each task. Given that most participants performed poorly on both tasks, with average accuracies around 23.03% in the homograph task and 62.04% in the homonym task, we only use accuracy data to represent their semantic/phonological inhibition performance for both tasks. This is because some participants' RTs were derived from only one or two valid responses, thus were not representative of their inhibition performance.

Author Recognition Test. We used a UK version of a Children's Author Recognition Test which was developed by Harlaar, Dale, & Plomin (2007). This version of the test included a list of 21 names of 'real' authors who wrote books for children and 21 foils. Each participant was asked to identify the names of real authors from the list.

The rest of the tasks used identical stimuli and procedure as in the adult study (i.e. expressive and receptive vocabulary, backward digit span, STOP-IT task). For all the individual difference measures, we present the number of participants, range, mean and standard deviation in Table 3.1, and the correlations among the measures in Table 3.2.

Table 3. 1 Descriptive statistics for individual differences variables by age groups

Measures	*N*	Range	Mean	*SD*
Expressive vocabulary				
8-10 years	31	47-75	59.48	7.83
11-13 years	49	45-70	56.24	6.15
14-16 years	32	43-69	54.22	5.99
Receptive vocabulary				
8-10 years	31	93-138	116.29	13.06
11-13 years	49	87-132	111.08	10.69
14-16 years	32	83-131	108.00	12.75
Reading				
8-10 years	31	0-15	6.52	3.23
11-13 years	49	1-12	7.65	2.96
14-16 years	32	4-16	8.66	3.42

WM				
8-10 years	31	12-29	19.45	4.19
11-13 years	49	13-36	22.35	5.37
14-16 years	32	16-33	23.03	4.53
STOP-IT				
8-10 years	28	175.6-591.8	315.49	93.78
11-13 years	47	208.0-430.1	306.35	52.32
14-16 years	31	202.3-342.1	266.34	38.86
Hmy Acc				
8-10 years	31	0.00-0.75	0.47	0.23
11-13 years	49	0.25-0.91	0.64	0.19
14-16 years	32	0.40-1.00	0.73	0.16
Hmg Acc				
8-10 years	31	0.00-0.40	0.16	0.12
11-13 years	49	0.00-0.80	0.24	0.25
14-16 years	32	0.00-0.80	0.31	0.22

Note: *Expressive vocabulary* stands for *WASI-II vocabulary subtest*, *Receptive vocabulary* stands for *PPTV*, *Reading* stands for *Children's Author Recognition Test*, *WM* stands for *Backward digit span*, *STOP* stands for *STOP-IT performance (SSRT), Hmy Acc* stands for *Homonym accuracy, Hmg* Acc stands for *Homograph accuracy,* respectively.

Table 3. 2 Correlations between individual difference measures by age groups (i.e., 8-10 years/11-13 years/14-16 years)

	Expressive vocabulary	Receptive vocabulary	Reading	WM	STOP-IT	Hmy Acc	Hmg Acc
Expressive vocabulary		0.405*/0.322*/0.538**	0.236/0.501**/0.654**	0.089/0.286*/0.076	-0.110/-0.323*/0.314	0.055/0.215/0.114	0.036/0.312*/0.109
Receptive vocabulary	0.405*/0.322*/0.538**		0.331/0.409**/0.637**	-0.302/0.141/0.128	-0.009/-0.094/0.169	0.156/0.373**/0.260	0.129/0.390**/0.046
Reading	0.236/0.501**/0.654**	0.331/0.409**/0.637**		0.209/0.437**/0.067	-0.162/-0.361*/0.240	0.063/0.204/0.204	0.379*/0.395**/-0.016
WM	0.089/0.286*/0.076	-0.302/0.141/0.128	0.209/0.437**/0.067		-0.188/-0.268/-0.290	0.085/0.379**/-0.242	-0.045/0.347*/-0.143
STOP-IT	-0.110/-0.323*/0.314	-0.009/-0.094/0.169	-0.162/-0.361*/0.240	-0.188/-0.268/-0.290		-0.204/-0.121/0.026	-0.062/-0.338*/0.202
Hmy Acc	0.055/0.215/0.114	0.156/0.373**/0.260	0.063/0.204/0.204	0.085/0.379**/-0.242	-0.204/-0.121/0.026		0.264/0.472**/0.154
Hmg Acc	0.036/0.312*/0.109	0.129/0.390**/0.046	0.379*/0.395**/-0.016	-0.045/0.347*/-0.143	-0.062/-0.338*/0.202	0.264/0.472**/0.154	

3.2.3 Data coding and analysis

The procedure of data coding was identical to those described in study 1 (see pages 50-53). For comprehension task, we report analyses with accuracy data and reaction times; for production, we report analyses with accuracy, passive/active utterance choice, agent omission and verbal fluency. To minimize the influence of outliers in analysing continuous DVs (i.e. comprehension RTs, production fluency), we first excluded extreme values (duration longer than 5000 ms) and then removed values deviated 3.5 *SD*s from condition mean by age group[1] (i.e., for animate and inanimate conditions separately). This resulted in the removal of around 1.62% of correct responses for comprehension RTs, and around 2.54% of correct passive descriptions for production fluency.

For analyses, we ran a series of LMEMs for comprehension and production performance separately. All models included maximal random-effects structures. We included *age group*, *animacy condition or reversibility rating* and the interaction between the two to investigate whether there was any developmental change in language performance. Reversibility rating of individual items, rather than high vs. low reversibility condition (as in study 1) was included as a fixed factor, mainly because our participants were only tested with 10 animate items, and this design does not have enough power to see statistical differences across reversibility conditions. In the age-based analyses, the fixed factor of age group was contrast-coded: a first contrast compared adolescents (coded as -2) with both younger (coded as 1) and older children (coded as 1), and a second contrast compared older children (coded as -1) with younger children (coded as

[1]The reason to not exclude outliers by individual means is because each participant only have up to 10 valid data points per condition, thus extreme outliers which greatly affects individual means cannot be identified using this exclusion criteria (in fact, all the original RTs fall within 2.5 SDs of individual means).

1). Any significant interactions were interpreted using post hoc comparisons with holm-adjusted tests (‘emmeans’ package in R, and function ‘emtrends ‘for interactions with covariates)

To examine the unique contribution of different cognitive skills, measures of individual difference (as well as interactions with animacy condition and age group) were entered in prior selected orders: vocabulary measures were entered first for the role of lexical knowledge, WM were entered secondly to account for the influence of storage capacity, reading experience and inhibition skills were entered last to investigate whether exposure to complex structures benefit children and adolescents’ language performance, and crucially, whether interference resolution ability explains additional variance after controlling for the effects of vocabulary (as for adults in study 1). At each step, we removed non-significant main effects or interactions, and then re-run the model until all remaining individual predictors demonstrated significant main effects or significant interactions with either animacy condition or age group.

3.3 Results

3.3.1 Accuracy of complex phrase tasks

The accuracy for children and adolescents was generally very high for both comprehension (animate target: M=96.94%, *SD*=0.06; inanimate target: M=97.66%, *SD*=0.05) and production tasks (animate target: M=94.91%, *SD*=0.08; inanimate target: M=88.39%, *SD*=0.13), so most of the trials were included for further analyses. Although the accuracy in production was significantly higher for animate-targeted phrases (Wilcoxon test: $z=5.0$, $p<0.01$), similar with our adult’s data, most errors were due to failing to inhibit the agent first tendency in production, resulting in describing the agent of the event rather than the patient/target entity.

Below we also examined whether there were any age differences in their accuracy performance. In general, each of the three age groups performed similarly in the comprehension task (see Table 3.3), but there was a general developmental improvement in their accuracy performance in production (see Table 3.4); adolescents were more accurate than children (p=0.03), and older children were more accurate than younger children (p<0.01). We do not report further individual difference analyses here due to convergence issues with production accuracy. Models with comprehension accuracy did not identify any relevant predictors.

Table 3. 3 Coefficient estimates of a LMEM predicting comprehension accuracy from age group and head-noun animacy

	Coefficient	SE	*t*-score	*p*-value
Intercept	4.25	0.46	9.21	< 0.01*
Animacy	-0.31	0.59	-0.53	0.60
Adolescents vs. children	-0.11	0.17	-0.63	0.53
Old children vs. young children	-0.14	0.26	-0.53	0.60
Adolescents vs. children*Animacy	0.22	0.22	1.02	0.31
Old children vs. young children*Animacy	0.20	0.35	0.57	0.57

Table 3. 4 Coefficient estimates of a LMEM predicting production accuracy from age group and head-noun animacy

	Coefficient	SE	*t*-score	*p*-value
Intercept	2.83	0.36	7.87	< 0.01*
Animacy	0.74	0.44	1.68	0.09
Adolescents vs. children	-0.23	0.11	-2.16	0.03*
Old children vs. young children	-0.47	0.16	-2.95	<0.01*
Adolescents vs. children*Animacy	-0.06	0.15	-0.37	0.71
Old children vs. young children*Animacy	0.11	0.20	0.55	0.58

3.3.2 Comprehension RT

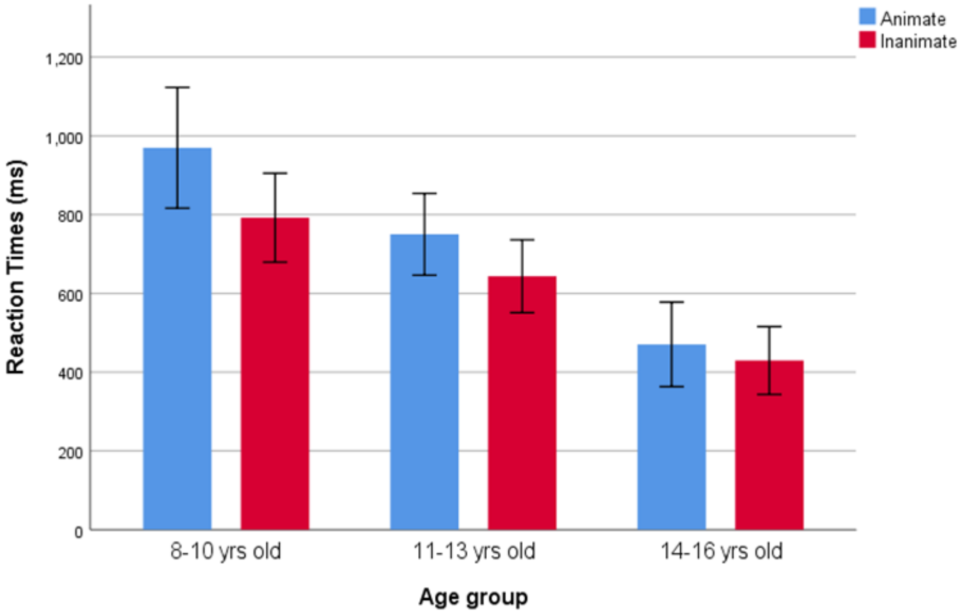


Figure 3. 1 Mean comprehension RTs (with standard error bars) for each animacy condition by age group.

Animacy effects. The first model examined the animacy effect across age-groups to establish whether the difficulty associated with the animacy manipulation varied with age (see Table 3.5). There was a marginally significant main effect of animacy condition (p=0.06), with processing of animate-targeted phrases slower than inanimate-targeted phrases. There was also a significant main effect of age, with adolescents responding significantly faster than children (p<0.01), and old children responding faster than young children (p=0.02). The animacy*age interaction was only significant in the comparison between adolescents and children, with children demonstrating greater difference between animacy conditions when compared to adolescents (p=0.02); and there was no significant difference in animacy interference between young and old children (p=0.50).

We followed up this interaction with post hoc analysis and found both children groups took longer to comprehend animate-targeted phrases as compared to inanimate-targeted phrases (younger children: z=-2.006, p=0.05; older children: z=-1.934, *p*=0.06), whereas adolescents did not (z=-0.642, *p*=0.52). This null finding in adolescents contradicts with previous findings with adults in study 1. We suspect this could be due to the fact that adolescents found the task easy[2] and there were relative few items in each condition, as we selected the simplest 10 items from the stimuli to adapt the level of task difficulty to younger children. Moreover, half of the animate-head items were of low-reversibility. So, the study may also not have enough power to see statistical differences across animacy conditions, even though there were numerical differences.

Table 3. 5 Coefficient estimates of a LMEM predicting comprehension RTs (ms) from age group and head-noun animacy

	Coefficient	SE	*t*-score	*p*-value
Intercept	622.08	42.75	14.55	< 0.01*
Animacy	94.83	46.78	2.03	0.06
Adolescents vs. children	93.53	20.30	4.61	0.00*
Old children vs. young children	82.31	33.31	2.47	0.02*
Adolescents vs. children*Animacy	32.30	13.80	2.34	0.02*
Old children vs. young children*Animacy	15.23	22.64	0.67	0.50

[2]On average, our adolescents' RTs were 470.46 (*SD*=303.97) for animate condition, and 429.45 (*SD*=243.46) for inanimate condition. And their performance was even faster than adults' RTs, which were 574.90 (*SD*=400.12) for animate condition, and 446.72 (*SD*=358.60) for inanimate condition.

Reversibility effects. Despite the overall increased difficulty with processing animate-targeted phrases, we also investigated whether children and adolescents are sensitive to the degree of agent-patient role competition (assessed by the reversibility rating collected in study 1) within the animate condition. Thus, we run a LMEM including reversibility rating and age group as the predictors, even though there were fewer items in this data set. The results in table 3.6 reported a significant main effect of reversibility rating (*p*=0.03) and non-significant interactions between the rating and any age-group comparison[3], suggesting that all our participants' performance were affected by the reversibility manipulation: the more likely the two animate nouns compete for the agent role, the harder they are to be comprehended within active structures.

Table 3. 6 Coefficient estimates of a LMEM predicting comprehension RTs (ms) from age group and reversibility ratings

	Coefficient	SE	*t*-score	*p*-value
Intercept	801.02	54.52	14.69	< 0.01*
Reversibility	-39.80	16.27	-2.45	0.03*
Adolescents vs. children	145.18	26.55	5.47	0.00*
Old children vs. young children	70.79	43.48	1.63	0.11
Adolescents vs. children *Reversibility	-9.57	5.53	-1.73	0.08
Old children vs. young children*Reversibility	11.92	9.04	1.32	0.19

[3] Although the interaction with the age comparison between adolescents and children was marginal significant (*p*=0.08), post hoc analysis revealed all age groups did not differ significantly in their relationships with reversibility ratings: young children vs. old children (*t*=1.224, *p*=0.45); young children vs. adolescents (*t*=-0.418, *p*=0.68); old children vs. adolescents (*t*=-1.703, *p*=0.28).

Relations with individual differences. We next examined the influence of individual difference in predicting general comprehension speed or susceptibility to comprehension interference (i.e., an interaction with animacy condition). The final model is presented in Table 3.7. In general, there was a main effect of homograph accuracy (p=0.01) and digit span (p=0.01), in which participants with better semantic inhibition and WM span were faster in processing complex phrases. Moreover, the influence of both predictors (homograph accuracy: p=0.01; digit span: p=0.03) interacts with the age-group comparison between young and older children, but not with animacy condition. To further investigate this interaction, we then refer to the separate by age-group models including individual predictors (see table 3.8), and found that the RTs of young children were predicted by homograph accuracy and digit span (see figure 3.2), whereas old children and adolescents' RTs were not. Notably, the main effect of the age comparison between young and old children became non-significant (p=0.93), after the influence of homograph accuracy and digit span were added. This may suggest that semantic inhibition and working memory were driving the developmental improvement in processing complex phrases overall.

In sum, children's comprehension RT was affected by the degree of agent-patient competition between the noun concepts, resulting in significant animacy effect and significant correlation with the reversibility rating. This pattern mirrors the adult findings from study 1. Not surprisingly, there was general developmental improvement in comprehension performance, such that complex phrases were processed faster by older participants. However, the degree of animacy-based competition did not change with age in children, and was not explained by any of the measured cognitive abilities. Nevertheless, we found that semantic inhibition and working memory capacity significantly predicted comprehension RTs for young children. This may account for developmental improvements in children, such that old children comprehend complex phrases more efficiently than young children because they developed better semantic inhibition and memory capacity, as they grow older.

Table 3. 7 Coefficient estimates of a LMEM predicting comprehension RTs (ms) from age group, head-noun animacy and individual difference measures

	Coefficient	SE	*t*-score	*p*-value
Intercept	574.96	45.45	12.65	<.01*
Animacy	94.61	46.52	2.03	0.06
Adolescents vs. children	55.46	22.05	2.52	0.01*
Older children vs. young children	-3.48	38.62	-0.09	0.93
Homograph	-461.93	168.66	-2.74	0.01*
WM span	-15.89	6.03	-2.64	0.01*
Adolescents vs. children*Animacy	32.25	13.81	2.34	0.02*
Old children vs. young children*Animacy	15.00	22.65	0.66	0.51
Adolescents vs. children*Homograph	-173.09	106.76	-1.62	0.11
Old children vs. young children*Homograph	-636.10	226.13	-2.81	0.01*
Adolescents vs. children*WM span	-0.07	4.37	-0.02	0.99
Old children vs. young children*WM span	-15.34	7.17	-2.14	0.03*

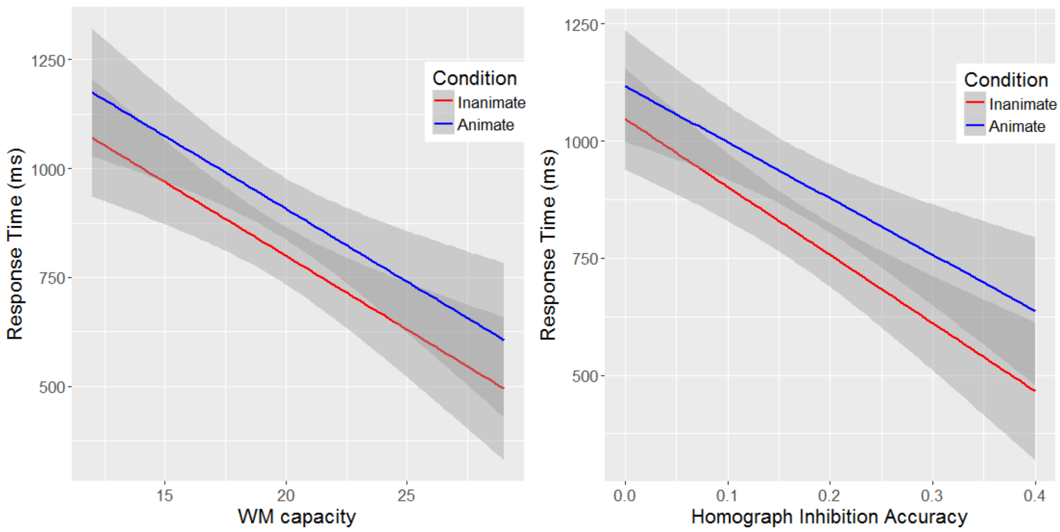


Figure 3. 2 WM capacity (left panel) and Homograph Inhibition (right panel) predict young children's comprehension RTs (ms). Shading indicates standard error.

Table 3. 8 Separate by age-group models of the effects of head-noun animacy and cognitive predictors in predicting comprehension RTs (ms)

Young children (8-10 yrs old):

	Coefficient	SE	*t*-score	*p*-value
Intercept	796.32	52.47	15.18	< 0.01*
Animacy	142.78	65.36	2.18	0.04*
Homograph Accuracy	-1258.58	385.16	-3.27	< 0.01*
WM span	-31.11	10.99	-2.83	0.01*

Old children (11-13 yrs):

	Coefficient	SE	*t*-score	*p*-value
Intercept	634.05	53.74	11.80	< 0.01*
Animacy	110.55	48.64	2.27	0.03*

Adolescents (14-16 yrs):

	Coefficient	SE	*t*-score	*p*-value
Intercept	434.25	55.33	7.85	<0 .01*
Animacy	34.07	45.03	0.76	0.46

3.3.3 Production task

3.3.3.1 Utterance structures

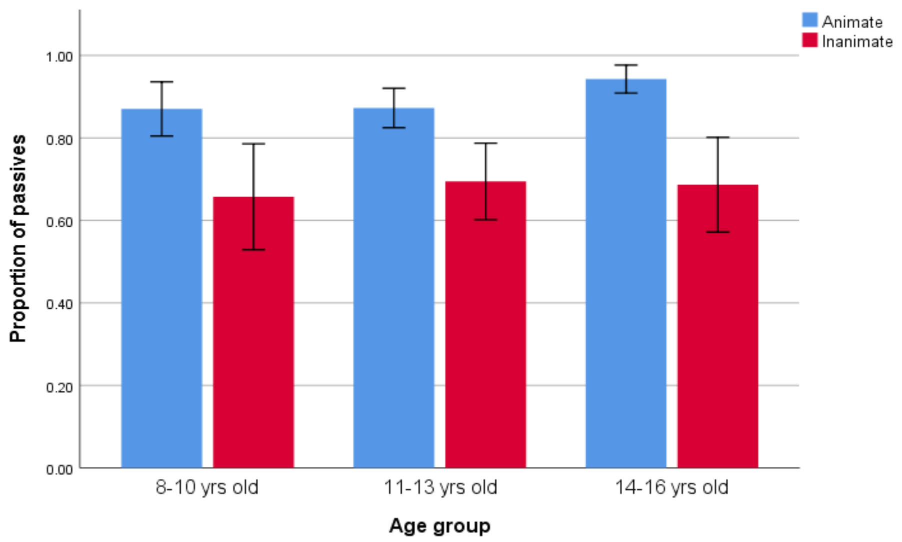


Figure 3. 3 Average proportion of passives produced (with standard error bars) for each animacy condition by age group.

Animacy effects. As shown in table 3.9, sentence structure was significantly predicted by head-noun animacy (p<0.01): participants produced more passives in response to animate-target pictures as compared to inanimate-target pictures. There was also a significant interaction between animacy condition and age group for adolescents and children (p=0.02), such that the animacy effect was greater for adolescents as compared to the two children groups. This is because our adolescents group produced more passives in responses to animate targets as compared to the children groups, and their production frequency for animate targets is comparable to the numbers reported in many previous studies using adult participants (Gennari, Mirković, et al., 2012; Jessica L Montag & MacDonald, 2015). As shown in figure 3.3, adolescents produced 94.27% (*SD*=0.10) passives when describing animate targets, and 68.67% (*SD*=0.32) passives when describing inanimate targets. Old and young children produced, respectively, 87.24% (*SD*=0.17) and 87.03% (*SD*=0.18) passives when describing animate entities, and 69.45% (*SD*=0.32) and 65.74% (SD=0.36) when describing inanimate entities.

Table 3. 9 Coefficient estimates of a LMEM predicting sentence structure (active/passive) from age group and head-noun animacy

	Coefficient	SE	*z*-score	*p*-value
Intercept	1.58	0.34	4.65	<0.01*
Animacy	1.73	0.31	5.59	<0.01*
Adolescents vs. children	-0.02	0.20	-0.11	0.91
Old children vs. young children	-0.23	0.33	-0.71	0.48
Adolescents vs. children *Animacy	-0.33	0.15	-2.30	0.02*
Old children vs. young children *Animacy	0.23	0.22	1.04	0.30

Reversibility effects. The results in table 3.10 reported a significant main effect of reversibility rating (p=0.01) and non-significant interactions between the rating and any age-group comparison. This suggests all age groups' structure preference are related to the degree of semantic competition between the noun concepts, and they did not differ significantly in the size of reversibility effects. We chose not to perform analyses of individual difference for choice of utterance type, given that the models failed to converge when inhibition predictors were entered.

Table 3. 10 Coefficient estimates of a LMEM predicting sentence structure (active/passive) from age group and reversibility ratings

	Coefficient	SE	*z*-score	*p*-value
Intercept	3.88	0.47	8.18	<0.01
Reversibility	-0.32	0.13	-2.45	0.01*
Adolescents vs. children	-0.53	0.23	-2.26	0.02*
Old children vs. young children	0.28	0.30	0.95	0.34
Adolescents vs. children *Reversibility	0.05	0.06	0.81	0.42
Old children vs. young children *Reversibility	-0.11	0.07	-1.49	0.14

3.3.3.2 Agent omission

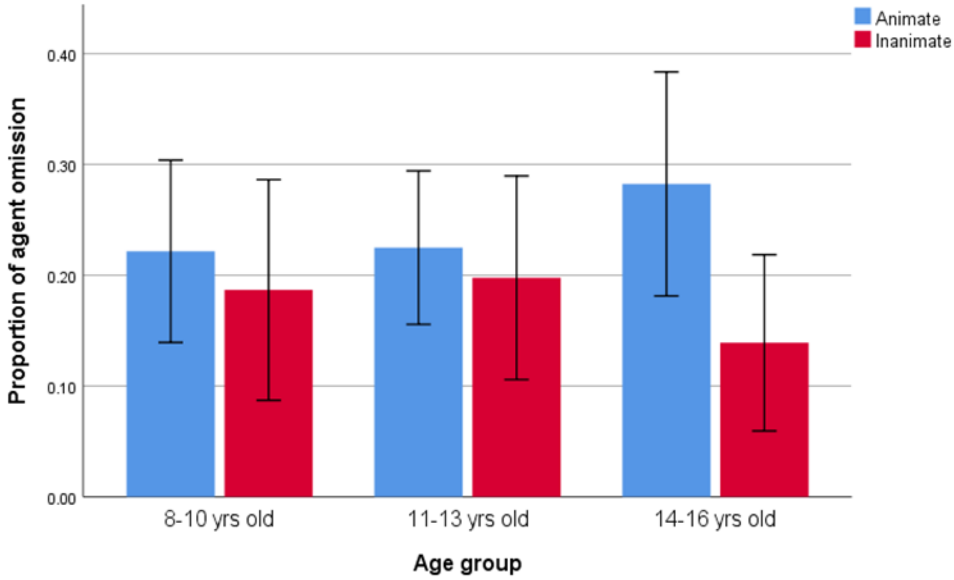


Figure 3. 4 Average proportion of agent omission produced within passive utterances (with standard error bars) for each animacy condition by age group.

Animacy effects. We now turn to analyses of one fine-grained detail of passive utterances: agent omission. The proportions of agent omission by animacy condition is shown in figure 3.4. We were unable to perform analyses of reversibility effects with individual difference predictors for this coding, because the occurrence of agent omission was not very high within passive utterance and each participant only produced 10 descriptions per condition, which has left us with not enough data to obtain reliable results.

Table 3.11 summarize a model including animacy condition and age groups in predicting agent omission. In general, the marginal significant animacy effect (p=0.08) and marginal significant interaction between animacy condition and the age comparison between adolescents and children (p=0.09) were driven by a significant animacy effect in the adolescents group (z=-3.061, p<0.01), but less significant results in the two children groups (young children: z=-0.952, *p*=0.34; old children: *z*=-1.762, *p*=0.08). This result may reflect an age-related difference in abilities to use certain production options to reduce planning burden. This is because agent omission is often considered as a decision that speakers make online to reduce planning competition, especially in circumstances where two similar nouns (e.g. two animate nouns) need to be planned in sequence. Our results may suggest that children lack experience or knowledge with this aspect of language flexibility, and they may have thought that as in the comprehension task, all the noun elements needed to be mentioned to provide a complete descriptions. Thus, they did not drop agents as strategically as adolescents, who omitted agents more often when planning of animate nouns led to greater competition.

Table 3. 11 Coefficient estimates of a LMEM predicting agent omission from age group and head-noun animacy

	Coefficient	SE	*z*-score	*p*-value
Intercept	-3.56	0.59	-6.00	<0.01*
Animacy	1.19	0.68	1.74	0.08
Adolescents vs. children	0.09	0.23	0.38	0.71
Old children vs. young children	0.02	0.37	0.06	0.95
Adolescents vs. children*Animacy	-0.30	0.17	-1.72	0.09
Old children vs. young children *Animacy	-0.10	0.28	-0.37	0.71

3.3.3.3 Verbal fluency

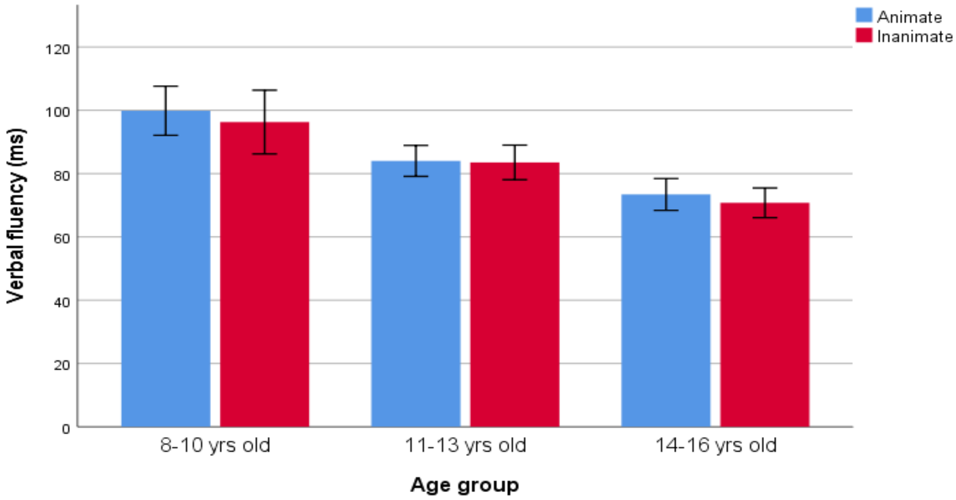


Figure 3. 5 Mean verbal fluency (with standard error bars) for each animacy condition by age group.

Animacy and reversibility effects. As shown in table 3.12, there was a significant age effect, such that older speakers produced more fluent descriptions than younger participants (adolescents vs. children: p<0.01; older children vs. young children: p=0.02). Also, the main effect of animacy condition was not significant (p=0.36), but the interaction between age group and animacy was marginally significant in the comparison between old and young children (p=0.07), with young children elicited greater difference between conditions as compared to older children. Indeed, the post hoc analysis revealed that the younger children were the only group to elicit a significant animacy effect in production fluency (t=-2.129, p=0.04). This might suggest that as compared to older participants, younger children were more vulnerable to the presence of production interference in animate condition.

For the analyses of reversibility effects, we found no significant main effect of reversibility ratings in verbal fluency, and non-significant interactions with any age-group comparisons. Again, this null result is likely due to the lack of power in our design to obtain reliable results.

Table 3. 12 Coefficient estimates from a mixed-effects model predicting verbal fluency (ms) from age group and head-noun animacy

	Coefficient	SE	*t*-score	*p*-value
Intercept	82.35	2.45	33.61	<0.01*
Adolescents vs. children	4.95	1.17	4.22	<0.01*
Old children vs. young children	4.67	1.98	2.36	0.02*
Animacy	2.49	2.14	1.16	0.26
Adolescents vs. children *Animacy	1.15	0.91	1.26	0.21
Old children vs. young children *Animacy	2.93	1.58	1.85	0.07

Relations with individual differences. The model in table 3.13 revealed significant influences of 3 individual predictors on verbal fluency: expressive vocabulary and STOP-IT (motor inhibition) and WM span. First, the effect of expressive vocabulary predicted production fluency for both conditions (p=0.04), such that speakers with better vocabulary knowledge produced more fluent descriptions in general. However, when running separate models for each age group, vocabulary only contributed to adolescents' production performance (p<0.01, see table 3.14 and figure 3.6). Also, although receptive vocabulary was not identified as a significant predictor in the model including all participants, it did predict adolescents' production fluency as equally as expressive vocabulary, given that models including either predictor elicited very similar AIC scores (AIC for the model including expressive vocabulary is 4108.6, and including receptive vocabulary is 4110.0).

On the other hand, the effect of STOP-IT performance (p<0.01) would suggest that speakers with better motor inhibition produced more disfluent descriptions for both conditions. Follow-ups by age-group analyses revealed that this influence was driven by a significant main effect of STOP-IT measures in young children only (p=0.03). Although STOP-IT performance was also identified as an important predictor for adults' production fluency in study 1, it is not clear why we found the opposite pattern here, as an increase in inhibition skills seems to hinder production performance. We suspect that some of this negative relationship was driven by outliers, as there were two children who performed poorly on the STOP-IT task as compared to others (with SSRT of 591.8 ms and 544.4 ms; the average SSRT for the rest of the group is 300.66 ms, *SD*=59.53 ms). Indeed, after excluding these two children's data, the main effect of STOP-IT performance became non-significant (p=0.11). Moreover, the difference in general verbal fluency between individuals does not necessarily reflect the difference in the level of production difficulty they might encounter, because individuals are likely to vary in their habitual speaking rate (e.g. individuals who produced longer fluency measures might speak in a slow fashion, rather than being disfluent). This is why the critical prediction in fluency should be an interaction with animacy condition. In the absence of an interaction, the general correlation is difficult to interpret, and this also applies to the above correlation with vocabulary measures.

Finally, there was a significant three way interaction between animacy, WM span and the comparison between old and young children (p<0.01). Follow-up analyses by age group suggest that verbal fluency for young children (but not for older children) were marginally influenced by an interaction between animacy and WM span (p=0.06), such that young children with better WM capacity exhibited less difference between conditions (see figure 3.7).

To sum up the production results, children and adolescents' tendency to produce passives were influenced by the degree of agent-patient competition between noun concepts, resulting in significant animacy effects and correlations with reversibility ratings (demonstrating a similar pattern to adults). There were also age-related differences between the three groups. Specifically, there was a developmental improvement in verbal fluency and some age-related changes in the animacy effect: young children, unlike older participants, did not produce more agentless passives in animate condition as compared to inanimate condition, and were the only group to show a significant animacy effect in verbal fluency. This may imply that young children are less aware of or less skilled in taking advantage of agent dropping to ease planning of competitive concepts, and are more vulnerable to the presence of interference in production. Moreover, our individual difference analyses revealed significant influence of vocabulary knowledge and WM capacity in predicting production fluency, the former predicts general performance in adolescents, irrespective of animacy condition. The latter contributes to the animacy effect in young children, suggesting that young children with better WM capacity were less susceptible to production interference.

Table 3. 13 Coefficient estimates from a mixed-effects model predicting production fluency (ms) from age group, head-noun animacy and individual difference measures

	Coefficient	SE	*t*-score	*p*-value
Intercept	81.23	2.48	32.79	< 0.01*
Animacy	2.88	2.28	1.26	0.22
Adolescents vs. children	5.63	1.26	4.48	<0.01*
Older children vs. young children	4.53	2.08	2.18	0.03*
Expressive vocabulary	-0.48	0.23	-2.12	0.04
STOP-IT	-0.07	0.02	-3.02	<0.01*
WM span	-0.18	0.37	-0.48	0.63
Adolescents vs. children*Animacy	1.41	0.96	1.46	0.14
Older children vs. young children*Animacy	3.41	1.72	1.98	0.048*
Animacy * WM span	-0.30	0.29	-1.04	0.30
Adolescents vs. children* WM span	0.07	0.25	0.29	0.77
Older children vs. young children* WM span	0.37	0.43	0.87	0.39
Animacy*Adolescents vs. children* WM span	-0.16	0.20	-0.76	0.45
Animacy*Older children vs. young children* WM span	-1.03	0.36	-2.89	<0.01*

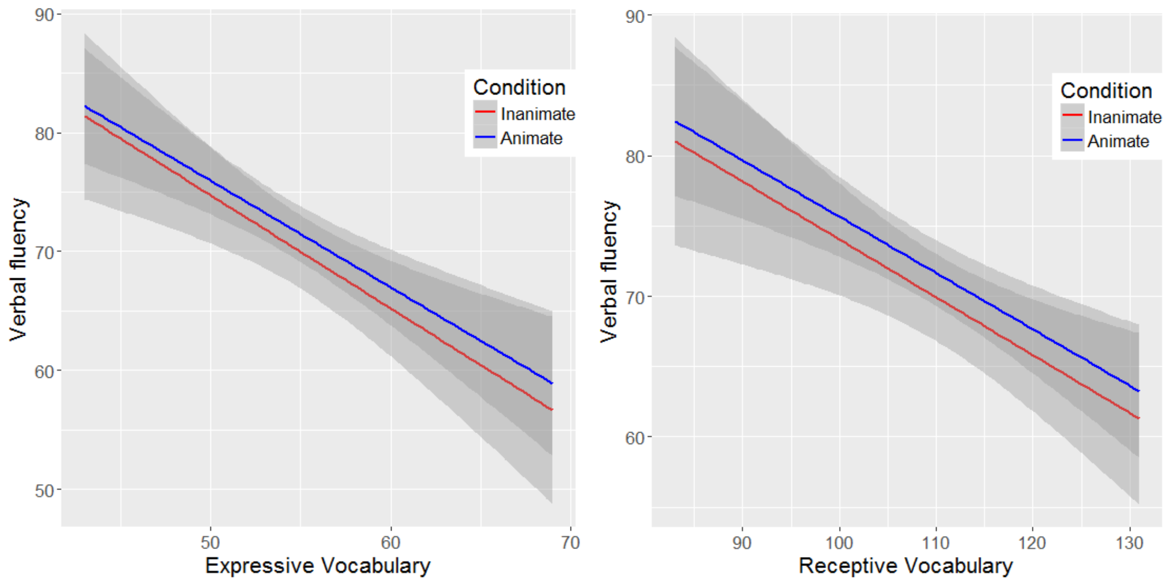


Figure 3. 6 An Expressive vocabulary (left panel) or Receptive vocabulary measure (right panel) predicts adolescents' verbal fluency (ms). Shading indicates standard error.

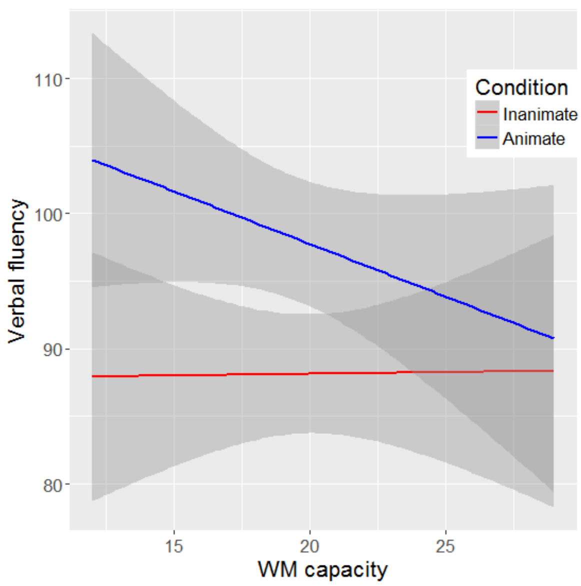


Figure 3. 7 Interaction between Animacy and a measure of WM capacity in predicting young children's verbal fluency (ms). Shading indicates standard error.

Table 3. 14 Separate by age-group models of the effects of head-noun animacy and cognitive factors in predicting verbal fluency (ms)

Younger children (8-10 yrs old):

	Coefficient	SE	*t*-score	*p*-value
Intercept	86.61	3.43	25.28	<0.01*
Animacy	12.47	3.97	3.14	<0.01*
STOP-IT	-0.08	0.04	-2.32	0.03*
WM span	0.07	0.81	0.09	0.93
WM span*Animacy	-1.38	0.74	-1.87	0.06

Older children (11-13 yrs):

	Coefficient	SE	*t*-score	*p*-value
Intercept	82.98	3.49	23.73	<0.01*
Animacy	0.77	2.48	0.31	0.76

Adolescents (14-16 yrs):

	Coefficient	SE	*t*-score	*p*-value
Intercept	72.38	2.80	25.84	< 0.01*
Animacy	0.12	2.77	0.04	0.97
Expressive vocabulary	-0.97	0.30	-3.22	<0.01*

	Coefficient	SE	*t*-score	*p*-value
Intercept	72.42	2.83	25.60	< 0.01*
Animacy	0.21	2.79	0.07	0.94
Receptive vocabulary	-0.43	0.15	-2.94	0.01*

3.4 Discussion

3.4.1 Developmental patterns in comprehension and production of complex phrases

Table 3. 15 Summary of results

Comprehension	**Main effects**	**Predictors**	**Interaction**
RT	Age Animacy	WM, Hmg Acc in young children	
	Reversibility		
Production	**Main effects**	**Predictors**	**Interaction**
Acc	Age		
Passives	Animacy		Animacy*Age
	Reversibility		
Ag Om	Animacy		Animacy*Age
Fluency	Age	Vocabulary in adolescents	Animacy*Age Animacy*WM in young children

This study was designed to reveal comprehension and production patterns with complex phrases in children and adolescents, and to examine the role of individual difference predictors in explaining their language performance beyond age. In general, there were developmental improvements in both comprehension and production performance, such that complex phrases were processed faster by older children and adolescents, and older participants produced more accurate and more fluent descriptions. There was also some evidence showing age-related differences in animacy effects, specifically the comparison between young children and adolescents in their production performance. For example, it was found that young children were less skilled in using certain strategies to reduce production interference as compared to adolescents, such as planning a passive structure or omit the agent when describing animate-targeted pictures. This is convergent with the direction of the developmental pattern identified by Montag & MacDonald (2015) in a comparison between 8 and 12 year olds' utterance choices. Although our young (8-10 yrs) and old children groups (11-13 yrs) did not differ significantly in the rates of passive utterances and agentless passives produced across conditions, this may be a reflection of using a wider age range for children groups in our study (as the number of participants recruited for each age group in our study is comparable to Montag & MacDonald (2015)'s sample). Another age-related difference in animacy effect was found in verbal fluency where young children were the only age group who produced more disfluent descriptions for animate-targeted pictures. This suggests that, as compared to older speakers, they are more affected by the presence of production interference in animate condition.

Similar to our analyses with adults' data, we also examined within animate items to investigate whether children and adolescents' language performance were also explained by specific semantic features of the noun phrases, rather than only categorical animacy. Our participants' performance in terms of comprehension RTs and production choices are consistent with what has been reported with adults in study 1, such that all age groups' performance for animate items were similarly predicted by the degree of agent-patient competition between the noun concepts, as more competition leads to greater difficulty in performance. However, we only tested children and adolescents' performance on 20 of the 42 animate items used in our adult study, and the selected items do not include many cases having low reversibility ratings (see appendix, i.e. high competition between involved nouns, such as "the player that the other player is hitting"). Thus, the non-significant correlations with reversibility ratings in children and adolescents' production fluency may be due to a lack of variability in difficulty among the items and lack of statistical power, and future studies should ideally include more items to capture a wider varying degree of semantic competition in production.

Importantly, this study is the first to compare children and adolescents' comprehension and production of the same type of complex phrases and to examine production performance to both preferred structures and verbal fluency. The above age group analysis results add to the growing body of evidence suggesting a reversibility-based semantic competition underlying comprehension and production processes, and this may manifest at an early age.

3.4.2 Relations with individual differences

Turning now to the role of individual differences, we found that 8-10 yrs olds' comprehension and production performance were associated with the same domains of cognitive skills. In particular, young children's processing speed of both animate and inanimate-head descriptions were predicted by WM span and semantic inhibition, whereas their production fluency was explained by motor inhibition and an interaction between WM and animacy condition. This is consistent with other studies that have found that variability in aspects such as WM or executive function is associated with different comprehension outcomes (WM: e.g. Montgomery, Magimairaj, & O'Malley, 2008; EF: e.g. Woodard et al., 2016). Also, it is mostly consistent with our findings in adults in study 1, as semantic inhibition is associated with comprehension of complex phrases whereas motor inhibition only contributes to production performance. However, the role of motor inhibition in young children's production fluency remains unclear at this stage, as the relationship is on the opposite direction to that predicted and is likely being driven by extreme value in motor inhibition task performance. Thus, future studies could include more children to capture a wider range of performance level in motor inhibition tasks to further explore its contribution to the development of production ability.

Moreover, WM capacity appears to be the only cognitive predictor that explains animacy effect, as 8-10 years with better WM capacity demonstrated less difference in verbal fluency between animacy conditions. Although this relationship is only marginally significant (p=0.06), this seems to suggest that young children's sensitivity to production interference is influenced by whether their memory capacity is sufficient to cope with additional demands with planning competitive concepts. This is because, same as in comprehension, production has inherent WM demands. Information regarding individual concepts or utterance plans must be maintained in

WM until being outputted, and similar representations (e.g. phonological, semantic similar) tend to interfere with one another, thus affecting production efficiency (Acheson & MacDonald, 2009; Smith & Wheeldon, 2004). Given that there is scarce investigation of sentence production in children, this result adds to the existing literature in showing the role of WM capacity in modulating production competition.

Note that although we found WM capacity underpins both comprehension and production processes, it does not predict the size of animacy effect in young children's comprehension RTs. This maybe because the type of complex phrases tested in the comprehension task (active phrases, e.g. *the teddy bear/man that the girl is hugging*) is generally difficult for children regardless of the animacy properties of the heads, considering its high syntactic complexity and low frequency in children's language input (Montag & MacDonald, 2015). One possibility is that, active phrase is particularly difficult for children as it contains more distant structural relations: the verb and the head noun which need to be analysed together are separated by a second noun; which yields a higher integration cost and taps WM resources (Gibson, 1998). Thus, WM capacity were found to predict the general processing speed of active phrases in young children.

Together, it appears that young children's comprehension and production of complex phrases rely largely on variations in WM capacity, and this is consistent with several studies which observed an independent influence of WM capacity in children's comprehension performance using a digit-span task (Blything & Cain, 2016; Blything, Davies, & Cain, 2015). Our data seems to suggest that the resource-constraint characteristic of the verbal WM affects language performance with non-canonical structures, possibly relating to the processes to maintain relevant information and integrating/planning verbs (the processes involved in both comprehension and production tasks). Also, this is most consistent with the single resource verbal WM account (Just & Carpenter, 1992): there is a pool of domain-specific verbal WM for processing both linguistic and non-linguistic verbally mediated information, as inferred by the nature of the WM task used in the current study. The backward digit span task assesses the storage and processing of verbal information that contains relatively low semantic demands, as compared to other commonly used WM measures, e.g. the reading or listening span which has explicit linguistic processing requirements. Given that we only included one type of verbal WM measure, our findings cannot determine which memory resources are more relevant to complex phrase processing. For example, an alternative verbal WM account would suggest that the WM effect should reflect children's long-term linguistic knowledge rather than only resource limits (MacDonald & Christiansen, 2002; Acheson & MacDonald, 2009; Wells et al., 2009). Many adult studies have found that linguistic WM measures are better predictors of sentence comprehension than WM measures tapping capacity-limited processes with little semantic information (Daneman & Merikle, 1996; Shah & Miyake, 1996). It is possible that resource constraint might be more important for children than adults, because general processing capacity may play a role at early language acquisition stages, but not once individuals have mastery over complex structures. At this time language performance depends on retrieval of long-term linguistic knowledge and the ability to represent this information in WM. These processes are mediated by language knowledge and experience, as assessed in many linguistic WM tasks at word or sentence level. Given that young children's language performance was not predicted by measures on vocabulary knowledge and reading experience, our data seems to be consistent with this prediction. Nevertheless, future research should include more complex measures of verbal WM to provide a more accurate assessment and to better understand the role of WM in sentence comprehension and production.

Finally, we found that adolescents' production fluency was best explained by vocabulary rather than other cognitive factors. But fluency did not differ across animacy conditions, so this general relation does not depend on condition. This suggests that vocabulary cannot be explained in terms of the quality of individual's lexical representation as other adults studies have argued (Van Dyke et al., 2014). Rather it points to a role of lexical retrieval ability in production. The absence of such relationship with comprehension times may be because the use of child-appropriate stimuli with adolescents has resulted in good performance and left little variance to explain. Again, this is mostly consistent with our adult results in study 1 in suggesting asymmetry between comprehension and production in terms of task-difficulty, as production of transitive events is more cognitively demanding than comprehension of the same events.

3.4.3. Conclusion

The present study is the first study to date to investigate the comprehension and production patterns of complex phrases in children and adolescents, using an individual difference approach. Complex phrases which induce semantic competition and load WM (especially in actives), take time to learn, so it is possible that young children who are less skilled in using these structures, might display systematically different performance patterns than older participants. In our study, we found that children as young as 8-10 yrs old are already sensitive to the degree of semantic competition occurring in animate-head phrases. However, there is evidence for age-related improvement in both comprehension and production performance, as well as individual's sensitivity to semantic interference. Young children in the present study were still far from adult-like in their performance, despite being generally slow and less accurate, they are also less skilled in using passive structures and agent omissions strategically to ameliorate planning interference, i.e. they did not omit agents more often for animate-targeted items, and produced fewer passives as compared to older speakers.

More importantly, our results revealed that the only cognitive predictor recruited by both comprehension and production processes in young children is WM capacity. This is consistent with the capacity-constraint account, and also provides important evidence for the relationship between WM capacity and sentence-production abilities, given that most studies only addressed this relationship in reading comprehension. Also, lack of other predictions (such as vocabulary and inhibition skills) to explain condition differences may imply that unlike adults, children and adolescents are unable to utilize these skills to resolve semantic competition in complex phrase. It might be that they possess comparatively underdeveloped cognitive skills, especially inhibition skills, thus theses skills are less likely to be engaged during online language processing. However, it is also possible that we do not have sufficient statistical power to observe similar effects to those of adults, as each group was relatively small compared to our adult sample. Thus, we emphasize the need for future studies to increase the sample size and use more items (especially animate-targeted items) to capture a wider range of difficulty levels and also increase statistical power to allow firm conclusions to be drawn regarding to individual difference analyses.

Chapter 4

Reversibility-based competition in production of complex phrases: evidence from eye-movements

4.1 Introduction

Study 1 and 2 provided evidence for the presence of common competition processes underpinning complex phrase comprehension and production across a wide age range, from childhood to adulthood. Specifically, noun reversibility, likely arising from competition between alternative syntactic roles of the noun phrases moderated comprehension and production difficulty, and participants' susceptibility to this competition varies as a function of individual skills. For adults, individual's ability to resolve this competition is associated with inhibition skills over above vocabulary knowledge; whereas children's language performance is predicted by available memory resources. This may reflect that the relative importance of different cognitive skills underpinning language processes tend to change with development.

These results are consistent with previous findings suggesting that the semantic interference between similar noun phrases (moderated by noun-animacy, similarity, reversibility/thematic fit between nouns and verbs, etc.) operates at the verb position in comprehension (Gennari & MacDonald, 2008; Gordon et al., 2006). For example, one eye-tracking reading study has reported that similarity-based interference manifests in sentence processing during verb encoding, where the grammatical relationship between the nouns and the verb must be decided. Also, this interference is only observed under circumstances where two similar nouns must be held together in WM before either one can be integrated with the verb (e.g. "The *banker* that the *barber* praised climbed the mountain just outside of town"; Gordon et al., 2006). This seems to suggest that the similarity or noun reversibility engenders competition in comprehension owning to activation of alternative noun-verb interpretation when verbs are processed.

However, there is scarce investigation on the actual time course of semantic interference in sentence production and its association with animacy or reversibility, particularly when the verbs are planned. This may be due to the time course of sentence production being more difficult to investigate as compared to comprehension processes in eye tracking studies. As discussed in chapter 1, the nature of the production task is very different from comprehension task. Comprehension involves making predictions based on current input and evidence of interference can be observed when these predictions conflict with each other or with the upcoming input. Thus, the time course of ambiguity resolution in comprehension potentially depends on where the ambiguous information and the disambiguating information is placed in a given sentence. For this reason, researchers often manipulate the sentence structures to change the positions of these elements.

In production, however, speakers need to engage some degree of advanced planning to develop a presumably unambiguous representation of the message they wish to convey before utterance (Levelt, 1993). For example, when asked to describe a character, as in "the dog being washed by a girl", speakers formulate a conceptual representation of the message about the two event characters (e.g. dog, girl) and the event relationship between them (e.g. who did what to whom), and then retrieve appropriate lexical items and a syntactic structure to map all of this information to language. They must select from a range of potentially suitable lexical terms (e.g. dog, puppy, etc.) to express individual concepts and select from a range of possible syntactic structures (e.g. active, passive). As successful production requires settling on one option and inhibiting the others, activation of similar alternatives leads to competition in speakers' mind. But the time course of competition in production is much more difficult to estimate, as the scope of these selection processes can be flexible and may overlap with each other: speakers may begin encoding of individual character and its relational information at different points of planning under different circumstances. To this end, eye-tracking methodology has been used by a number of studies to examine the difference in the time course of lexical and structural planning, by identifying speakers' eye movement patterns that are specific to individual characters and relations between them, e.g. fixations distributed on the action region or on both characters (e.g. Konopka & Mayer, 2014; Van de Velde, Meyer & Konopka, 2014; Konopka, 2018). And one often raised question concerns whether and why speakers prioritize a linear word-by-word lexical-based planning or a hierarchical relation-driven planning at the early stage of sentence formation (e.g. after picture onset and before articulation), that is whether the evidence supports the *Linear Incrementality* or *Hierarchical Incrementality* account in sentence production.

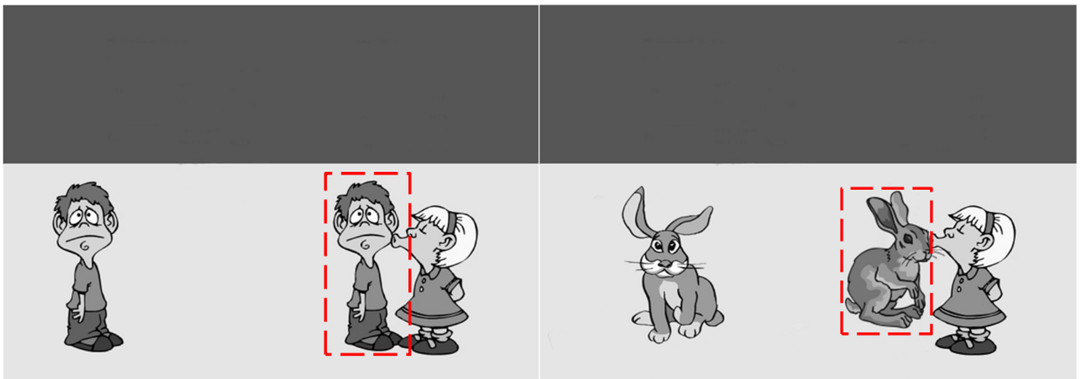
It has been shown that the flexibility of using different planning strategies (planning units specific to lexical encoding or also including relational information) partially depends on the ease of lexical and structural encoding, which is influenced by a number of factors including the accessibility and availability of lexical and structure alternatives (Bock, 1982; Bock & Warren, 1985; Myachykov, Scheepers, Garrod, Thompson, & Fedorova, 2013; Tanaka, Branigan, & Pickering, 2010) and lexical and structural priming (Bock, 1986; Konopka & Meyer, 2014; Pickering & Ferreira, 2008; Branigan, Pickering, McLean, & Stewart, 2006; Pickering & Branigan, 1998). Also, it depends on the complexity of the planned sentence and associated cognitive load. Wagner, Jescheniak and Schriefers, (2010) found that the scope of advanced planning is narrowed under increased cognitive load, by increasing the complexity of the target sentence or by including additional conceptual decision task. Thus, different production studies using different methodologies and targeting different utterance structures have reported divergent results: in some cases, the selection of critical information for sentence structure (such as verbs) is observed at initial stages before an utterance, which is consistent with *Hierarchical Incrementality* (Kempen & Huijbers, 1983; Schnur, Costa, & Caramazza, 2006; Konopka, 2019); whereas in others it is suggested to happen much later as a separate increment, as suggested by the *Linear Incrementality* account (Iwasaki, 2011; Schnur et al., 2006; Gleitman et al., 2007). Although results are mixed, it seems that early sentence planning is sensitive to the ease of encoding relational and non-relational information, and in order to maximize speech fluency, speakers can flexibly allocate resources to adopt different planning strategies to prioritize encoding of different information.

The current study examined the time course of semantic competition in complex phrase production. We focused on the encoding of relational information (i.e. syntactic role assignment) during verb planning, as the verb morphology should agree in form and meaning with the head subject (i.e. being kissed, or is kissing), and determines the agent/patient relationship between the entities. The hypothesis evaluated in this chapter is whether the competition of syntactic roles between highly reversible animate entities may reduce the ease of relational encoding, thus influences speakers' timing of verb planning. So far, the existence of semantic competition has been illustrated in many production studies and has been shown to constraint production performance. For example, it has been shown that in word production studies using a picture-word paradigm, naming latency of a target picture was significantly longer when the paired distractor word is semantically related to the target picture than when it is an unrelated word (Caramazza & Costa, 2000; Costa, Alario, & Caramazza, 2005). Similarly, at sentence or grammatical level, speakers mitigate similarity or reversibility-based competition via choices of utterance forms: passives were preferred to demote the accessibility of the competing entity (i.e. the agent) or entirely omit it in agentless passive, to ease planning interference (Gennari & MacDonald, 2009; Humphreys & Gennari, 2014). Also, in other languages which permit both SVO and VOS word order such as Tzeltal (primarily VOS based), SVO was preferred over VOS to ease planning interference by separating conceptually similar entities (Norcliffe, Konopka, Brown, & Levinson, 2015).

Only one eye-tracking study has explored the time course of competition processes in producing passive phrases (Humphreys et al., 2016). In a picture description study, participants apprehended pictures of various characters doing different actions and answered to questions about animate or inanimate patient characters. It was found that semantic similarity between characters modulates fixation on agents (i.e. the competing entity) before utterance and during the encoding of the verb phrase (i.e., during the utterance of the head noun phrase), with less similar agents/competitors tending to be fixated more than those that were more similar to the targets. Given that the fixation likelihood reflects the degree of activation/accessibility of the entities on speakers' mind (Huettig & Hartsuiker, 2008), the results suggest greater semantic-syntactic competition where a more similar/interfering competitor is present. That is, a competitor was fixated less (more inhibited) to make it less accessible in planning, especially during the encoding of the verb phrase at the point in which the syntactic roles of the nouns must be decided. However, this study only examined fixations on the animate entities (the subject of the embedded verb) because there were size differences between the animate and inanimate characters being referred to by the head-noun. This comparison does not examine fixation differences between animate target entities that differ in reversibility. Thus, it is less clear whether fixations on these entities may elicit a different pattern of results. Also, many eye-tracking production studies, including the above one, have not investigated the role of individual differences in predicting fixation patterns during online production. Thus, an important contribution of the current study is to provide data on how individual cognitive skills (e.g. vocabulary knowledge, executive control) become involved in resolving semantic competition in complex phrase production, and to test the reliability of any time-course effects being observed on different entities/regions of interest.

4.1.2 The present study

Motivated by previous production studies summarized above, we examined how the time course of complex phrase production is influenced by the reversibility between animate items and followed a similar procedure to that utilized in study 1 and a previous eye-tracking study (Humphreys et al., 2016) for comparison. Briefly, we created new pictures involving only animate agents and patients (e.g. boy, woman, dog, horse), and monitored participants' eye movements as they describe the patient character in each picture (e.g. *the boy/rabbit* being kicked by the girl, see figure 4.1). The use of new picture-items including only animate entities helps to minimize the difference in the degree of visual salience and conceptual accessibility between animate and inanimate patients/targets (although not entirely eliminated), thus allowing informative comparisons of all entities across conditions. Within these picture-items, we manipulated semantic-syntactic reversibility between the entities, such that half of the items included highly reversible agents and patients (i.e. high-reversibility condition) where both entities, e.g. *girl* and *boy*, are equally likely to be agents of the action, e.g. *kiss*. The other half of the items included less reversible entities (i.e. low-reversibility condition) e.g. *rabbit* and *girl*, where the patients are less eligible to take the agent role, e.g. *rabbit* is unlikely to do the action *kiss*, thus less competition is expected. By creating high and low-reversibility items matched on depicted event/action, it also helps to minimize the difference in event codability between conditions, which has been suggested to play an important role in the flexibility of sentence formation. It was found that when the event action is ambiguous and can be described with various verbs (i.e. low-codability events), speakers are more likely to fall back on linear incrementality, encode the head subject first and postpone relational encoding as which becomes difficult to complete, as compared to describing a high-codability event where the event action can be easily apprehended (e.g. Konopka, 2019; Konopka & Mayer, 2014).



High-reversibility condition Low-reversibility condition

Figure 4. 1 Examples of picture stimuli in the eye tracking production task

4.1.3 Study objectives and hypotheses

Then what pattern of results should we expect regarding fixation differences across agents and patients between reversibility conditions? At first, previous evidence has shown that when planning descriptions, speakers normally look at the characters in the order of mention, as language production is generally assumed to be incremental (Griffin & Bock, 2000). Thus, we would expect to see speakers fixate the patient/target before naming the first noun, and attend to the agent before naming the by-phrase agent. Also, during verb encoding, we expected fixation patterns across reversibility condition would resemble the patterns observed across animacy conditions in Humphreys et al.'s study. Thus, agents/competitors in the high-reversibility condition would be more inhibited/less fixated than those in the low-reversibility condition, due to greater semantic-syntactic competition in the former.

Moreover, in Humphreys et al.'s study, the similarity-based competition is mainly observed during utterance of the head noun phrase. It was suggested that at this point, the upcoming verb phrase was encoded (e.g. the boy that ***is being painted***), which must agree in form and meaning with the head noun (e.g. whether "the man" is the agent or the patient of "paint"). As a result, the semantic competition between similar noun phrases must occur at this point of planning. The authors did not report any significant differences after utterance of the head noun and before naming the verb (a period that may include pronoun and auxiliary e.g. *that is being*). We expect that fixation pattern during this period is also worth investigation, generally because how much advanced planning of the verb phrase was generated during the utterance of the head noun phrase is unclear. It is possible that speakers may experience difficulties with planning the verb phrase and adopt an incremental strategy. They may plan smaller units during the utterance of the head noun (e.g. they only planned the pronoun, auxiliary ***that/who is*** instead of the whole verb phrase ***is being kissed***) and postpone encoding of the verb (e.g. *kissed*) until the utterance of the pronoun and auxiliaries (e.g. *that/who is*). This is because, unlike *content words* such as nouns and verbs, pronouns and auxiliaries are *function words* without any intrinsic meanings, and are generally more accessible and less difficult to plan. Planning of function words may not always occur at a grammatical or conceptual level prior to articulation, but are sometimes uttered to gain additional planning time for upcoming difficult materials e.g. repetition typically involves function words (Griffin, 2003; Maclay & Osgood, 1959). Thus, in the current study, we analyzed fixation patterns after head noun offset and before verb onset (i.e. N1 offset -Verb onset) to investigate whether evidence of semantic competition can be observed here, before the utterance of the verb. But we do not expect to see fixation differences across reversibility conditions after the verb is uttered, as the competition should have been resolved by this point.

More importantly, little is known about the role of individual skills in predicting the time course of semantic competition during phrase planning. Given that vocabulary and inhibition skills played an important role in complex phrase production as demonstrated in study 1, we examined whether these relationships also exists in speakers' fixations to agents and patients. Specifically, we would expect that if the semantic competition is more difficult to resolve in the high-reversibility condition, reversibility should interact with inhibition performance during the encoding of verb phrase where the syntactic competition occurs, such that the likelihood of fixating on the target/patient or inhibiting the competitor/agent would be explained by inhibition measures to a greater extent for high-reversibility condition.

4.2 Experimental Methods

4.2.1 Participants

70 native English speakers (12 males: mean age=20.54, *SD*=2.01; 58 females: mean age=20.21, *SD*=1.30) from the University of York participated in the study for course credit or payment. 3 participants' eye movement data were removed due to poor calibration, and 4 participants' data were excluded due to too many errors in the production task (i.e. they always produce full sentences such as "*the sheep is being kicked by a boy*", instead of complex phrases like "*the sheep being kicked by a boy*"). This left us with 63 participants' data for analysis (10 males: mean age=20.75, *SD*=2.15; 53 females: mean age=20.23, *SD*=1.33).

4.2.2 Materials

Eye tracking production task. 20 scenes were created using graphic software and clip art obtained from the internet, describing events of 20 different actions (see appendix B). Each scene contained three or more animate characters, and at least two of them are involved in an action such as *carrying, lifting, painting*, etc. In one version of each scene, the agent acts on an animate patient who is also eligible to do the same action towards the agent (i.e. high-reversibility condition); and in another version, the action is performed on an animate patient who is not capable to do this action (i.e. low-reversibility condition). For example, in a scene describing the action "*kiss*" (see figure 4.1), in one version a girl is kissing a boy; and in another version a girl is kissing a rabbit. Each scene also contained at least one exemplar of the target/patient entity as the contrast entity, for example, the scene describing "*kiss*" also contained an additional *boy/rabbit* at the background.

***Agent-role likelihood/reversibility ratings*.** Note that although the picture items were categorized as high-reversibility and low-reversibility items, there are still variability in the degree of agent-role likelihood/reversibility across the items. Considering that eye tracking data is closely associated with the properties of visual stimuli in the scenes (e.g. sizes, positions of characters), we used ratings provided with the presentation of the actual experimental pictures (instead of verbal descriptions of the scenes as done in study 1). Two online questionnaires were created, in which participants were asked to rate the likelihood of the opposite event from that being shown in a given picture. They were presented with the actual experimental pictures which had red squares highlighting the patients of the actions, and were asked to rate the likelihood/ plausibility of each highlighted patient performing the agent role and acting on the current agent. The ratings were given on a 1 (extremely unlikely) to 7 (extremely likely) point Likert scale. Pictures of the same scene were included in different versions of questionnaires which were conducted with different participants, to ensure each participant only rated each scene once. 33 participants completed the first version of the questionnaire, and 34 participants completed the second version of the questionnaire. As expected, the ratings for high-reversibility items (*mean*=4.28, *SD*=1.19) were significantly higher than the ratings for low-reversibility items (*mean*=1.47, *SD*=0.56): $t(19)=9.991$, $p<0.001$. This suggests that the reverse agent-patient relationship is more plausible in high-reversibility condition, which might lead to greater syntactic competition during phrase planning.

4.2.3 Design and Procedure

The design and procedure were similar to the production tasks described in study 1 (see pages 47-48). Briefly, the 40 pictures were allocated to 2 different lists and conducted with different participants, with each scene only appear once in either list. We also included 19 fillers in each list which elicited different kinds of responses (e.g. *the boy carrying a teddy bear*), and ensure one filler occur between any two experimental pictures to reduce structural priming in production.

The whole experiment lasted about an hour. Participants first completed the eye tracking production task. They were seated in front of a 22-inch display monitor, with their eyes approximately 60 cm away from the monitor. In each trial, participants inspected a picture for two seconds, then a red box appeared to highlight one character in the picture and stayed on the screen for one second. Participants were instructed to describe the highlighted character, and focus on the action going on rather than the appearance or the position of the character, and then press a key to proceed to the next trial. Their verbal responses were recorded through a microphone positioned in front of them. Their eye movements were recorded by an Eye Link II head-mounted eye tracker, sampling at 250 Hz. After the production task, all participants completed 3 cognitive tasks adopted from study 1 (see pages 48-50 for descriptions of these assessments): the vocabulary subtests from the WASI-II, the STOP-IT task and the homograph task. For the homograph task, two participants' data was not recorded due to program malfunctioning.

Descriptive statistics and the correlation matrix between cognitive measures are shown in table 4.1 and 4.2 separately. None of the measures were correlated with each other, suggesting that they tap on different underlying cognitive skills.

Table 4. 1 Descriptive statistics for individual differences measures

	N	**Range**	***Mean***	***SD***
Vocabulary	63	40-68	54.11	4.33
STOP	63	185.50-332.30	265.01	31.91
Homograph	61	-194.43-520.34	176.04	174.33

Note: *Vocabulary* stands for *WASI-II vocabulary subtest, STOP* stands for *STOP-IT performance (SSRT)* respectively.

Table 4. 2 Correlations between individual differences measures

	Vocabulary	**STOP**	**Homograph**
Vocabulary	1.00	0.33	0.13
STOP	0.33	1.00	0.22
Homograph	0.13	0.22	1.00

*p<0.05; **p<0.01

4.2.4 Data coding and analysis

4.2.4.1 Coding of verbal responses

Verbal responses for experimental trials were transcribed by the experimenter and research assistants. Incorrect responses (e.g. those that did not uniquely identify the target character) or descriptions that did not include the targeted structure were first excluded from analyses. In the remaining responses, the majority of utterances were passives (high-rev: 99.3%, low-rev: 99.5%) and there were no significant differences between reversibility conditions (by participants: z=-1.225, p=0.221; by items: z=-0.730, p=0.465). Further analyses with eye tracking data only included passive responses, given that different word orders in active and passive structures would make condition non-comparable.

In each correct passive response, the onsets and offsets of relevant elements were marked using the Praat software. Relevant elements included *the first determiner*, *first noun*, *relative pronoun*, *verb auxiliary*, *main verb*, *by-preposition*, *second determiner* and *the second noun*; and only present elements were marked. The timings of all these markers were then aligned with eye movement data starting at the position where recording starts (i.e. when red box appears).

4.2.4.2 Data analyses

Eye movement data analyses. The eye movement data was analyzed in relation to participants' fixations on the regions of interest in each scene. The regions of interest were drawn on each picture defining the areas of different entities, including the *target/patient*, *the competitor/agent*, and *the contrast* (i.e. the distractor in the background). For example, in the event of *kiss* in figure 4.1, the target/patient is the boy or the rabbit, the agent is the girl and the contrast is the additional boy or the rabbit in the scene.

To test which planning stages are influenced by the reversibility manipulation and various individual difference abilities, analyses of fixation data were performed on different time windows before and during the speech separately, including the ***SOT***, utterance of the ***head noun phrase***, ***N1 offset -Verb onset*** and the ***main verb***. The reason to not include a by-phrase time window is because the by-phrase is always the last element speakers need to utter, thus fixation patterns during this time window are not informative regarding to speakers' planning of the subsequent utterance.

For each time window, we determined whether a particular entity was fixated or not from the eye-movement record, and these binomial fixation data was then entered into logistic linear mixed-effects models in R (version 3.4.1, bobyqa optimizer and maximum iterations set at 100,000). All the initial models included the maximal random-effects structure: by subject and by item intercepts, by subject and by item random slopes for all fixed factors. In cases of non-convergence, we removed the random slope parameter which accounted for the least amount of variance until convergence was achieved (Barr, Levy, Scheepers, & Tily, 2013). We first examined whether fixations in each time window differed between reversibility conditions, and across different entities, by including "entity"[4] (three categories: *agent* as the reference category, *patient*, *contrast*), reversibility condition and their interactions as fixed effects. Significant interactions were interpreted using post hoc comparisons with holm adjusted tests ('emmeans' package in R).

To examine the unique contribution of different cognitive skills, we then run separate models to predict fixation likelihood on agents and patients. This is because a model including fixation data from all entities, reversibility condition and its interaction with any cognitive measure was too complicated to converge. Also, the contrast entity was generally unlikely to be attended during the speech (because it is not a relevant character), and fixations on this character did not differ across conditions, thus we do not report separate analyses to predict fixation likelihood on the contrast. In each of the model conducted, we entered z scored cognitive measures and their interactions with the reversibility condition in prior selected orders: expressive vocabulary was entered first to account for the role of lexical knowledge, measures of inhibition skills (stop-it SSRT, homograph inhibition RTs) were entered secondly to examine whether competition resolution skills explain additional variance after controlling for the effects of vocabulary. At each stage, non-significant interaction or main effects of cognitive measures were pruned to identify the simplest most explanatory model.

[4]By including entity as a fixed effect, it helps to minimize the number of separate comparisons conducted across regions of interests.

Agent omission and SOT analyses. We also conducted analyses with the coding of agent omission and length of SOT using logistic and linear mixed-effects models (LMEMs) separately, following the same procedure in study 1 for comparison.

4.3 Results

4.3.1 Reversibility effects in agent omissions

Based on previous results, we expected that the likelihood of agent omission should be higher in the high-reversibility condition, given that speakers experience greater interference in planning highly reversible nouns, thus they may inhibit one of the nouns (i.e. the agents) during planning and also completely omit it from the description. Thus, if any cognitive measures should predict the likelihood of agent omission, we would expect an interaction between reversibility condition and inhibition skills. As predicted, the LMEM model in table 4.3 reported a significant main effect of reversibility condition ($p<0.001$), and the average proportion of agent omission by items also correlated with the reversibility ratings ($r(40)=0.311$, $p=0.05$). This suggests that speakers' tendency to omit the interfering agent is modulated by the degree of reversibility-based competition between the noun concepts. Importantly, there was also significant interaction between homograph inhibition and reversibility condition ($p=0.04$). As compared to poor inhibitors, good inhibitors were less affected by the reversibility manipulation and thus displayed less differences between conditions (see figure 4.2).

Table 4. 3 Main effects and interactions for reversibility condition and individual difference measures in predicting agent omission

	Coefficient	**SE**	***z-score***	***p*-value**
Intercept	0.63	0.34	1.83	0.07
Condition	-1.02	0.29	-3.54	<0.001*
Homograph	0.06	0.26	0.24	0.81

Homograph*Condition	-0.45	0.21	-2.10	0.04*

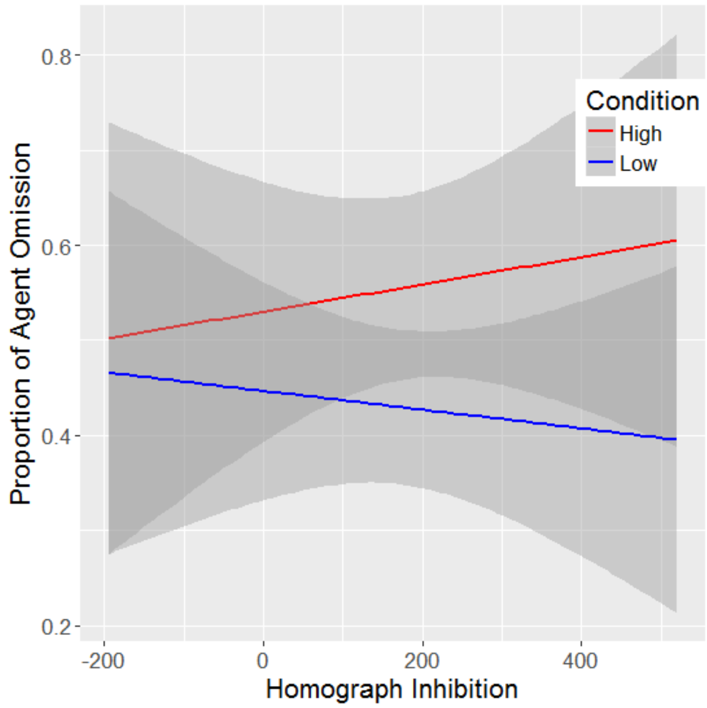


Figure 4. 2 Interaction between reversibility condition and homograph inhibition scores in predicting proportion of agent omission

4.3.2 Reversibility effects in speech onset time (SOT)

Similar to study 1, there were significant main effects of reversibility condition (*p*=0.01, see table 4.4) and homograph inhibition (*p*=0.004). SOTs were longer for high-reversibility items than low-reversibility items, and good inhibitors required less initiation time in their descriptions, suggesting that SOTs were relevant to reversibility-based competition, and SOTs for animate-targeted descriptions vary as a function of individual inhibition skills. More importantly and unique to the current study, there was also main effect (*p*=0.04) and marginal significant interaction between expressive vocabulary and reversibility condition (*p*=0.068, see figure 4.3). Individuals with better vocabulary had shorter SOTs overall and were also less affected by the reversibility manipulation, thus displaying less difference between reversibility conditions as compared to individuals with poorer vocabulary.

Table 4. 4 Main effects and interactions for reversibility condition and individual difference measures in predicting SOT (ms)

	Coefficient	SE	*t-score*	*p*-value
Intercept	1494.18	58.70	25.53	<0.001*
Condition	-146.10	52.91	-2.76	0.01*
Vocabulary	-95.48	44.60	-2.14	0.04*
Homograph	103.98	36.46	2.85	0.006*
Vocabulary*Condition	51.09	27.46	1.86	0.068

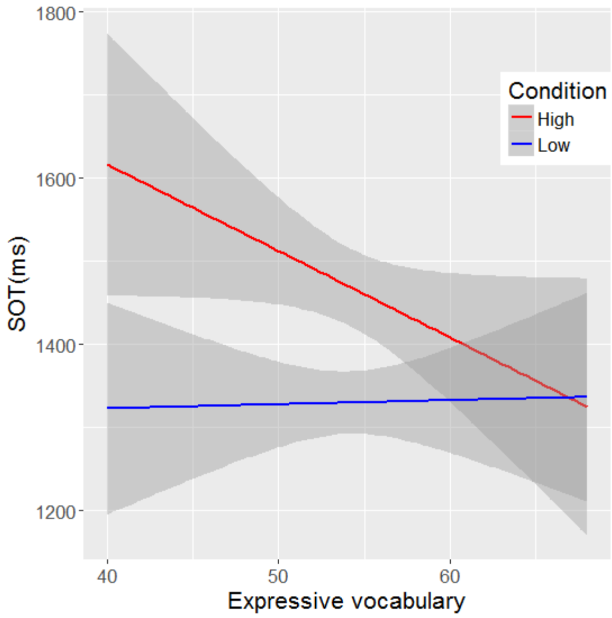


Figure 4. 3 Interaction between reversibility condition and expressive vocabulary in predicting production SOTs (ms)

4.3.3 Eye movement during SOT

Reversibility effects. Now we move on to the analyses with fixation data during SOT. Fixations before utterance generally reflect initial apprehension of main aspects of the event, and the first entity to be named is fixated the most (Griffin & Bock, 2000). Indeed, we found speakers were more likely to fixate on the patient (the first named entity) than the other entities, and in turn, the agent was more likely to be fixated than the contrast (see figure 4.4 and table 4.5). There was no significant fixation difference between reversibility conditions across the entities, suggesting the reversibility effect is not reflected here.

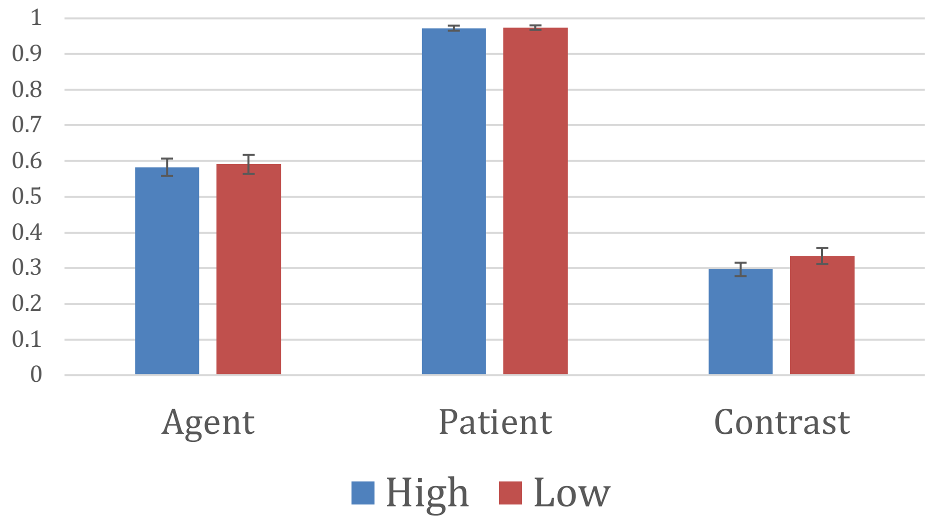


Figure 4. 4 Average proportion of trials fixated on entities during SOT

Table 4. 5 Results of models predicting fixation likelihood before utterance of the head noun

		Coefficient	**SE**	***z-score***	***p*-value**
All fixations	Intercept	0.37	0.13	2.97	<0.001*
	Agent vs. Contrast	-1.63	0.17	-9.84	<0.001*
	Agent vs. Patient	3.73	0.55	6.81	<0.001*
	Condition	0.06	0.15	0.37	0.71
	Agent vs. Contrast*Condition	0.13	0.22	0.59	0.56
	Agent vs. Patient*Condition	-0.16	0.72	-0.23	0.82
Fixation on agent	Intercept	0.42	0.17	2.48	0.01*
	Vocabulary	0.13	0.12	1.13	0.26
	STOP	0.19	0.12	1.62	0.11

Condition	0.06	0.17	0.34	0.73
Vocabulary*Condition	-0.30	0.15	-2.07	0.04*
STOP*Condition	-0.31	0.15	-2.14	0.03*

Individual difference. As shown in table 4.5, logistic mixed models indicated significant predictors for fixation likelihood on the agent, but none of the measures predict fixations on the patient. Specifically, fixations on agents were explained by significant interactions between reversibility condition with expressive vocabulary (*p*=0.04) and with STOP-IT performance (*p*=0.03), but these relationships exert opposite influences (see figure 4.5).

To interpret the interaction with expressive vocabulary, paired sample t-tests were conducted with speakers having high vocabulary (*N*=35, *Range*: 55-68) and low vocabulary (*N*=27, *Range*: 40-53) separately, using a median split. The results show that speakers with poor vocabulary are less likely to fixate on agents in the high-reversibility condition as compared to low-reversibility condition: *t*(26)=-1.872, *p*=0.073. This may suggest they experience greater difficulties in retrieving the target patient, as indicated by the SOT duration analyses, and thus they are less likely to fixate the competitor in the high-reversibility condition. Speakers with good vocabulary, on the other hand, do not fixate differently on agents across conditions: *t*(35)=1.044, *p*=0.304. This is also consistent with the finding that they show no difference in SOT durations.

The interaction with STOP-IT performance indicates that better inhibitors (i.e. low SSRT, *N*=32, *Range*: 185.5-266.5) are less likely to fixate on the agents/competitors in high-reversibility condition as compared to low-reversibility condition: *t*(31)=-1.982, *p*=0.056; suggesting that they are more likely to inhibit activation of highly-interfering competitors and sustain attention to the patients/targets. On the other hand, poor inhibitors (i.e. high SSRT, *N*=31, *Range*: 270.4-332.3) are affected by the reversibility manipulation: they are more likely to attend to highly similar and reversible competitors as compared to less reversible ones, $t(30)=1.878$, $p=0.070$. Overall, in addition to vocabulary, it appears the inhibition plays a role at this early stage of planning.

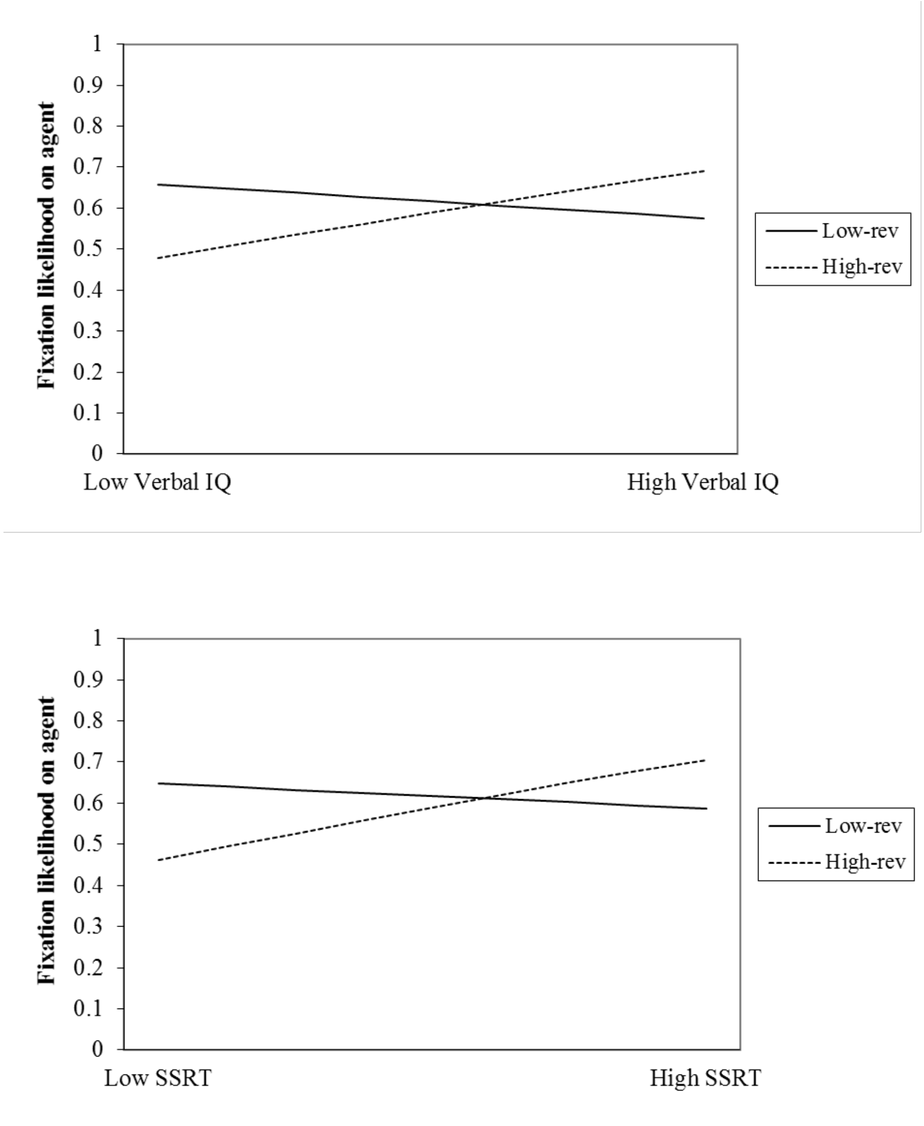
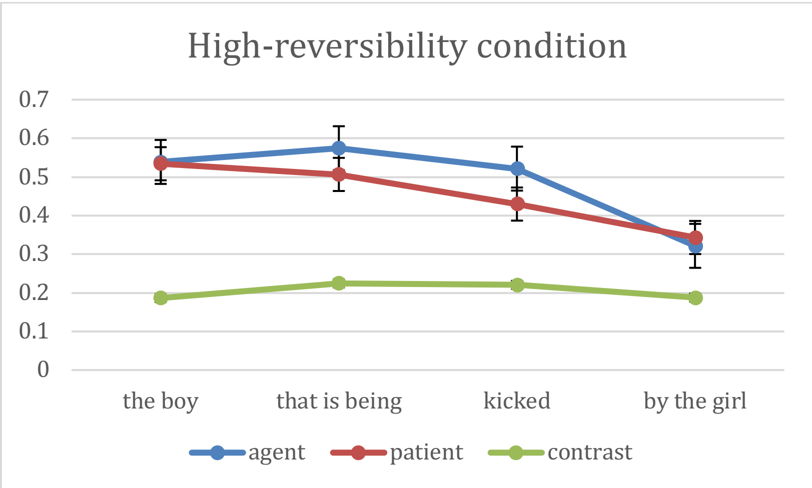
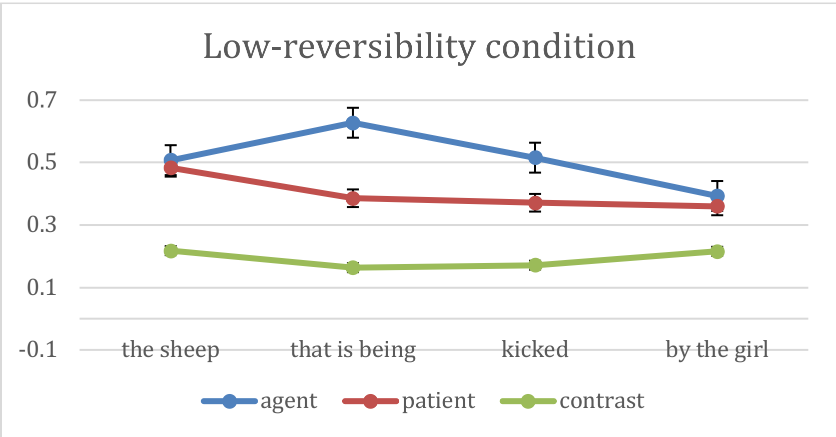


Figure 4. 5 Interactions between reversibility condition and vocabulary (upper panel), and between condition and motor inhibition (lower panel) in predicting fixation likelihood on the agent/competitor during SOTs.

4.3.3 Eye movement during speech

We then move on to analyses of fixation data during the speech. Figure 4.6 shows the overall fixation pattern on different entities during the utterance of critical phrases, for high-reversibility and low-reversibility conditions respectively, and a comparison between conditions. In general, relevant entities (i.e. agent and patient) were more likely to be fixated than irrelevant ones (i.e. contrast), and the greatest difference between fixations on agents and patients seems to be during N1 offset -Verb onset where the syntactic roles of the nouns play a role in deciding the verb morphology before naming the verb. To further examine statistical differences in fixation patterns across conditions and its relationship with cognitive measures, we next report fixation analyses during pre-defined time windows separately.





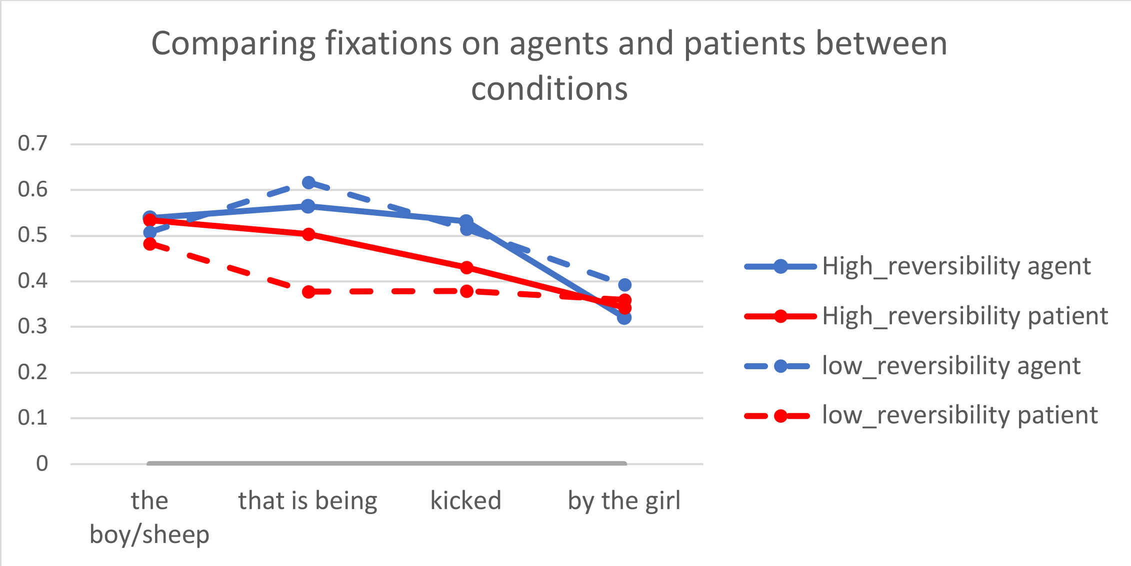


Figure 4. 6 Average proportion of trials fixated on entities when critical phrases were uttered (error bars represent standard error)

4.3.3.1 Head noun phrase

The average length of this time window was 546.23 ms, and it does not differ between reversibility conditions: $t(62)=1.592$, $p=0.116$. Following previous findings, during the utterance of the first determiner and the head noun, aspects of the upcoming verb phrase should be planned and early competition between agent and patient entities might be expected.

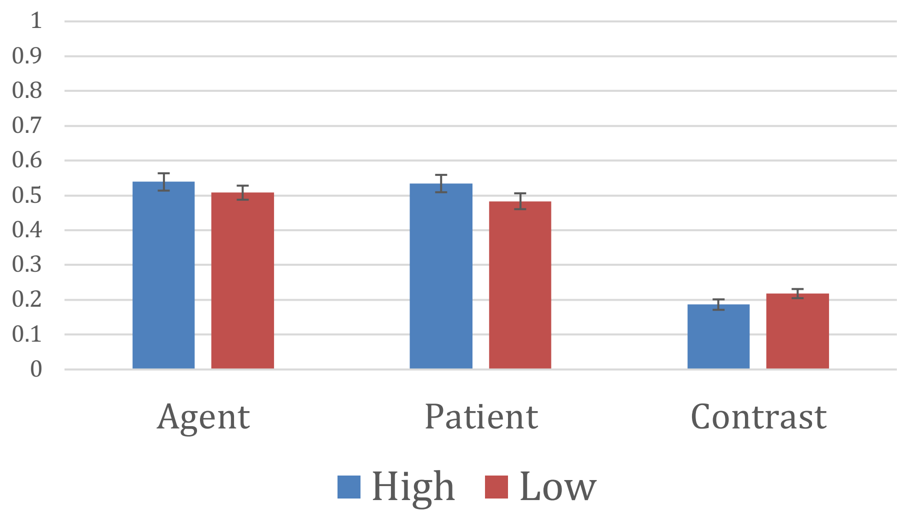


Figure 4. 7 Average proportion of trials fixated on entities during utterance of head noun phrase

Reversibility effects. As shown in table 4.6 and figure 4.7, there was no significant difference in fixation likelihood between agents and patients across conditions. This seems to suggest that while uttering the noun phrase and beginning to plan the verb phrase, both the agent and patient are equally relevant in speaker's mind, suggesting that the relationship between the entities is probably being considered.

Moreover, despite the contrast being less active than other entities, there was a significant interaction between condition and entities. This was because the contrast was more fixated in the low-reversibility condition (z=-2.377, p=0.02), whereas there was no significant difference between conditions for agent and patient (agent: z=1.044, p=0.30; patient: z=-1.671, p=0.09). This may be because the contrast in high-reversibility condition is always more interfering as it shares more features with both the agent and patient characters in the scene, thus less attention was allocated to it to reduce the availability of potential competitors. Further individual difference analyses did not identify any significant predictors for fixation likelihood on the agent or patient entity.

Table 4. 6 Results of models predicting fixation during the utterance of head noun phrase

		Coefficient	**SE**	***z-score***	***p*-value**
All fixations	Intercept	0.18	0.11	1.66	0.10
	Agent vs. Contrast	-2.49	0.25	-10.00	<0.001*
	Agent vs. Patient	-0.03	0.14	-0.24	0.81
	Condition	-0.14	0.14	-1.04	0.30
	Agent vs. Contrast*Condition	0.69	0.28	2.47	0.01*
	Agent vs. Patient*Condition	-0.07	0.18	-0.41	0.68

4.3.3.2. N1 offset -Verb onset

As hypothesized, fixations during N1 offset -Verb onset may reflect planning of the upcoming auxiliary + verb where syntactic role assignment comes into play, and evidence of competition is expected here. For this time window, the average duration was significantly longer for high-reversibility items (*mean*=725.60, *SD*=297.66) as compared to low-reversibility items (*mean*=559.19, *SD*=191.85): $t(62)=5.569$, $p<0.001$, and also significantly correlated with the reversibility ratings: $r(40)=0.425$, $p=0.006$. However, speakers did not produce more pronouns for high-reversibility-items as compared to low-reversibility items $t(66)=-1.235$, $p=0.221$. This suggests that the length of this duration is relevant to reversibility-based competition, but is not reflected in the tendency to produce pronouns to gain additional planning time for difficult items, and maybe relevant to other measures such as fixations to characters (as reported below) during this time window.

Reversibility effects. During this time window, speakers differ marginally significantly in their fixation odds on relevant entities, with agents being more likely to be fixated than patients ($p=0.06$). Also, there was a significant interaction between condition and the comparison between agent and patient ($p<0.001$). Post hoc analyses (which ones) revealed that in the high-reversibility condition, the fixation difference between agent and patient ($z=1.878$, $p=0.0604$) was less significant, as compared to the difference in low-reversibility condition ($z=8.920$, $p<0.001$) where the agents were more fixated. This seems to suggest that competition between the two animate entities was more difficult to resolve in high-reversibility items, thus speakers tend to fixate both entities to encode the syntactic relationship between them. Also, similar to the previous finding, the competitor/agent was marginally less likely to be fixated in the high-reversibility than the low-reversibility condition (post hoc analysis for the above interaction: $z=(-1.806)$, $p=0.07$).

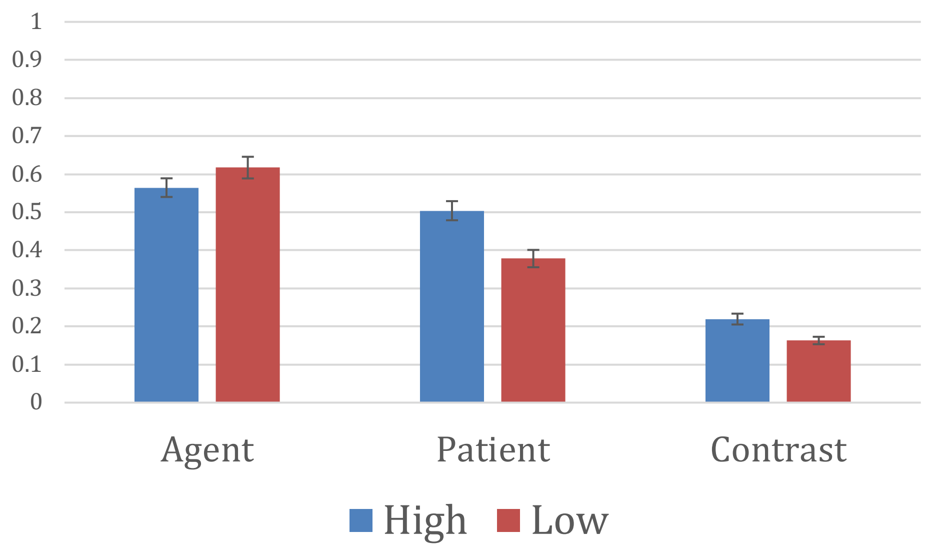


Figure 4. 8 Average proportion of trials fixated on entities during utterance of the pronoun and auxiliary

Table 4. 7 Results of models predicting fixation during the utterance of pronoun and auxiliary

		Coefficient	**SE**	***z-score***	***p*-value**
All fixations	Intercept	0.26	0.14	1.81	0.07
	Agent vs. Contrast	-2.24	0.16	-14.06	<0.001*
	Agent vs. Patient	-0.24	0.13	-1.88	0.06
	Condition	0.25	0.14	1.81	0.07
	Agent vs. Contrast*Condition	-0.37	0.23	-1.64	0.10
	Agent vs. Patient*Condition	-0.89	0.18	-5.01	<0.001*
Fixation on patient	Intercept	-0.02	0.24	-0.09	0.92
	Homograph	-0.36	0.12	-2.92	0.003*
	Condition	-0.73	0.21	-3.43	<0.001*
	Homograph*Condition	0.30	0.16	1.94	0.05*

Individual difference. As shown in table 4.7, the likelihood of fixating the patient/target was predicted by a significant main effect of homograph inhibition (*p*=0.003) and there was a significant interaction of this measure with reversibility condition (*p*=0.05). That is, as semantic inhibition is better, fixation odds on the patient/target increase, and more so in high-reversibility condition (see figure 4.9). This suggests that speaker's likelihood to enhance activation of the targets during semantic-syntactic role encoding is linked to individual's semantic inhibition skills. None of the measures predicts fixation odds on the agent.

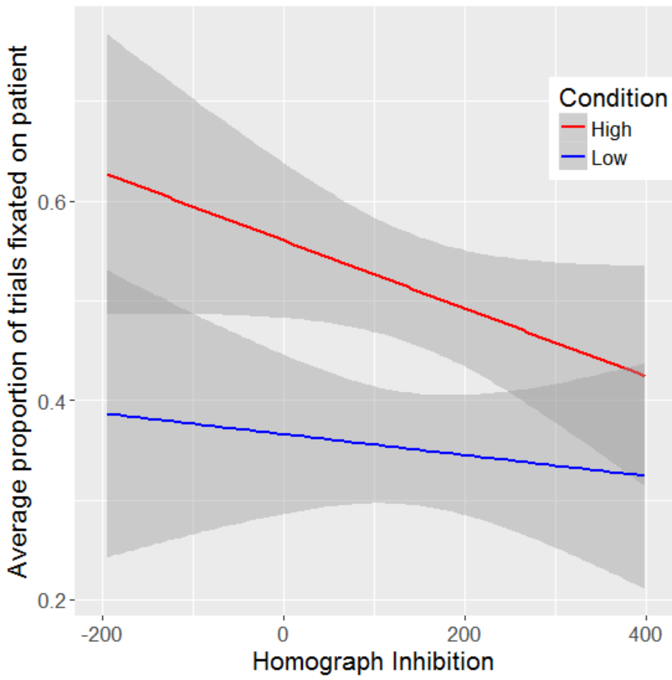


Figure 4. 9 The interaction between reversibility condition and homograph inhibition in predicting fixation likelihood on the patient

4.3.3.3. Main verb

Fixation during the utterance of the main verb should reflect the planning of the upcoming by-phrase if there is any, thus we expect the agents would be more likely to be fixated than other entities. All descriptions with and without by-phrases were included in the analyses, given that the proportion of agent omission is very high for both conditions (high: 56.82%; low: 42.21%) and excluding those without by-phrases would result in removing too much data.

Reversibility effects. For this time window, the average duration was significantly longer in the high-reversibility condition (*mean*=468.55, *SD*=92.27) as compared to low-reversibility condition (*mean*=447.77, *SD*=84.38): $t(62)=3.189$, $p=0.002$. As many descriptions did not include by-phrases, this may simply reflect wrap-up effects (or continuing interference effect) of previous verb planning processes. As predicted, during this time window, agent was more likely to be fixated than the contrast ($p<0.001$) and the patient ($p=0.07$), and there was no significant fixation difference across entities between reversibility conditions (see figure 10 and table 8).

Individual difference. As shown in table 4.8, the likelihood of fixating the agent (i.e. target of the upcoming by phrase) was predicted by a significant main effect of STOP-IT performance ($p=0.03$), with better inhibitors (i.e. low SSRT) more likely to fixate/enhance activation of the agents/targets.

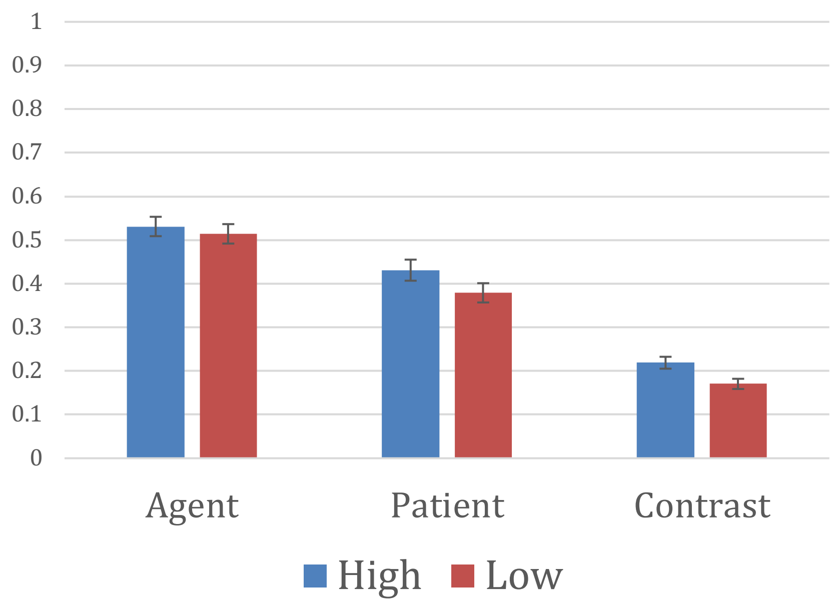


Figure 4. 10 Average proportion of trials fixated on entities during utterance of the main verb

Table 4. 8 Results of models predicting fixation during the utterance of the main verb

		Coefficient	**SE**	***z-score***	***p*-value**
All fixations	Intercept	0.12	0.14	0.85	0.39
	Agent vs. Contrast	-2.26	0.25	-8.88	<0.001*
	Agent vs. Patient	-0.51	0.28	-1.79	0.07
	Condition	-0.04	0.15	-0.30	0.77
	Agent vs. Contrast*Condition	0.08	0.29	0.29	0.77
	Agent vs. Patient*Condition	-0.38	0.25	-1.51	0.13
Fixation on agent	Intercept	0.11	0.13	0.87	0.38
	Condition	-0.03	0.13	-0.24	0.81
	STOP-IT	-0.16	0.07	-2.17	0.03*

4.4 Discussion

4.4.1 Summary of non-fixation results

The study aimed to investigate the time course of reversibility-based competition during passive phrase planning, and how cognitive skills become involved in individual's sensitivity to production competition as reflected in eye-movement patterns. Overall, for the non-fixation data, we found that speakers experienced greater competition when planning high-reversibility animate nouns as compared to low-reversibility nouns, such that they required longer SOTs, longer planning time for verb morphology encoding, and were also more likely to omit the by-phrase agents to strategically ameliorate planning interference.

These findings were in line with study 1 in suggesting phrase production is modulated by noun reversibility, but the analyses of individual difference have identified slightly different predictors or relationships for SOT and agent omission. For SOT, in addition to a general influence of homograph inhibition, we also found vocabulary knowledge interacted with reversibility conditions, such that high-reversibility items had a stronger relationship with vocabulary knowledge than low-reversibility items. This may be due to the fact that the current study has increased power to detect such a relationship with vocabulary by increasing the number of picture-items for comparisons (i.e. 20 items per reversibility condition vs. 7 items per condition), and by utilizing a better controlled manipulation of noun-reversibility across conditions. In the current study, the grouping of high vs. low items was based on whether the target/patient entity was eligible to compete for the agent role under the same scenarios, rather than reversibility ratings of animate items featuring different sentences as in study 1.

For agent omission, although the same interaction between homograph inhibition and reversibility conditions was observed, this relationship was in the opposite direction as reported in study 1. In the present study, poor inhibitors were more affected by reversibility manipulation and were more likely to omit agents for high-reversibility than low-reversibility items. In study 1, in contrast, we found that poorer inhibitors omit agents less in the high-reversibility than the low-reversibility condition. One explanation could be due to differences in the cognitive effort induced by the two production tasks. Speakers may experience greater competition in sentence structure selection in study 1's production task, as they were heavily primed with the active alternatives in the preceding comprehension task, which was not administrated in the current study. Thus, for speakers from study 1, part of their planning effort in producing passives (including agentless passives) was to avoid planning two competing nouns together in actives, thus the agent concept needed to be strongly inhibited/ignored at early stage of planning. Good inhibitors would inhibit the agent relatively more strongly than those with poorer inhibition, and thus would tend to omit the agent by-phrase more often in the high-reversibility condition. In the current study, in contrast, speakers were not explicitly made aware of the active alternatives, and participants from the beginning always adopted the same structure (i.e. passives) due to priming within the experiment. For poor inhibitors, the difficulty of planning the by-phrase agents might be harder for high-reversibility items, because this may cause greater interference with the currently active element, thus would show more agent omissions. Thus, individual differences in agent omissions may reflect either early inhibition processes (as in Study 1) or later encoding processes (as in the present study), as argued by (Hsiao et al., 2014).

Table 4. 9 Summary of results

Non-fixation data	**DV**	**Fixed factor**	**Predictors**	**Interaction**
	Ag Om	Reversibility (High > Low)		Hmg*Reversibility
	SOT	Reversibility (High > Low)	Vocabulary Hmg	Vocabulary*Reversibility
Fixation data				
Time window	**ROI**	**Fixed factor**	**Predictors**	**Interaction**
SOT	All fixs	Entity (Patient>Agent>Contrast)		
	Agent			Vocabulary*Reversibility STOP*Reversibility
Head Noun	All fixs	Entity (Patient=Agent>Contrast)		Agent vs. Contrast*Reversibility
N1 offset + Verb onset	All fixs	Entity (Agent>Patient>Contrast)		**Agent vs. Patient*Reversibility**
	Patient	Reversibility (High>Low)	Hmg	**Hmg*Reversibility**
Verb	All fixs	Entity (Agent>Patient>Contrast)		
	Agent		STOP	

Note: *Ag Om* stands for *agent omissions*, *Vocabulary* stands for *WASI-II vocabulary subtest*, *Hmg* stands for *homograph, STOP* stands for *STOP-IT performance (SSRT)* respectively.

4.4.2 Summary of fixation results

The fixation pattern suggests that the speakers fixate the entities in the order of mention: fixation was directed to the patient before naming it, then neither patient nor agent was preferentially focused during the utterance of the head noun, which may suggest apprehension of the grammatical relationship between the animate nouns. This was followed by the agent being more likely to be fixated before the verb is uttered, suggesting the encoding of the subsequent by-phrase agent or simply that the speakers attend to the action executor to retrieve a name for the verb, e.g. in descriptions such as “the boy being kissed by the girl”, speakers need to fixate to the “girl” character for accurate description of the action.

More importantly for the purpose of this experiment, we observed reversibility-based fixation difference not just on agents. A significant fixation difference between agents and patients across reversibility condition is only observed during the N1 offset + Verb onset time window (see bold in table 4.9). In particular, speakers were more likely to fixate the agents for low-reversibility items before naming the verb, whereas their fixations on agents and patients differ less significantly for high-reversibility items and they only start fixating the agents after the verb onset. This pattern of results seems to suggest syntactic competition being more difficult to resolve in the high-reversibility condition, leading to less clear divergence between fixations on agents and patients, as the syntactic roles of the entities cannot be easily decided between highly reversible nouns. In contrast, the clearer divergence between fixations on agents and patients (i.e. agents being more fixated) in the low-reversibility condition may indicate that the syntactic roles can be easily assigned between the nouns, thus the agent/competitor is not inhibited as it does not interfere as much with the patient role of the target. The agents being more fixated for low-reversibility items may also suggest that less reversible entities engender less planning interference during verb morphology encoding, which leads to an early shift toward encoding of the subsequent by-phrase agent. This is consistent with the idea that the degree of incrementality during production can be influenced by the variability in the ease of lexical or grammatical encoding. For example, it was shown that in simple SVO sentence, easier encoding of the subject entity under semantic priming condition results in an earlier shift of gaze and attention to the next to-be-planned element (i.e. the objects), as compared to fixation patterns observed under no-priming condition (Ganushchak, Konopka, & Chen, 2017). Similarly, speakers prioritize encoding of relational information (e.g. event action) when such process becomes more accessible and less cognitive demanding under syntactic priming, or when the to-be-described event gist is easy to encode (Konopka & Meyer, 2014). Thus, our results confirm previous findings in that the timing of encoding different information is likely to vary significantly as a function of task demands (e.g. Konopka & Meyer, 2014; Lee, Brown-Schmidt & Watson, 2013; Wagner, Jescheniak & Schriefers, 2010), with a new variable: the semantic competition between highly similar and reversible animate nouns also modulates the ease of apprehending the syntactic relationship between the entities, thus influences the timing of encoding this information during verb planning.

Critically, during this time window, the likelihood of fixating the patients/targets were predicted by homograph inhibition performance. That is, good inhibitors were more likely to fixate on the patients/targets in the high-reversibility than the low-reversibility conditions. Recall that the fixations on patients were generally more likely for the high-reversibility than the low-reversibility condition (see Figure 4.8), whereas fixation on agents did not differ across conditions. In the context of this interaction, the role of inhibition on patient fixations suggests that good-inhibitors were able to fixate and maintain activation of the target noun during syntactic role assignment in the context of an active competitor, but such maintenance was not much required in the low-reversibility condition where the role assignments is easier. In contrast, poorer inhibitors fixate on both characters equally on both conditions, suggesting they experienced difficulty in this process in all conditions. Nevertheless, more eye-tracking studies also examining individual differences are necessary to fully understadings the present relationships.

Overall, the fixation data is consistent with the view that reversibility-based competition occurs at the point where the verb morphology must be planned to indicate a semantic-syntactic role of the target noun, and that semantic inhibition skills underpins speakers' ability to maintain the target representation. However, in the current study, the time course of reversibility-based fixation difference is observed preceding the verb utterance, rather than during the production of head noun phrase as demonstrated in the Humphreys et al.'s study. This might be related to the lack of statistical power for comparisons within animates in that study compared to the present one, which only included animate entities, rather than animate vs inanimate entities. This meant that Humphreys et al. could only correlate fixations with similarity/reversibility ratings within animate entities, rather than compare across categorical conditions. Nevertheless, both studies suggest that it is at the point of planning the verb phrase that reversibility and noun similarity effects are observed, so they are manifested in a slightly different time window.

4.4.3 Conclusion

Taken together, the above production results suggests that reversibility based competition manifests at verb positions, and is particularly relevant to individual's semantic inhibition skill. This then parallels previous comprehension findings: the semantic-syntactic competition in comprehension occurs at the verb position, when similar nouns need to be held together in WM and assigned with appropriate syntactic roles with the input verb (Gennari & MacDonald, 2008; Gordon et al., 2006). Thus, despite the fact that production involves processes which are absent in comprehension (e.g. self-controlled, accessibility-driven plans), together these results point to shared competition resolution process across tasks, which occurs at verb planning and verb comprehension, and guides syntactic function assignment.

The results of this study is also compatible with predictions made by linear incrementality in production (e.g. Gleitman, January, Nappa, & Trueswell, 2007; Levelt, 1982), as the difference in fixation patterns between conditions were not observed during the SOT, but shortly before the verb was uttered. This suggests that complex phrase planning proceeds in a word-by-word, or concept-by-concept fashion; and relational encoding of the nouns and the verb (i.e. syntactic assignment) did not occur early to generate a conceptual representation of the utterance during SOT. This can be interpreted as a consequence of using a red square to highlight the target entity in the current design, which increased accessibility of the head noun and guided speakers to concentrate more on lexical encoding of the head noun before utterance. Nevertheless, the differences in SOTs suggested some competition at the selection of the first none, which was simply not reflected in fixations to characters, and not relevant to resolution of reversibility-based competition.

Altogether, our results provide a temporally fine-grained view of passive phrase production and suggest verb planning is guided by the reversibility-based competition between agents and patients. One direction for future research is to include separate entity and action regions, i.e. the part of picture that provides crucial information about what action is being depicted (e.g. "the boy’s hand holding the paint brush"). This will hopefully allow for a clearer interpretation of the eye tracking data, as in our design, fixations on agents cannot unambiguously indicate whether the person or the action that the person is doing was being considered in speaker’s mind.

Chapter 5

Discussion

The aim of this thesis was to understand the relationship between complex phrase comprehension and production by examining whether the two tasks draw on shared mechanisms and resources and whether these vary over development. The answer to this question is not only important in its own right, for example, for understanding the overall architecture of the cognitive system serving language from childhood to adulthood; but also for elucidating key components of any language model. However, comprehension and production have been typically investigated and modelled separately, and most research has been heavily biased towards comprehension, especially with children and adolescents. The current work aimed to fill this gap by assessing comprehension and production of complex phrases, both of which are known to induce semantic competition in adults. Moreover, it has been unclear whether competition and processing difficulty more generally are caused by common processes recruited by comprehension and production or whether distinct processes happen to show parallel behavioural effects. Thus, the goal of the present work was to investigate the extent to which comprehension and production engage common or distinct cognitive processes and resources for semantic competition.

5.1 Summary of results

The present studies used picture-based paradigms to investigate semantic competition in complex phrase comprehension and production, and their relationships with individual cognitive skills. Study 1 tested the hypothesis that resolution of semantic competition when comprehending and producing complex phrases would depend upon similar cognitive resources (i.e., vocabulary knowledge and inhibition) in adults. The results showed evidence of semantic competition in both tasks, and the degree of competition was linked to animacy configuration and semantic reversibility between noun concepts (whether the nouns are both animate entities and may share agent/patient roles). It was found that participants experienced greater difficulties when producing and comprehending complex phrases containing two animate nouns with highly reversible roles: they produced more agentless passives (prefer easier structures in production) and also took longer to comprehend. Also, in two RT measures we collected for production performance (which was not assessed in previous studies), SOT and verbal fluency, were significantly poorer for high-reversibility phrases as compared to low-reversibility phrases. Together, these findings suggested that semantic similarity and reversibility between noun concepts elicit competition during sentence comprehension *and* production. Further to this, we found common vocabulary and homograph inhibition skills underpinning individual's sensitivity to comprehension and production competition, and a motor inhibition skill only contributed to production performance in addition to vocabulary and homograph inhibition influences, highlighting common as well as distinct cognitive processes and resources.

Study 2 adapted the paradigm for children and adolescents by reducing the number of testing trials and using simpler pictures and homograph words. Similar to Study 1, we observed the same animacy and reversibility-based effects on comprehension RTs and preference for passives in production, but also the data illustrated some age-related differences across the three age groups. Despite the condition effects, there was a general developmental improvement in comprehension times, production accuracy and fluency, as older participants were faster in processing and also produced more accurate and more fluent descriptions. The sizes of animacy effects also varied across age groups in production: young children were the only age group showing poorer fluency for describing events with two animate entities as compared to those with one animate and one inanimate entity; and as compared to adolescents, young children produced fewer passives in the animate condition, and their agent omissions showed no effect of animacy. This suggested that despite young children being sensitive to the degree of semantic competition in processing, they differed from older speakers in the size of production interference they experience and their strategic use of production options (use of passives, agent omission) to alleviate planning interference. Critically, as compared to adults, we identified different kinds of cognitive skills in predicting young children's language performance: WM capacity explained both comprehension RTs and production fluency. This supports the capacity-constraint account (Just & Carpenter, 1992; Swanson, 1996), which suggests that WM storage and processing capacity play an important role in children's language processing.

Study 3 investigated the time course of semantic competition in adults' phrase production using eye-tracking methodology. Given that previous findings have reported that resolution of semantic competition in comprehension operates at the point in which the verb is encoded, this study aimed to examine whether verb planning (particularly when establishing syntactic functions at the verb) also plays an important role in producing phrases containing competitive nouns. It was found that fixations on agents and patients differed between high and low-reversibility phrases at the point where the syntactic roles were considered for alternative nouns, i.e. before the utterance of the verb. Less reversible nouns engendered less competition, resulting in clear divergence of fixations between agent and patient, and led to earlier encoding of the subsequent by-phrase agent. The results suggest that, similar to previous comprehension findings, the semantic reversibility between the noun concepts played a role during verb planning. Moreover, complementary to the findings of Study 1, the likelihood of fixating the target/patient was predicted by speakers' homograph inhibition skill, particularly for the high-reversibility items. This suggests that inhibition processes were engaged in the homograph task and are recruited for maintaining attention to the target when resolving production competition.

Taken together, these findings contribute to our understanding of the relationship between comprehension and production in several ways and go beyond the simple claim proposed by many language models (e.g. the dual-path model) that similar representations or prediction-related mechanisms play a role in both processes. Below we sketch out the theoretical implications of our findings for models of comprehension and production in mature and developing language systems.

5.2 Implications for comprehension processes in adults

Our results suggest that sentence comprehension involves some form of competition, at least in complex structures. Unlike reading research focusing on the role of working memory capacity in comprehension (Caplan & Waters, 1999; Marcel A Just & Carpenter, 1992; Tan, Martin, & Van Dyke, 2017; J A. Van Dyke & Johns, 2012; Julie A Van Dyke et al., 2014), the present studies follow a growing body of research exploring how executive or cognitive control measures operate on sentence comprehension. Working memory researchers for example have emphasized the role of retrieval and working memory interference in sentence comprehension (Lewis, Vasishth, & Van Dyke, 2006; J A Van Dyke & McElree, 2006), whereas others point to the role of general cognitive control mechanisms in sentence comprehension (Hsu & Novick, 2016; Novick et al., 2013; Nozari, Trueswell, & Thompson-Schill, 2016). Thus, a growing body of evidence suggests that some aspects of executive functions must operate in comprehension, particularly when this process becomes difficult and less automatic or consistent with typical predictions.

It is nevertheless too early to pinpoint a specific control mechanisms involved in sentence comprehension, in part because many executive tasks correlate with each other, and at the same time, no task is a pure measure of inhibition or interference, instead involving additional processes (Friedman & Miyake, 2004; Miyake et al., 2000). Consider for example, working memory interference. This concept is linked to difficulty in cue-dependent retrieval when distractors are also available (J A. Van Dyke & Johns, 2012). In the homograph task, participants first read out a homograph in its dominant meaning (e.g., *wind* in the context of *blow*), and in a later trial occurred within a short time period, they read out the same written form in the context of a different meaning (e.g., *turn).* At this point, they might be primed to pronounce the word in the dominant way, but the context indicates a less-frequent meaning and pronunciation. Arguably, the inhibition of a prepotent response must take place. Yet, it is possible to conceptualize this process as working memory interference. The second context is a cue to retrieve the less dominant meaning/pronunciation, which surely competes with the earlier and highly available meaning/pronunciation.

A clue to what specific process might be relevant in comprehension can be inferred from differences between the homograph and homonym inhibition tasks. The homonym task asked participants to decide whether a word was related to the previously presented context-word. When subordinate meanings were targeted, the previously computed dominant meaning of a word should have been inhibited. However, this measure did not predict comprehension performance, as homograph inhibition did. This superior sensitivity of the homograph task might stem from its being more cognitively demanding, compared to the homonym task, as it requires inhibition at two different levels of linguistic information. Indeed, the differences between subordinate and dominant meanings were generally larger in the homograph than in the homonym task, suggesting more difficulty. Moreover, deciding whether two words are related might not require as much inhibition as stopping a prepotent pronunciation and selecting a specific meaning, since semantic relatedness can be judged even if the two meanings are simultaneously entertained. Thus, it might be the specificity of the selection process in the homograph task that is relevant for complex phrase comprehension, particularly, when highly reversible animate nouns are involved.

Previous accounts of comprehension difficulty in these structures are consistent with this possibility. In comprehending phrases such as *the man that the girl is hugging*, there is difficulty not only in maintaining the two nouns in memory but also in establishing who is acting on whom, particularly when the two nouns are equally good candidates for the agent/subject of the verb (Gennari & MacDonald, 2009; Gordon et al, 2001, 2006; Humphreys et al, 2016). The nouns thus compete with each other and one must be selected to link to the appropriate syntactic and semantic role with the verb. This process requires the selection of one noun and inhibition of the alternative one, a process that is consistent with the specific selection and inhibition suggested by the homograph tasks. It is also broadly compatible with models of sentence comprehension positing some form of competition or interference among alternative cues, such as working memory and probabilistic constraint based models (Lewis & Vasishth, 2005; Lewis et al., 2006; MacDonald et al., 1994; McRae, Spivey-Knowlton, & Tanenhaus, 1998).

5.3 Implications for production processes in adults

Our results suggest that phrase production engages motor and semantic-phonological inhibition processes, particularly in the animate-target and high-reversibility condition. This possibility is consistent with much production research reporting interference or competition at various levels of linguistic representation such as lexical and structure competition (G S Dell & O'Seaghdha, 1994; V. S. Ferreira, 1996; Konopka, 2012; Meyer, 1996; Shao, Meyer, & Roelofs, 2013; Slevc, 2011), and particularly, in the production of referential expressions (Allum & Wheeldon, 2009; Arnold, 2010; Arnold & Griffin, 2007; Fukumura & van Gompel, 2010; Fukumura, van Gompel, Harley, & Pickering, 2011; Konopka, 2012; Smith & Wheeldon, 1999, 2004). For example, when two conceptually similar nouns are planned in the same phrase (e.g., *the hammer and the axe are moving up*), they elicit longer speech onset times due to semantic interference or competition, as found here for phrases containing two animate nouns (Smith & Wheeldon, 1999, 2004; Konopka, 2012). In phrases containing subordinate clauses, however, speakers are not restricted to produce two nouns in a sequence as in noun conjunctions and can flexibly opt for an alternative structure. Competition between similar animate nouns thus results in the inhibition of one noun and the selection of the other to be uttered first, which is then made the subject of the upcoming verb in a passive structure. This competition explains why animate-target phrases are overwhelmingly produced in passives rather than actives and their agents omitted (study 1). Competition also played a role within animate items as a function of reversibility (study 3): the speech onset and verb planning times were longer, fluency was poorer, and fixations on patients were more likely in the high-reversibility than low-reversibility condition.

Although inhibition skills were associated with producing animate-head and high-reversibility passives as expected, they also predicted production performance with inanimate-head phrases, suggesting a role for inhibition in this condition too (study 1). Within the context of production research, this finding is not particularly surprising because competitive processes e.g., lexical and structural competition, are an inherent aspect of production regardless of animacy. In visual contexts in particular, the presence of alternative characters make referential targets less accessible (Arnold & Griffin, 2007; Fukumura & van Gompel, 2010; Fukumura et al., 2011). In our task, when talking about inanimate objects (e.g. *the teddy bear* being hugged by the girl), speakers not only had to inhibit the more conceptually and visually salient animate character interacting with it (e.g. *the girl*), but likely also alternative structures such as agent-initial phrases, e.g. *the girl (that's) hugging the teddy bear*, which are the most frequent phrase structure in the language (Roland et al., 2007).

Several pieces of evidence support these inhibition processes. First, when describing inanimate-targets after event apprehension, speakers fixate on agents as much as the object-target before speech (Humphreys et al., 2016), suggesting that there might be uncertainty regarding which character to name first. Second, in study 1, we found more errors for inanimate-target than animate-target phrases and these were cases in which speakers incorrectly naming the agent first rather than the highlighted object. Third, inanimate objects are not only less salient than human agents, but their names were less frequent and accessible, thus favouring the initial naming of agents (J K Bock, 1987; J K Bock & Levelt, 1994; J K Bock et al., 1992; J K Bock & Warren, 1985). Therefore, some degree of competition in planning complex phrases surely occurred in the animate-target condition independently of reversibility-induced conflict: Speakers must access less accessible forms, while avoiding interference from competing characters and alternative agent-first plans. This interference explains why participants with poor inhibition overwhelmingly used passives in all conditions, as they experience more difficulty overall. The difference between animate and inanimate-target phrases therefore does not depend on whether or not competition occurs, but in the degree and nature of the competition taking place.

Finally, phrase production also appears to have recruited motor inhibition processes, as indicated in significant interactions between STOP-IT performance and animacy or reversibility conditions (studies 1 & 3). Note that the STOP-IT task not only requires motor inhibition but also monitoring and flexible adjustments to competing task demands, namely, responding as fast as possible and stopping the response when instructed at varying delays. Indeed, performance in this task correlates with interference control more generally (Friedman & Miyake, 2004). Thus, the present results may indicate not only motor inhibition processes, but also more general executive functions such as monitoring and resistance to distracting interference, which may operate at different levels of linguistic representations (phonology, syntax and semantics). Nevertheless, more sentence production research is clearly needed to more thoroughly elucidate the nature of the executive functions involved.

5.4 Implications for the relationship between comprehension and production in adults

Previous models of sentence production and comprehension have suggested a common architecture for production and comprehension. The proposed models differ substantially in their architectures, although they agree that production and prediction in comprehension are intimately linked. The production-as-covert-simulation model argues that in comprehension, the listener covertly imitates what has been heard and engages the simulator and production implementer to predict the next word (Pickering and Garrod, 2007, 2013, Pickering and Mani, 2018). The dual-path architecture, in contrast, does not assume covert imitation and a production implementer distinct from the comprehension implementer. Instead, it implements word-by-word production and comprehension after learning in such a way that the prediction of the next word in comprehension is the same process as the prediction of the next word in production (Chang et al, 2006; Chang & Dell, 2014). In both processes, the sequencing pathway is involved, and its hidden layer modulates prediction by acting as the repository of what has been learned up to that point. Unlike the covert-simulation model, this architecture thus embodies the common knowledge base and prior experience underpinning production and comprehension. However, both models aim to account for prediction rather than processing difficulty, and mostly specify the mechanisms taking place in relatively easier processes, e.g., when linguistic elements are highly predictable or primed from the context. It therefore remains unclear what mechanisms operate in difficult processing, e.g., in cases of conflict between production choices and interpretations, and importantly, whether competition resolution mechanisms would be shared across sentence production and comprehension.

The present work addressed this possibility and suggests that the same inhibition-related cognitive skills underpin both production and comprehension of complex referential phrases. To account for this commonality, it might be possible to extend the existing production-comprehension models so that comprehension difficulty emerges from conflicting alternative expectations and/or the mismatch between predicted and actual input, as argued by probabilistic models of comprehension. For example, comprehending complex active phrases starting with *the man that* involves predicting alternative continuations consistent with the most frequent phrase pattern in the language, where the first noun is the agent/subject of an upcoming verb, thus eliciting conflict with the incoming input when a noun is heard (Gennari & MacDonald, 2008, 2009). When two animate nouns have been heard in *the man that the girl…* expectations might conflict with each other and verb interpretation is difficult, accounting for reversibility effects. Such processes only occur for animate targets but not inanimate ones, thus accounting for animacy effects. This putative extension requires that the notion of prediction is understood as probabilistic and experience-based, rather than an all-or-nothing process, e.g., one in which single elements are correctly anticipated.

However, the challenge for production-comprehension models is to specify how the production system would generate alternative predictions and account for production difficulty as well. The production-as-covert-imitation model was not designed to account for such processes, as it is mostly focused on explaining prediction in comprehension, and often, in highly probable cases. The dual-path model in contrast, could potentially explain these processes because it can account for production choices as a function of prior learning (F Chang, 2002; G S Dell, Chang, & Griffin, 2001). Note that the choices of active vs. passive in complex phrase production has been shown to be related to reading experience and the probability of the syntactic-semantic structures in the language, including animacy configurations (Gennari & MacDonald, 2009; J L Montag & MacDonald, 2015). Thus, the dual-path has the potential to explain how learning from linguistic experience modulates the prediction of alternative continuations and how conflicting alternative may elicit difficulty in both production and comprehension. This possibility is consistent with many separate production and comprehension computational models that explain competition in various domains (Gary S Dell, Burger, & Svec, 1997; Gary S Dell, Chang, & Griffin, 1999; Gary S Dell, Juliano, & Govindjee, 1993; Fitz et al., 2011; McRae et al., 1998; Nozari, Dell, & Schwartz, 2011), including inhibition mechanisms in production (Gary S Dell, Oppenheim, & Kittredge, 2008; Oppenheim, Dell, & Schwartz, 2010). It is therefore possible that predictive processes operate in both production and comprehension, sometimes generating conflicting alternatives compatible with the input, which are resolved on the bases of available evidence in a given context and prior language experience. Relatively easy as well as difficult sentences may thus be explained by the same language architecture.

Additional evidence for a common language architecture comes from neurobiological researcher arguing that at least verbal inhibition mechanisms are implemented in prefrontal cortex, and particularly, the left inferior frontal gyrus (Gennari, MacDonald, Postle, & Seidenberg, 2007; Hsu, Jaeggi, & Novick, 2017; Humphreys & Gennari, 2014; January et al., 2009; Martin, 2005; Novick, Trueswell, & Thompson-Schill, 2005a, 2010; Spalek & Thompson-Schill, 2008; Thothathiri, Schwartz, & Thompson- Schill, 2010; Whitney, Kirk, O'Sullivan, Lambon Ralph, & Jefferies, 2011). This region is part of the language processing network (Catani, Howard, Pajevic, & Jones, 2002; Hickok & Poeppel, 2007) and is engaged in both production and comprehension of sentences and words, particularly when comprehension involves resolution of ambiguities or the integration of multiple semantic features (Thompson-Schill, Bedny, & Goldberg, 2005).

However, our findings also indicated that production was additionally predicted by motor-related inhibition or action control, suggesting that some aspects of competition in production are not necessarily verbal in nature, and might be part of more general action execution processes. This possibility aligns with some neuro-biological evidence in that the production network appears to be larger than the comprehension network involving not only subcortical regions but also additional motor regions such as supplementary motor cortex (de Zubicaray, McMahon, Eastburn, & Wilson, 2002; Humphreys & Gennari, 2014) and the anterior cingulate cortex thought to underpin speech and action monitoring (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Nozari et al., 2011). Thus, it is likely that production engages additional mechanisms not present in comprehension and that these mechanisms are shared with other non-verbal tasks. From this perspective, production-comprehension models appear underdeveloped, as not all aspects of production are engaged in comprehension.

Similarly, our results also highlight that processing difficulty in part depends on individual differences in vocabulary and inhibition skills. These differences are linked to prior linguistic experience, prior practice with inhibition-related task (Hussey et al., 2017) and neuro-biological factors (e.g., as in language impairments). The covert-simulation account argues that in some cases such as children and old adults, prediction-by-simulation does not occur, thus potentially accounting for individual differences. However, the optionality of this prediction leaves unexplained the nature of the relationship between production and comprehension in many speakers. In contrast, the dual-path model can potentially vary the learning (the network training) and aspects of its architecture to begin to understand individual differences in processing. The explanatory power of such models for individual differences, however, remains to be determined.

In sum, the present results indicate that shared verbal inhibitory skills, as measured by homograph inhibition, underpin sentence production and comprehension, but that production additionally involves motor inhibition or control processes that appear task-specific. Current models of the relationship between production and comprehension are not sufficiently developed to account for common verbal inhibition as well as distinctive production mechanisms, which may in turn be shared with other cognitive processes such as action planning. However, common inhibition mechanisms are compatible with the dual-path architecture and connectionist models more generally, in that prior experience with the language modulates the conflicting alternative representations. Together with prior independent models of production or comprehension, our results instead suggest that competitive mechanisms may be intrinsic to the language system and thus integrative language processing models should go beyond a common system for prediction of upcoming elements and include competitive processes.

5.5 Implications for developmental processes in comprehension

Our results suggest that, as in adults, children's comprehension of complex phrases also involves some form of semantic competition. The degree of competition is modulated by animacy configuration and semantic reversibility between the noun concepts. This is consistent with previous findings which reported just like adults, young children are also sensitive to a number of syntactic-lexical constraint (such as feature similarity between noun concepts, e.g. animacy, noun phrase type) when they process complex phrases (Arnon, 2010; Arosio et al., 2011; Corrêa, 1995). On the other hand, due to the use of child appropriate stimuli, we found fast comprehension performance in adolescents that showed little differences across conditions. Since our adults in Study 1 did find animate-target phrases difficult to process, it is possible that our adolescents did not find animate items challenging, because they were not tested with many high-reversibility items where both nouns were equally good candidates for the event agent/patient (e.g. the man that the woman is hitting). Children, on the other hand, may have been more affected by explicit lexical features*, e.g. animacy*, in making their decisions as opposed to other semantic-syntactic features older participants would pay attention to, *e.g. noun-verb relationship*. Hence, they found phrases containing "the dog, the woman" more interfering than those containing "the book, the woman", despite both noun combinations implying low reversibility in agent-patient roles. This is consistent with the view that as compared to adults, children are generally less sensitive and skilled in using certain cues and strategies (e.g. temporal connectives, referential context, top-down context use) to facilitate comprehension, as these cues and skills are cognitively demanding and late developing. (Blything & Cain, 2016; Khanna & Boland, 2010; Snedeker & Trueswell, 2004).

Given that the size of the animacy effect is common to all children despite the relatively extended age range, the change in how animate-head active phrases are processed may occur right before the period of young adolescents. This may reflect ongoing development of children's knowledge of different syntactic-lexical cues in comprehension, driven by increasing exposure to these features in complex structures across contexts. We failed to find a relationship between comprehension and our reading experience measure, maybe because the structures tested in our experiment do not match the most frequent types of active phrases in children's text input, which often contains inanimate heads and pronoun subjects (e.g. *the book I read*) (Montag & MacDonald, 2015). Children's mastery of the unusual animate-head phrases may instead owe to exposure with other sources of language input (e.g. adult-directed speech, academic texts), and also not necessarily experience with the same sentence forms, but "neighbouring" forms that share similar noun combinations. Future research is needed to identify the contributions of different sources of linguistic experience.

Although the size of the animacy effect was not predicted by any measures of individual difference tested, we found young children's processing speed of all active phrases was predicted by individual memory capacity and semantic-phonological inhibition skills (homograph inhibition). The former relationship is in support of the memory-capacity account of sentence comprehension, which attributes processing difficulty to the memory capacity of the individual (e.g. Just & Carpenter, 1992; Swanson, 1996). The latter relationship is consistent with previous findings suggesting that during processing temporarily ambiguous sentences, children's inability to revise incorrect representations is linked to their immature inhibition/shifting skills (Choi & Trueswell, 2010; Novick, Trueswell, & Thompson-Schill, 2005b; Woodard et al., 2016). Hence, we interpret these relationships as reflecting extra time needed to construct and/or revise a mental representation in children with low memory capacity and inhibition skills. That is, these children are less capable of maintaining two nouns in working memory before establishing who is acting on whom. It is also possible that this result reflects children's preference for processing canonical word order (Slobin & Bever, 1982), so that non-canonical structures (e.g. *the man that the girl is hugging*) are initially analysed as canonical Subject-Verb-Object (e.g. the man *doing something*) until they hear the second noun. Children must revise their initial interpretation, which requires extra time for revision if they have low memory capacity or inhibition skills.

These correlations also suggest a role for memory capacity and inhibition for inanimate-head phrases, a form that is assumed to be non-demanding and non-interfering in adult comprehension. As previously discussed, it may reflect young children's lack of linguistic experience and knowledge with interpreting the unusual active structure. One possibility, as proposed by locality dependency theory (Gibson, 1998), it might simply be more difficult for children to link the subordinate verb with the head-noun due to intervening material between them, i.e., another noun. More distant structural relations impose more working memory load, and this would account for a general difficulty with complex object-extracted structures. On the other hand, they would simply have more experience with structures in which nouns and verbs occur close to each other, and thus a combination of memory load and lack of experience may account for their performance.

5.6 Implications for developmental processes in production

We found that for speakers from all age groups, their tendency to produce passives is influenced by animacy and reversibility-based competition between noun phrases, which is consistent with one prior study that reported animacy effects in both children (8 and 12-year-olds) and adults' production of passives (Montag & MacDonald, 2015). This suggests that children as young as 8 years old displayed an appreciation of the functional differences between actives and passives and are capable of using passive structures to reduce planning burden. However, when focusing on the *overall rate of passive usage* or a fine-grained production option, *agent omission*, it seems that children before the age of 14 weight these factors somewhat differently as compared to adolescents.

At first, we found children produced significantly less passives (thus more actives) for animate-targeted pictures as compared to adolescents. Given that our participants were primed with active structures in the comprehension task, this may reflect a greater vulnerability to priming effects in children. Syntactic priming has long been regarded as an implicit method to probe into the nature of speakers' syntactic representations, as repetition of sentence structures without overlap of lexical items suggests they have representations of these structures independent of lexical content (Branigan, 2007; Pickering & Branigan, 1999; Pickering & Ferreira, 2008). Consistent with other studies reporting stronger priming effect in less competent language users such as children (Branigan & Messenger, 2016; Messenger, Branigan, & McLean, 2011), non-native speakers (Flett, 2006; Flett, Branigan, & Pickering, 2013) and aphasics (Hartsuiker & Kolk, 1998), this finding may suggest children have weaker preferences for or fewer available syntactic representations when describing transitive events, owing to less exposure and knowledge with alternative structures. Thus, they are more susceptible to effects of recent linguistic experience and produced more primed actives as compared to adolescents.

Secondly, we found children did not omit more agents in describing animate-head phrases as compared to inanimate-head phrases. Adolescents, on the other hand, appeared to omit agents more strategically: they are more likely to drop the agents when describing animate-head passives, where the semantic competition is more pronounced as compared to when the head noun is inanimate. Again, this is possibly due to the lack of linguistic experience and knowledge with alternative production options (such as agent omissions) in children, hence they tend to mention all the characters in their descriptions, following the same principle of the primed active structures (where all characters were mentioned in a given description) in the comprehension task.

Another focus of the children's study was to examine whether the size of semantic competition in production varies with age and individual cognitive abilities. Although in young children, the animacy effect interacted with WM capacity, older children's performance was not related to individual difference measures, except for a simple correlation with vocabulary in adolescents. This might be because we did not have sufficient items and participants (statistical power) to observe effects in the older groups. If the members of a given age-group do not sufficiently differ from each other on individual measures, these measures are unlikely to predict performance. Nevertheless, the fact that young children's verbal fluency for animate-head passives was hindered relative to their inanimate-head descriptions suggests greater difficulty and semantic competition in young children.

Also, reversibility interacted with individuals' digit span: young children with better verbal working memory capacity showed smaller differences between conditions. Although there is less evidence suggesting a role of memory capacity in children's sentence production, compared to comprehension, memory capacity should also influence production performance as information must be maintained in WM before previously planned elements are being outputted, and elements in WM are therefore be susceptible to interference. Consistent with this idea, one recent production study with children has reported higher memory capacity (measured by digit span task) is associated with better performance in producing sentences placing additional load in WM, such that those describing events in reverse chronological order, in which speakers must maintain information about the first occurring event in WM during planning, e.g. *He ate the burger, after he put on the sandals* (Blything & Cain, 2019). Similarly, in our study, WM played a role as young children had a tendency to mention/plan all the characters in their descriptions as compared to older speakers, hence are less likely to completely inhibit the agent noun when planning the sentence subject, and planning of animate-head descriptions engendered greater competition and increased the memory resources required for sentence production.

Finally, adolescents' production fluency was predicted by vocabulary measures, suggesting an influence of lexical knowledge in production. However, the absence of such a relationship in younger age groups and in comprehension RTs contrasts with previous findings reporting significant influences of linguistic knowledge in language acquisition (Lee, 2011; Ouellette, 2006). Nevertheless, other studies with children have failed to report correlations between measures of vocabulary and language abilities (e.g., Blything & Cain, 2016; de Ruiter, Theakston, Brandt, & Lieven, 2018). Such inconsistencies may stem from across-experiment differences in the vocabulary assessments used, the type and complexity of sentence structures tested, and the manner in which comprehension and production performance are assessed.

5.7 Implications for the development of comprehension and production

Our results are consistent with many findings that even young children have abstract knowledge of complex syntactic structures, which is adopted in both comprehension and production processes (Messenger, Branigan, & McLean, 2012; Messenger & Fisher, 2018). However, they are still far-from achieving adult-like performance. In comprehension, the animacy effect was pronounced for all children but processing times increased with age. In production, some aspects of passive structure (agent omission) may be more difficult to master than others and acquired later in development. This provides supporting evidence for one study suggesting that acquisition of passives is a staged process, in which understanding of non-canonical thematic-role mapping is acquired later than knowledge of constituent structure (Messenger et al., 2012).

Also, as in adults, our results provided evidence for the role of domain-general processes across different modalities, as memory capacity was found to underpin both comprehension and production processes in young children, consistent with previous L2 (second language) studies reporting WM capacity is positively related to both comprehension and production outcomes (for a meta-analysis, see Linck, Osthus, Koeth, & Bunting, 2014). Thus, it seems that memory capacity plays an important role in individuals' language proficiency, and this also explains why the influence of WM was absent in older participants who have more linguistic experience and knowledge. Of particular note, we selected a verbal WM measure with minimum linguistic demands to better disentangle the effects of memory and language, as other measures with semantic content (e.g. reading/listening span) should be more strongly related to children's language performance. However, researchers are becoming increasingly aware of the association between WM capacity and long-term linguistic knowledge, and a thorough WM explanation needs to consider how limited WM interfaces with long-term knowledge of language. For example, Boyle, Lindell, and Kidd (2013) included measures tapping different components of WM, and found an episodic buffer measure (a temporary store which receives input from long-term memory and WM components, e.g. sentence repetition task) was a stronger predictors of children's comprehension of complex structures, as compared to a measure of central executive (responsible for directing and allocating attention and resources of the WM system, e.g. backward digit span). This suggests that the contribution of WM to comprehension should also reflect differences in children's long-term linguistic knowledge, instead of only capacity limits. Thus, future research should include more complex measures of memory to provide a more accurate assessment when examining the relationship between memory and children's language skills.

However, this is not to dismiss the asymmetry between comprehension and production, as it has long been recognised and debated in language acquisition research: comprehension performance lags behind production performance (Diessel, 2004; Grimm, Müller, Hamann, & Ruigendijk, 2011). Some have argued that in adults, the mechanisms underlying comprehension and production processes have become closely aligned; but in children, the same syntactic form may be processed differently. One particular example is related to children's underdeveloped perspective-taking skills: it was found that comprehension of Turkish morphological forms is delayed by the difficulty of reasoning about other people's information resources, and which is not necessarily recruited for production in children, as they often plan with their own perspectives and sources (Ünal & Papafragou, 2016). Thus, it suggests an inherent perspective-taking asymmetry between comprehension and production in language acquisition. In our study, perspective-taking may be less likely to contribute to the differences in cognitive demands imposed by comprehension and production, because our children were encouraged to describe the target character in a way in which the experimenter/listener can distinguish it from the remaining competitors in the scene, and their descriptions were always corrected if they fail to do so. Nevertheless, it would be interesting to see whether similar pattern of results arise in other aspects of language use (e.g. use of personal pronouns) and in nature interaction, and more importantly, to what extent the emergence of a comprehension and production asymmetry is related to recruitment of domain-general processes and linguistic knowledge.

5.8 General conclusions and future directions

In sum, the studies presented in this thesis indicate that verbal inhibitory skills as measured by homograph inhibition underpin adults' sentence comprehension and production (evident in both eye-tracking and behavioural data), thus suggesting that a common inhibitory mechanism supports production and comprehension beyond the role of vocabulary knowledge. Importantly however, production additionally recruited motor inhibition or control processes that were not recruited for comprehension of the same linguistic material. Current models of the relationship between production and comprehension cannot account for both common verbal inhibition across production and comprehension as well as distinctive production mechanisms, which may in turn be shared with other cognitive processes such as action planning. Together with prior independent models of production or comprehension, our results instead suggest that competitive mechanisms may be intrinsic to the adult language system and thus integrative language processing models should go beyond a common system for prediction of upcoming elements and include competitive processes. However, our interpretations regarding production processes are sometimes speculative, especially regarding the eye-movement data, as previous investigation of on-line sentence production is scarce. In the current study, we mainly examined production of the most frequently produced structure: passives. Clearly, more research targeting other alternatives and using on-line methodologies (including eye-tracking, EEG) is needed to thoroughly elucidate the nature of the cognitive processes recruited by production. Investigation of less frequently produced forms could be achieved by using the syntactic priming paradigm and examine whether the degree of priming effects is predicted by individual cognitive differences or linguistic knowledge.

On the other hand, our findings with children take a first step to support engagement of common domain-general process in young children's comprehension and production of complex phrases: the underdevelopment of WM at early ages is associated with poorer language performance. Together with adults' results, this suggests a discontinuity of cognitive functioning from childhood to adulthood, such that WM influence is more strongly implicated during early rather than late stages of language development, where other mechanisms and influences come into play. For instance, the influence of vocabulary knowledge and inhibition-based processes (particularly for the more difficult condition) may be more likely to be observed once individuals have acquired adequate knowledge of syntactic parsing and have gained increased exposure to different structures (we did report a vocabulary effect in adolescents' production fluency). Future research, with more statistical power, is much needed to further investigate the extent of these influences in older children and adolescents' language performance. Also, the present design (cross-sectional design) cannot permit conclusions to be made regarding causality. Training studies and large-scale longitudinal investigation are also needed to examine the development of cognitive control processes as predictors for later comprehension and production outcomes. Finally, we emphasize the need for future studies to test the generalization of the results with more and different measures of memory, linguistic knowledge and experience, etc.

Appendices

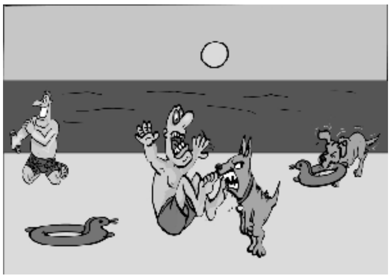
Appendix A: Complex phrases and corresponding picture items used in Study 1-2

Item	Condition	Reversibility Rating	Reversibility Group	Comprehension sentence	Items used for children
bite	animate	2.383458647	Medium	The man that the dog is biting.	Yes
bite	inanimate			The rubber ring that the dog is biting.	Yes
brush1	animate	3.954887218	Low	The dog that the girl is brushing.	Yes
brush1	inanimate			The car that the man is brushing.	Yes
brush2	animate	3.576441103	Low	The horse that the girl is brushing.	
brush2	inanimate			The suit that the man is brushing.	
carry1	animate	4.957393484	Low	The child that the woman is carrying.	
carry1	inanimate			The box that the man is carrying.	
carry2	animate	3.609022556	Low	The sheep that the farmer is carrying.	
carry2	inanimate			The ladder that the man is carrying.	
carry3	animate	4.904761905	Low	The baby that the woman is carrying.	Yes
carry3	inanimate			The books that the woman is carrying.	Yes
film1	animate	1.807017544	Medium	The boy that the woman is filming.	Yes
film1	inanimate			The statue that the woman is filming.	Yes
film2	animate	3.350877193	Low	The dog that the woman is filming.	
film2	inanimate			The car that the woman is filming.	
hit1	animate	0.007518797	High	The player that another player is hitting.	Yes
hit1	inanimate			The ball that a player is hitting.	Yes
hit2	animate	-0.563909774	High	The man that another man is hitting.	
hit2	inanimate			The gong that a man is hitting.	
hit3	animate	1	Medium	The boy that the girl is hitting.	
hit3	inanimate			The gong that the girl is hitting.	
hold1	animate	2.155388471	Medium	The girl that the woman is holding.	
hold1	inanimate			The vase that the woman is holding.	
hold2	animate	5.308270677	Low	The baby that the woman is holding.	
hold2	inanimate			The ball that the boy is holding.	
hug1	animate	-0.055137845	High	The man that the girl is hugging.	
hug1	inanimate			The teddy bear that the girl is hugging.	
hug2	animate	4.568922306	Low	The dog that the girl is hugging.	Yes
hug2	inanimate			The teddy bear that the girl is hugging.	Yes
kick	animate	-0.766917293	High	The girl that the boy is kicking.	Yes
kick	inanimate			The ball that the boy is kicking.	Yes
kiss1	animate	0.438596491	Medium	The girl that the woman is kissing.	
kiss1	inanimate			The trophy that the girl is kissing.	
kiss2	animate	-0.706766917	High	The boy that the girl is kissing.	
kiss2	inanimate			The picture that the girl is kissing.	
lick	animate	2.709273183	Medium	The baby that the dog is licking.	Yes
lick	inanimate			The guitar that the dog is licking.	Yes

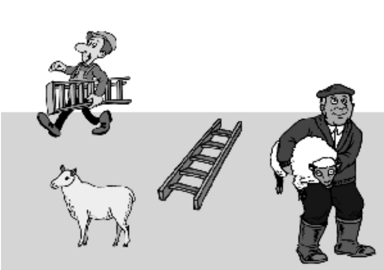
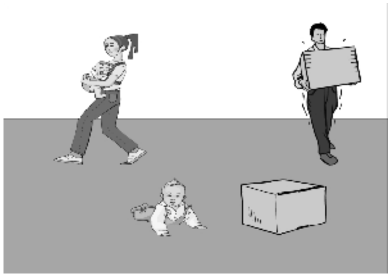
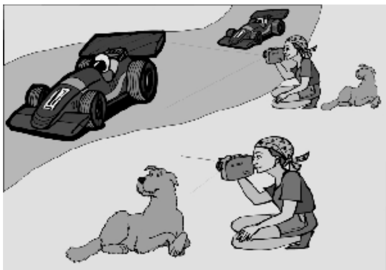
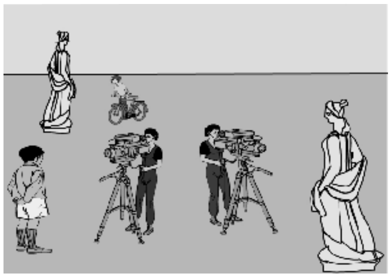
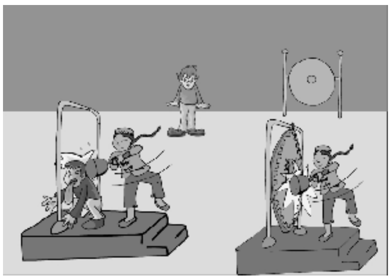
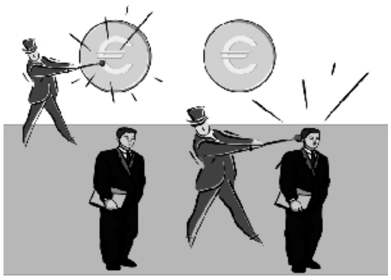
lift1	animate	2.967418546	Low	The boy that the man is lifting.	
lift1	inanimate			The weight that the man is lifting.	
lift2	animate	2.69924812	Medium	The girl that Santa is lifting.	
lift2	inanimate			The present that Santa is lifting.	
pinch	animate	2.097744361	Medium	The man that the lobster is pinching.	Yes
pinch	inanimate			The rubber ring that the lobster is pinching.	Yes
pull1	animate	0.280701754	Medium	The boy that another boy is pulling.	Yes
pull1	inanimate			The truck that the boy is pulling.	Yes
pull2	animate	0.072681704	High	The girl that the man is pulling.	
pull2	inanimate			The suitcase that the man is pulling.	
punch	animate	2.070175439	Medium	The man that the woman is punching.	Yes
punch	inanimate			The punch bag that the woman is punching.	Yes
push1	animate	-0.837092732	High	The man that another man is pushing.	Yes
push1	inanimate			The pram that the girl is pushing.	Yes
push2	animate	-0.07518797	High	The man that the woman is pushing.	
push2	inanimate			The trolley that the man is pushing.	
push3	animate	-0.030075188	High	The man that the boy is pushing.	
push3	inanimate			The trolley that the boy is pushing.	
scratch	animate	1.120300752	Medium	The man that the cat is scratching.	Yes
scratch	inanimate			The sofa that the cat is scratching.	Yes
shoot1	animate	0.195488722	High	The man that the woman is shooting.	
shoot1	inanimate			The target that the woman is shooting.	
shoot2	animate	4.751879699	Low	The deer that the man is shooting.	Yes
shoot2	inanimate			The apple that the man is shooting.	Yes
splash	animate	0.223057644	High	The boy that the girl is splashing.	Yes
splash	inanimate			The duck that the girl is splashing.	Yes
spray1	animate	-0.43358396	High	The woman that the man is spraying.	
spray1	inanimate			The statue that the man is spraying.	
spray2	animate	2.967418546	Low	The dog that the man is spraying.	
spray2	inanimate			The car that the man is spraying.	
spray3	animate	0.087719298	High	The man that the woman is spraying.	
spray3	inanimate			The flowers that the man is spraying.	
squirt	animate	0.167919799	High	The woman that the man is squirting.	
squirt	inanimate			The target that the man is squirting.	
step	animate	2.393483709	Medium	The man that the dog is jumping on.	Yes
step	inanimate			The mat that the boy is jumping on.	Yes
stroke	animate	4.714285714	Low	The dog that the woman is stroking.	Yes
stroke	inanimate			The teddy that the girl is stroking.	Yes
tie	animate	0.280701754	Medium	The man that the boy is tying up.	Yes
tie	inanimate			The shoes that the boy is tying up.	Yes
touch	animate	0.586466165	Medium	The man that the woman is touching.	Yes
touch	inanimate			The computer that the woman is touching.	Yes

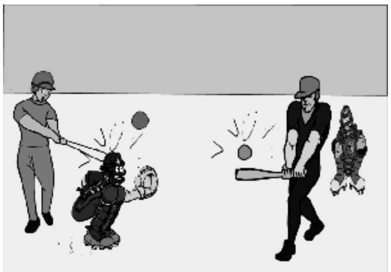
wash1	animate	4.365914787	Low	The dog that the girl is washing.	Yes
wash1	inanimate			The car that the man is washing.	Yes
wash2	animate	4.398496241	Low	The baby that the woman is washing.	
wash2	inanimate			The dish that the man is washing.	

Picture items



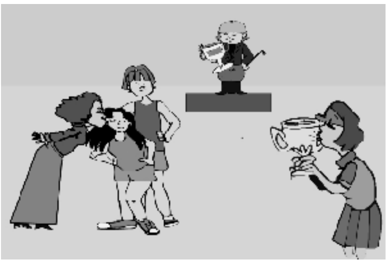
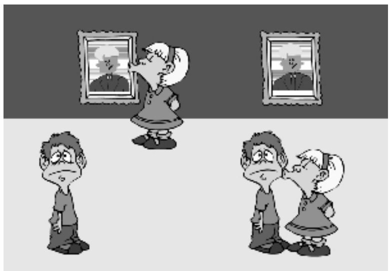
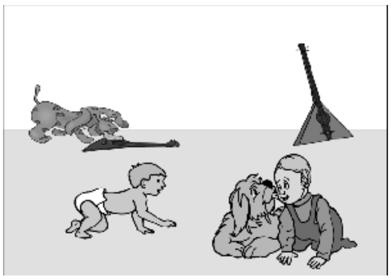
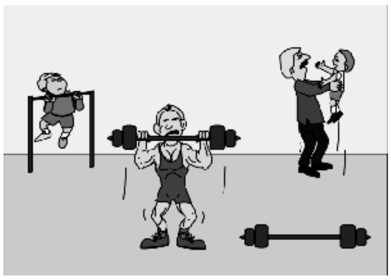


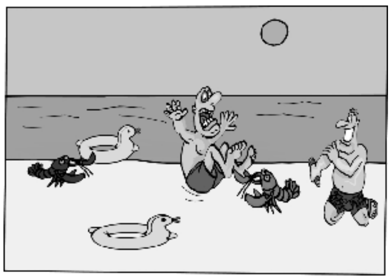







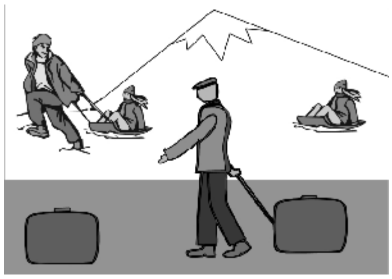
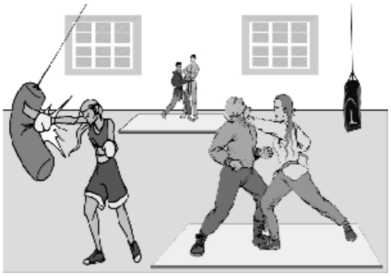


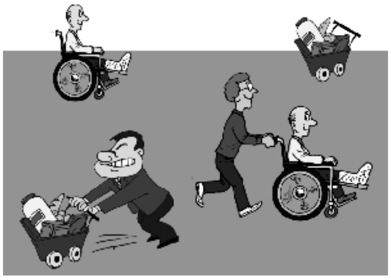
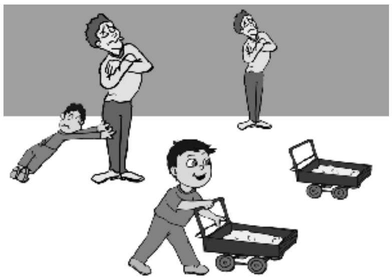
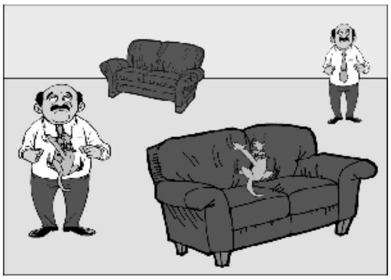
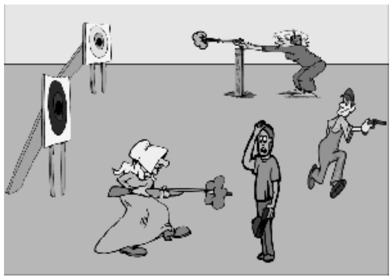


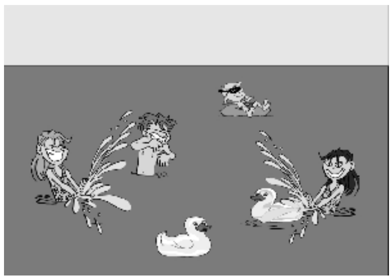


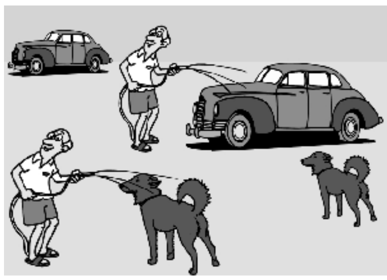





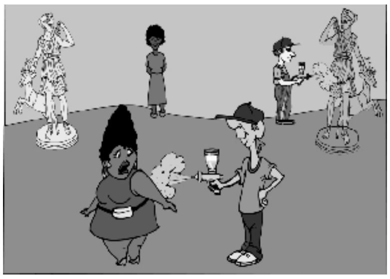


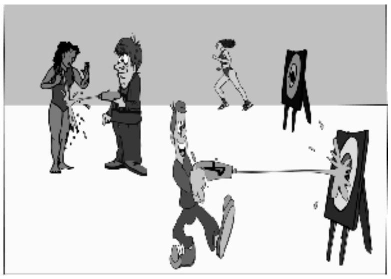




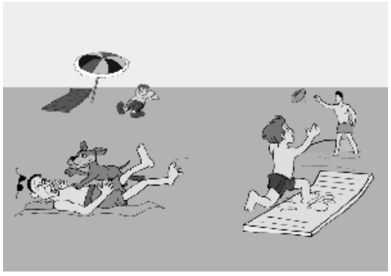





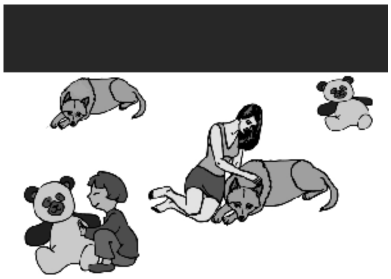


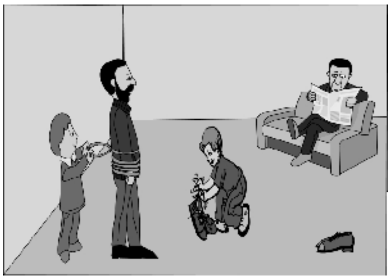


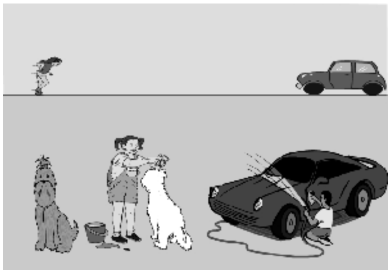


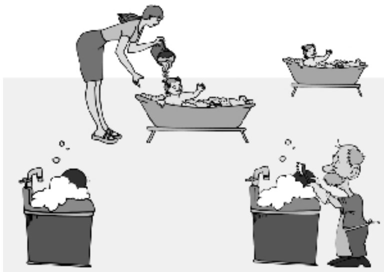




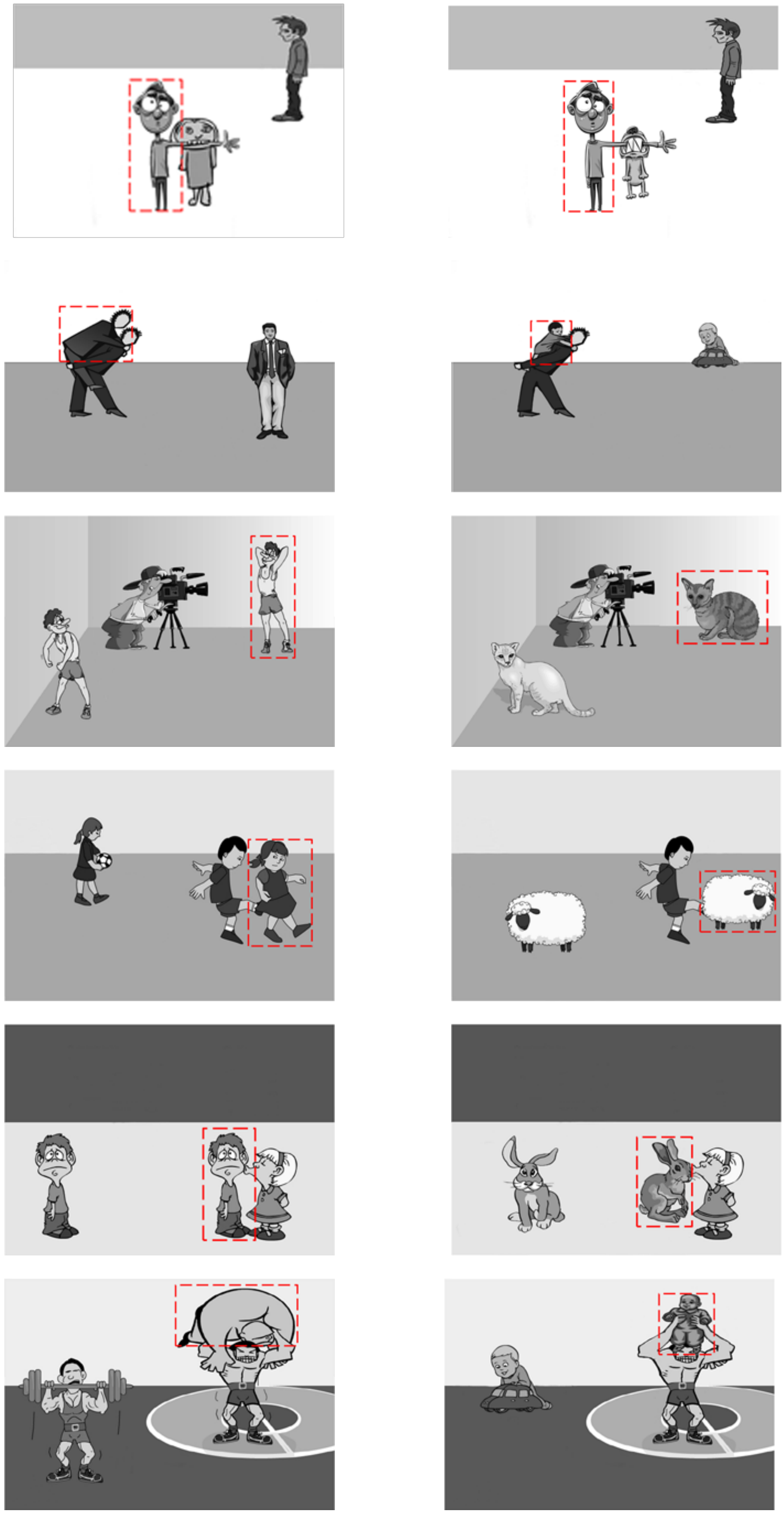


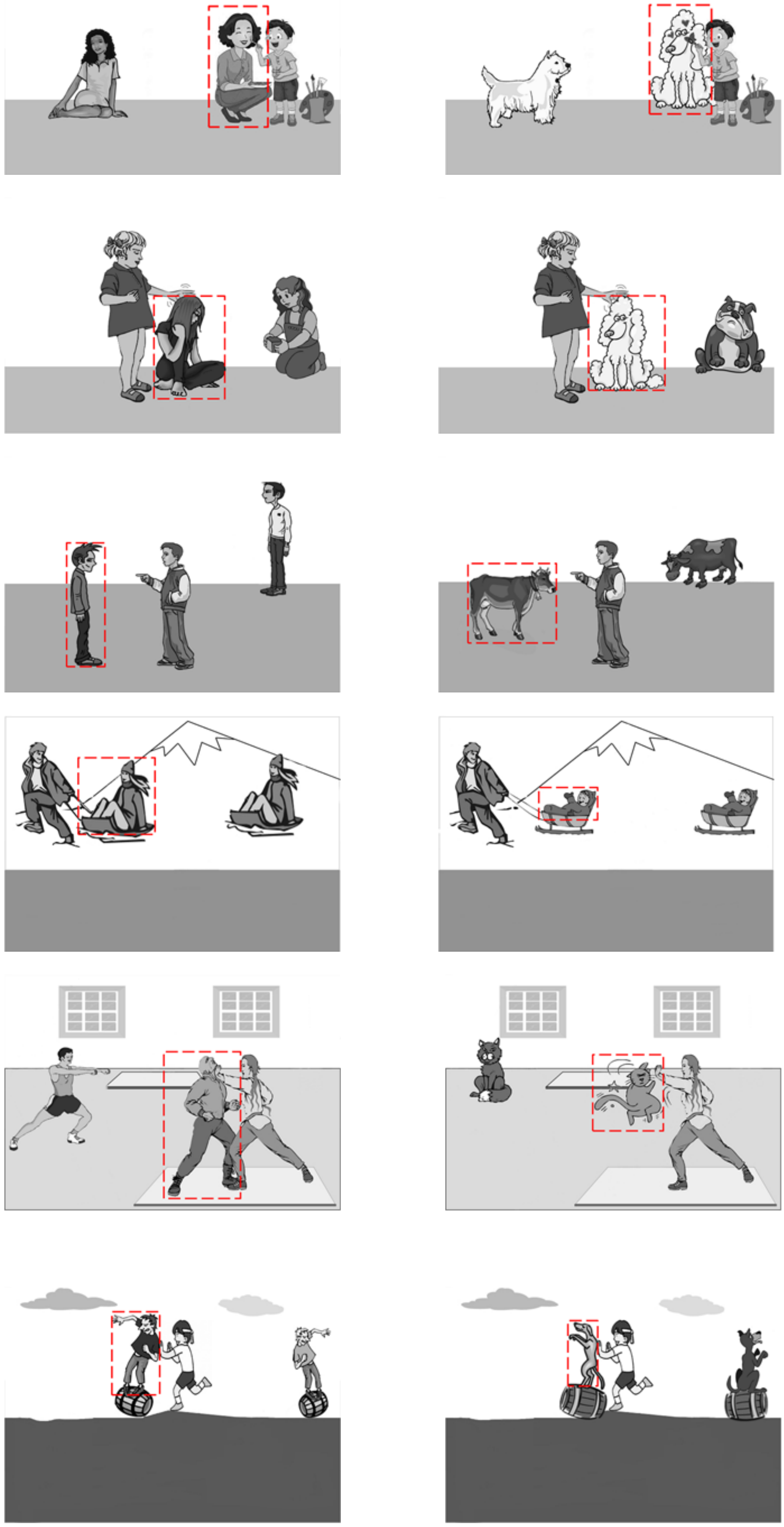


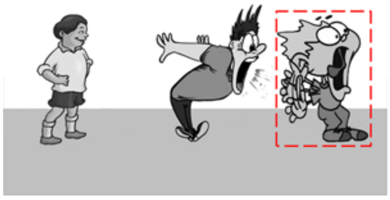
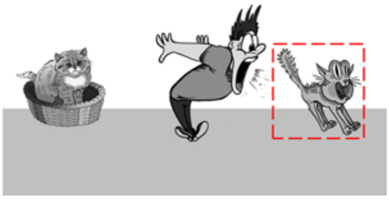
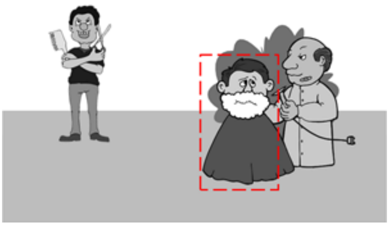
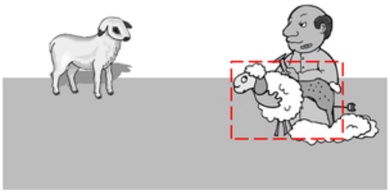
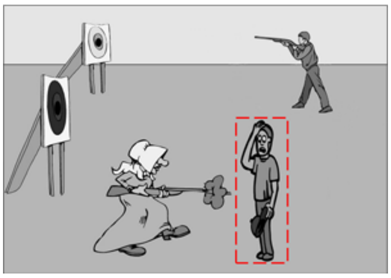
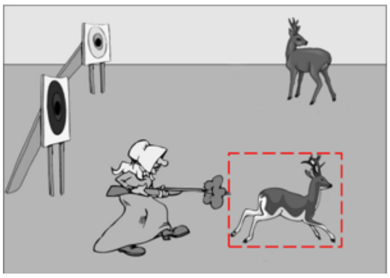


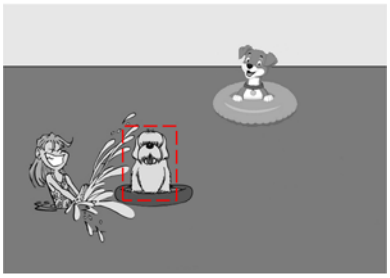


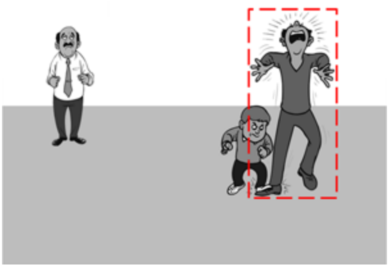
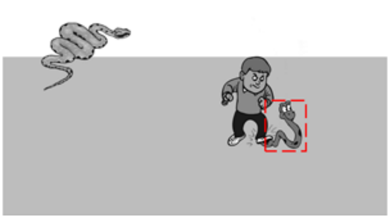
Appendix B: Picture items used in Study 3.

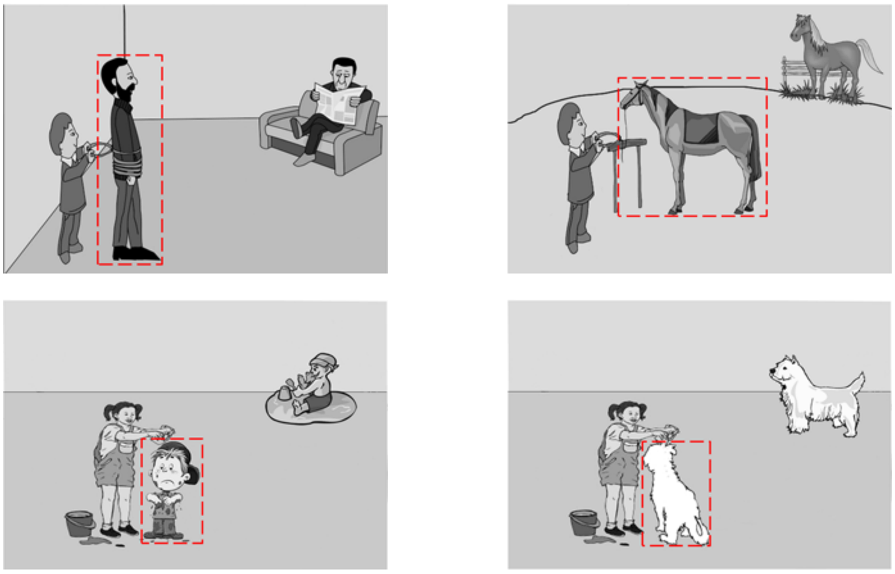








Appendix C: The homonym and homograph items used in Study 1-3.

Homograph stimuli

Prime (dominant)	Prime (subordinate)	Target	Items used for children
GUITAR	FISH	BASS	
ARROW	CURTSEY	BOW	Yes
HELP	MACHINE	CONSOLE	
STORM	LEAVE	DESERT	Yes
BIRD	SWIM	DOVE	Yes
EXIT	AWE	ENTRANCE	
SMELL	ANGER	INCENSE	
WRONG	SICK	INVALID	
FOLLOW	PENCIL	LEAD	Yes
SECOND	SMALL	MINUTE	Yes
THING	DISAGREE	OBJECT	Yes
ALLOW	LICENSE	PERMIT	
GIFT	DEMONSTRATE	PRESENT	
VEGETABLE	MAKE	PRODUCE	
WORK	PREDICT	PROJECT	
ALBUM	WRITE	RECORD	Yes
DECLINE	GARBAGE	REFUSE	
BOAT	FIGHT	ROW	Yes
MATH	SUBMIT	SUBJECT	
RIP	CRY	TEAR	Yes
MAD	DEFEAT	UPSET	
BLOW	TURN	WIND	Yes
HURT	WIND	WOUND	Yes
CLOSE	SUGGEST	INTIMATE	

Homonym Stimuli

Prime (dominant)	Prime (subordinate)	Target	Items used for children
BAT	DRESS	BALL	Yes
MONEY	RIVER	BANK	Yes
DOG	LOG	BARK	Yes
BALL	CAVE	BAT	
SHOE	CAR	BOOT	
LIGHT	FLOWER	BULB	Yes
BUS	FOOTBALL	COACH	Yes
ICE	MEDICINE	COLD	Yes
MOON	WORM	EARTH	
WINDOW	CUP	GLASS	
BREAD	TRAFFIC	JAM	

FIRE	GAME	MATCH	
CHEESE	COMPUTER	MOUSE	Yes
HAMMER	FINGER	NAIL	
ENVELOPE	MUSIC	NOTE	Yes
ALMOND	SCREW	NUT	
HAND	TREE	PALM	
TENNIS	NOISE	RACKET	
FINGER	TELEPHONE	RING	Yes
SHOVEL	ACE	SPADE	
BARN	DRINK	STRAW	
CUP	PLATE	TEA	Yes
CROSS	CLOCK	TICK	
SEA	HAND	WAVE	Yes

Trial structure and design

For the homograph and homonym tasks, the trials structure was similar and only differed in the participants task and dependent measure (speech onset time or response time). In each trial, participants first read or named the prime word (e.g., *blow, money*), which was presented for 2000ms. After an inter-trial time of 300ms, they named the target ambiguous word (e.g., *wind*) or made a relatedness judgment to the target (e.g., *bank*). Ambiguous words and unambiguous filler items were presented in random order within a block.

Appendix D: Additional Results for Study 1

Speech onset times (SOTs) from phrase production task:

SOTs were obtained from the recorded audio files by identifying the beginning of the first sound relative to the beginning of the red square presentation on the screen (see Design and Procedure from the main article). Initial disfluencies, if any, were included in the SOT, i.e., the beginning of a fluent noun phrase (e.g., *the man*) was taken as the SOT, rather than initial hesitations that could include a long *the*. Here, we compared SOTs among animate target phrases to examine whether reversibility influenced the SOTs, given that all animate words were comparable in conceptual salience, word frequency and length.

It is well known that inanimate nouns are less conceptually accessible than animate ones (Bock & Warren, 1985; Griffin, 2001; Griffin & Bock, 2000; McDonald et al., 1993), leading to longer speech onset times in naming and sentence production. Due to design constraints, the inanimate nouns in our study were also less frequent and longer than animate ones such as *boy, girl* and *man,* and inanimate references often required the planning of two words, e.g., *teddy bear*, *rubber duck*. Because we are interested in sentence-level competition processes rather than lexical accessibility to compare to comprehension, SOTs might be less informative in this respect, as they would be compounded with lexical retrieval variables. Indeed, the SOTs for inanimate-target phrases (mean: 1688, SD: 764) were longer than those for animate-target phrases (mean: 1578, SD: 709), as previously reported. Therefore, we only examined reversibility effects within animate-target phrases, which contained words of similar lexical characteristics.

As shown in Table SM1, there was a significant main effect of reversibility, suggesting that SOTs were longer for high-reversibility than low reversibility items. Similar to the analysis of verbal fluency, there was also a significant main effect of Homograph Inhibition Scores such that those with poorer inhibition required longer initiation time to name animate-target pictures than those with poorer inhibition scores. Although there was no interaction and no effect of other predictors, the high- and medium-reversibility condition showed a stronger relationship to inhibition scores than low-reversibility items (Figure SM1). Overall, these results suggest that speech onset times take longer for more reversible items and vary as a function of inhibition scores.

Table SM1. Model results predicting production SOT (ms) from reversibility and individual difference measures

	Coef.	**SE**	***t***	***p-value***
Intercept	1654.12	83.16	19.89	<0.001
High v Low reversibility	-.180	94.87	-1.90	0.06
Homograph Inhibition	159.95	52.47	3.04	0.003

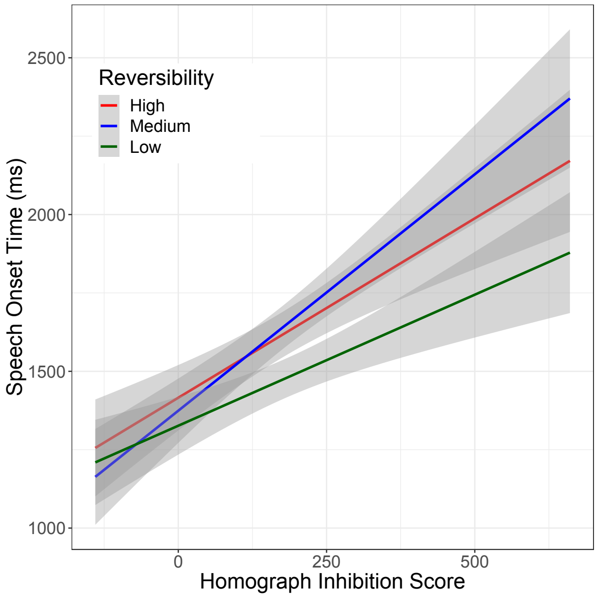


Figure SM1: Relationship between Speech Onset Times in production and Homograph inhibition scores as a function of Reversibility.

References

Acheson, D. J., & Hagoort, P. (2013). Stimulating the brain's language network: syntactic ambiguity resolution after TMS to the inferior frontal gyrus and middle temporal gyrus. *Journal of cognitive neuroscience*, *25*(10), 1664-1677.

Acheson, D. J., & MacDonald, M. C. (2009). Verbal working memory and language production: Common approaches to the serial ordering of verbal information. *Psychological Bulletin*, *135*(1), 50.

Acheson, D. J., MacDonald, M. C., & Wells, J. B. (2008). Supplementary materials from the article "New and updated tests of print exposure and reading abilities in college students". *Behavior Research Methods*, *40*(1), 278–289.

Allum, P. H., & Wheeldon, L. (2009). Scope of lexical access in spoken sentence production: Implications for the conceptual-syntactic interface. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(5), 1240–1255. https://doi.org/http://dx.doi.org/10.1037/a0016367

Arnold, J. E. (2010). How speakers refer: The role of accessibility. *Linguistics and Language Compass*, *4*(4), 187–203. https://doi.org/10.1111/j.1749-818X.2010.00193.x

Arnold, J. E., & Griffin, Z. M. (2007). The effect of additional characters on choice of referring expression: Everyone counts. *Journal of Memory and Language*, *56*, 521–536.

Arnon, I. (2010). Rethinking child difficulty: The effect of NP type on children's processing of relative clauses in Hebrew. *Journal of Child Language*, *37*(1), 27–57.

Arosio, F., Guasti, M. T., & Stucchi, N. (2011). Disambiguating information and memory resources in children's processing of Italian relative clauses. *Journal of Psycholinguistic Research*, *40*(2), 137–154.

Baddeley, A. D., & Hitch, G. J. (2000). Development of working memory: Should the Pascual-Leone and the Baddeley and Hitch models be merged?. *Journal of experimental child psychology*, *77*(2), 128-137.

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, E., Bretherton, I., & Snyder, L. S. (1991). *From first words to grammar: Individual differences and dissociable mechanisms* (Vol. 20). Cambridge University Press.

Bedard, A.-C., Nichols, S., Barbosa, J. A., Schachar, R., Logan, G. D., & Tannock, R. (2002). The development of selective inhibitory control across the life span. *Developmental Neuropsychology*, *21*(1), 93–111.

Bedny, M., McGill, M., & Thompson-Schill, S. L. (2008). Semantic adaptation and competition during word comprehension. *Cerebral Cortex*, *18*(11), 2574–2585.

Belke, E. (2008). Effects of working memory load on lexical-semantic encoding in language production. *Psychonomic Bulletin & Review*, *15*(2), 357–363.

Blything, L. P., & Cain, K. (2016). *Children ' s Processing and Comprehension of Complex Sentences Containing Temporal Connectives : The Influence of Memory on the Time Course of Accurate Responses*. *52*(10), 1517–1529.

Blything, L. P., & Cain, K. (2019). The role of memory and language ability in children's production of two-clause sentences containing before and after. *Journal of Experimental Child Psychology*, *182*, 61–85.

Blything, L. P., Davies, R., & Cain, K. (2015). Young children's comprehension of temporal relations in complex sentences: The influence of memory on performance. *Child Development*, *86*(6), 1922–1934.

Bock, J K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review*, *89*(1), 1.

Bock, J K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, *18*(3), 355–387.

Bock, J K. (1987). An effect of the accessibility of word forms on sentence structure. *Journal of Memory and Language*, *26*, 119–137.

Bock, J K, Dell, G. S., Chang, F., & Onishi, K. H. (2007). Persistent structural priming from language comprehension to language production. *Cognition*, *104*(3), 437–458.

Bock, J K, & Levelt, W. J. M. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics*. Academic Press.

Bock, J K, Loebell, H., & Morey, R. (1992). From conceptual roles to structural relations: Bridging the syntactic cleft. *Psychological Review*, *99*, 150–171.

Bock, J K, & Warren, R. K. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition*, *21*, 47–67.

Bock, J Kathryn, & Warren, R. K. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition*, *21*(1), 47–67.

Borella, E., Carretti, B., & Pelegrina, S. (2010). The specific role of inhibition in reading comprehension in good and poor comprehenders. *Journal of Learning Disabilities*, *43*(6), 541–552.

Bornstein, M. H., Hahn, C.-S., & Haynes, O. M. (2004). Specific and general language performance across early childhood: Stability and gender considerations. *First Language*, *24*(3), 267–304.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict Monitoring and Cognitive Control. *Psychological Review*, Vol. 108, pp. 624–652. https://doi.org/10.1037//0033-295X.108.3.624

Brandt, S., Kidd, E., Lieven, E., & Tomasello, M. (2009). The discourse bases of relativization: An investigation of young German and English-speaking children’s comprehension of relative clauses. *Cognitive Linguistics*, *20*(3), 539–570.

Branigan, H. (2007). Syntactic priming. *Language and Linguistics Compass*, *1*(1–2), 1–16.

Branigan, H. P., & Messenger, K. (2016). Consistent and cumulative effects of syntactic experience in children’s sentence production: Evidence for error-based implicit learning. *Cognition*, *157*, 250–256.

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue.

Cognition, *75*(2), B13--B25.

Branigan, H. P., Pickering, M. J., & McLean, J. F. (2005). Priming prepositional-phrase attachment during comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 468.

Branigan, H. P., Pickering, M. J., McLean, J. F., & Stewart, A. J. (2006). The role of local and global syntactic structure in language production: Evidence from syntactic priming. *Language and Cognitive Processes*, *21*(7–8), 974–1010.

Branigan, H. P., Pickering, M. J., & Tanaka, M. (2008). Contributions of animacy to grammatical function assignment and word order during production. *Lingua*, *118*(2), 172–189.

Braze, D., Tabor, W., Shankweiler, D. P., & Mencl, W. E. (2007). Speaking Up for Vocabulary. *Journal of Learning Disabilities*, *40*(3), 226–243. https://doi.org/10.1177/00222194070400030401

Brown-Schmidt, S., & Konopka, A. E. (2008). Little houses and casas pequeñas: Message formulation and syntactic form in unscripted speech with speakers of English and Spanish. *Cognition*, *109*(2), 274–280.

Buchanan, T. W., Laures-Gore, J. S., & Duff, M. C. (2014). Acute stress reduces speech fluency. *Biological Psychology*, *97*(1), 60–66. https://doi.org/10.1016/j.biopsycho.2014.02.005

Cain, K. (2006). Individual differences in children’s memory and reading comprehension: An investigation of semantic and inhibitory deficits. *Memory*, *14*(5), 553–569.

Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences*, *22*(1), 77–126.

Caramazza, A., & Costa, A. (2000). The semantic interference effect in the picture-word interference paradigm: does the response set matter? *Cognition*, *75*(B51), B51--B64.

Carretti, B., Borella, E., Cornoldi, C., & De Beni, R. (2009). Role of working memory in explaining the performance of individuals with specific reading comprehension difficulties: A meta-analysis. *Learning and Individual Differences*, *19*(2), 246–251.

Catani, M., Howard, R. J., Pajevic, S., & Jones, D. K. (2002). Virtual in vivo interactive dissection of white matter fasciculi in the human brain. *Neuroimage*, *17*(1), 77–94.

Catts, H. W., Hogan, T. P., & Adlof, S. M. (2005). Developmental changes in reading and reading disabilities. In *The connections between language and reading disabilities* (pp. 38–51). Psychology Press.

Chang, F. (2002). Symbolically speaking: A connectionist model of sentence production. In *Cognitive Science* (Vol. 26). https://doi.org/10.1016/S0364-0213(02)00079-4

Chang, Franklin, Dell, G. S., & Bock, J. K. (2006). Becoming Syntactic. *Psychological Review*, *113*(2), 234–272.

Choi, Y., & Trueswell, J. C. (2010). Children's (in) ability to recover from garden paths in a verb-final language: Evidence for developing control in sentence processing. *Journal of Experimental Child Psychology*, *106*(1), 41–61.

Corrêa, L. M. S. (1995). An alternative assessment of children's comprehension of relative clauses. *Journal of Psycholinguistic Research*, *24*(3), 183–203.

Costa, A., Alario, F.-X., & Caramazza, A. (2005). On the categorical nature of the semantic interference effect in the picture-word interference paradigm. *Psychonomic Bulletin & Review*, *12*(1), 125–131.

Cowan, N., Elliot, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. A. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology, 51,* 42–100.

Cowan, N., Rouder, J. N., Blume, C. L., & Saults, J. S. (2012). Models of verbal working memory capacity: What does it take to make them work? *Psychological Review*, *119*(3), 480.

Crone, E. A., Wendelken, C., Donohue, S., van Leijenhorst, L., & Bunge, S. A. (2006). Neurocognitive development of the ability to manipulate information in working memory. *Proceedings of the National Academy of Sciences*, *103*(24), 9315–9320.

D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology*, *19*(5), 381–385.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450–466.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, *3*(4), 422–433.

de Ruiter, L. E., Theakston, A. L., Brandt, S., & Lieven, E. V. M. (2018). Iconicity affects children's comprehension of complex sentences: The role of semantics, clause order, input and individual differences. *Cognition*, *171*, 202–224.

de Zubicaray, G. I., McMahon, K. L., Eastburn, M. M., & Wilson, S. J. (2002). Orthographic/phonological facilitation of naming responses in the picture–word task: An event-related fMRI study using overt vocal responding. *NeuroImage*, *16*, 1080–1093.

Dell, G S, Chang, F., & Griffin, Z. M. (2001). Connectionist models of language production: Lexical access and grammatical encoding. In *Christiansen, Morten H [Ed]*.

Dell, G S, & O'Seaghdha, P. G. (1994). Inhibition in interactive activation models of linguistic selection and sequencing. In D. Dagenbach (Ed.), *Using Smart Source Parsing. Inhibitory processes in attention, memory, and language*: *Vol. xiv* (pp. 409–453). San Diego, CA: Academic Press.

Dell, Gary S, Burger, L. K., & Svec, W. R. (1997). Language production and serial order: A functional analysis and a model. *Psychological Review*, *104*(1), 123–147. https://doi.org/http://dx.doi.org/10.1037/0033-295X.104.1.123

Dell, Gary S, & Chang, F. (2014). The p-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1634). https://doi.org/10.1098/rstb.2012.0394

Dell, Gary S, Chang, F., & Griffin, Z. M. (1999). Connectionist models of language production: Lexical access and grammatical encoding. *Cognitive Science: A Multidisciplinary Journal*, *23*(4), 517–542. https://doi.org/http://dx.doi.org/10.1016/S0364-0213%2899%2900014-2

Dell, Gary S, Juliano, C., & Govindjee, A. (1993). Structure and content in language production: A theory of frame constraints in phonological speech errors. *Cognitive Science: A Multidisciplinary Journal*, *17*(2), 149–195.

Dell, Gary S, Oppenheim, G. M., & Kittredge, A. K. (2008). Saying the right word at the right time: Syntagmatic and paradigmatic interference in sentence production. *Language and Cognitive Processes*, *23*(4), 583–608. https://doi.org/http://dx.doi.org/10.1080/01690960801920735

Diamond, A. (2013). Executive Functions. *Annual Review of Psychology*, *64*(1), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Dickinson, D. K., Golinkoff, R. M., & Hirsh-Pasek, K. (2010). Speaking out for language: Why language is central to reading development. *Educational Researcher*, *39*(4), 305–310.

Diessel, H. (2004). *The acquisition of complex sentences* (Vol. 105). Cambridge University Press.

Diessel, H., & Tomasello, M. (2000). The development of relative clauses in spontaneous child speech. *Cognitive Linguistics*, *11*(1/2), 131–152.

Diessel, H., & Tomasello, M. (2005). A new look at the acquisition of relative clauses. *Language*, 882–906.

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody picture vocabulary test*. Pearson Assessments.

Durrleman, S., Marinis, T., & Franck, J. (2016). Syntactic complexity in the comprehension of wh-questions and relative clauses in typical language development and autism. *Applied Psycholinguistics*, *37*(6), 1501–1527.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*(2), 211.

Farrant, B. M., & Zubrick, S. R. (2013). Parent--child book reading across early childhood and child vocabulary in the early school years: Findings from the Longitudinal Study of Australian Children. *First Language*, *33*(3), 280–293.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., … Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, i--185.

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology*, *42*(1), 98.

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, *25*(3), 348–368.

Ferreira, F., & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language*, *46*(1), 57–84.

Ferreira, V. S. (1996). Is it better to give than to donate? Syntactic flexibility in language production. *Journal of Memory and Language*, *35*(5), 724–755. https://doi.org/http://dx.doi.org/10.1006/jmla.1996.0038

Ferreira, V. S., & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology*, *40*(4), 296–340.

Ferretti, T. R., McRae, K., & Hatherell, A. (2001). Integrating verbs, situation schemas, and thematic role. *Journal of Memory & Language*, *44*(4), 516–547.

Fitz, H., Chang, F., & Christiansen, M. H. (2011). A connectionist account of the acquisition and processing of relative clauses. In E Kidd (Ed.), *The acquisition of relative clauses. Processing, typology and function* (pp. 40–60). Amsterdam, The Neatherlands: John Bejamins.

Flett, S. (2006). *A comparison of syntactic representation and processing in first and second language production*. University of Edinburgh.

Flett, S., Branigan, H. P., & Pickering, M. J. (2013). Are non-native structural preferences affected by native language preferences? *Bilingualism: Language and Cognition*, *16*(4), 751–760.

Frazier, L. (1987). Sentence processing: A tutorial review. *Attention and Performance*.

Frazier, L., & Clifton, C. (1989). Successive cyclicity in the grammar and the parser. *Language and*

Cognitive Processes, *4*(2), 93–126.

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: a latent-variable analysis. *Journal of Experimental Psychology. General*, *133*(1), 101–135. https://doi.org/10.1037/0096-3445.133.1.101

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

Friedmann, N., & Novogrodsky, R. (2004). The acquisition of relative clause comprehension in Hebrew: A study of SLI and normal development. *Journal of Child Language*, *31*(3), 661–681.

Fisk, J. E., & Sharp, C. A. (2004). Age-related impairment in executive functioning: Updating, inhibition, shifting, and access. *Journal of clinical and experimental neuropsychology*, *26*(7), 874-890

Fukumura, K., & van Gompel, R. (2010). The effects of animacy in the choice of referring expressions. *Language and Cognitive Processes*, *26*(June 2015), 37–41. https://doi.org/10.1080/01690965.2010.506444

Fukumura, K., van Gompel, R. P. G., Harley, T., & Pickering, M. J. (2011). How does similarity-based interference affect the choice of referring expression? *Journal of Memory and Language*, *65*(3), 331–344.

Gennari, S. P., & MacDonald, M. C. (2008). Semantic indeterminacy in object relative clauses. *Journal of Memory and Language*, *58*(2), 161–187.

Gennari, S. P., & MacDonald, M. C. (2009). Linking production and comprehension processes: The case of relative clauses. *Cognition*, *111*(1), 1–23. https://doi.org/10.1016/j.cognition.2008.12.006

Gennari, S. P., MacDonald, M. C., Postle, B. R., & Seidenberg, M. S. (2007). Context-dependent interpretation of words: evidence for interactive neural processes. *Neuroimage*, *35*(3), 1278–1286.

Gennari, S. P., Mirković, J., & MacDonald, M. C. (2012). Animacy and competition in relative clause production: A cross-linguistic investigation. *Cognitive Psychology*, *65*(2), 141–176.

Gernsbacher, M. A. (1993). Less skilled readers have less efficient supression mechanisms. *Psychological Science*, *4*(5), 294–299.

Gernsbacher, M. A., & Faust, M. E. (1991). The mechanism of suppression: a component of general comprehension skill. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *17*(2), 245–262. https://doi.org/10.1037/0278-7393.17.2.245

Gernsbacher, M. A., Sauer, E. A., Geye, H. M., Schweigert, E. K., & Hill Goldsmith, H. (2008). Infant and toddler oral- and manual-motor skills predict later speech fluency in autism. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *49*(1), 43–50. https://doi.org/10.1111/j.1469-7610.2007.01820.x

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*(1), 1–76.

Gleitman, L. R., January, D., Nappa, R., & Trueswell, J. C. (2007). On the give and take between event apprehension and utterance formulation. *Journal of Memory and Language*, *57*(4), 544–569.

Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., … others. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proceedings of the National Academy of Sciences*, *101*(21), 8174–8179.

Goodluck, H., & Tavakolian, S. (1982). Competence and processing in children's grammar of relative clauses. *Cognition*, *11*(1), 1–27.

Gordon, P. C., Hendrick, R., & Johnson, M. (2001a). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(6), 1411.

Gordon, P. C., Hendrick, R., & Johnson, M. (2001b). Memory Interference during language processing. *Journal of Experimental Psychology: Learning, Memory and Cogntion*, *27*(6), 1411–1423.

Gordon, P. C., Hendrick, R., Johnson, M., & Lee, Y. (2006). Similarity-based interference during language comprehension: Evidence from eye tracking during reading. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *32*(6), 1304–1321. https://doi.org/10.1037/0278-7393.32.6.1304

Gordon, P. C., & Lowder, M. W. (2012). Complex sentence processing: A review of theoretical perspectives on the comprehension of relative clauses. *Language and Linguistics Compass*, *6*(7), 403-415.

Gottlob, L. R., Goldinger, S. D., Stone, G. O., & Van Orden, G. C. (1999). Reading homographs: Orthographic, phonologic, and semantic dynamics. *Journal of Experimental Psychology Human Perception and Performance*, *25*(2), 561–574.

Griffin, Z M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition*, *82*(1), B1–B14. https://doi.org/10.1016/S0010-0277(01)00138-X

Griffin, Z M, & Bock, J. K. (2000). What the eyes say about speaking. *Psychological Science : A Journal of the American Psychological Society / APS*, *11*, 274–279. https://doi.org/10.1111/1467-9280.00255

Griffin, Zenzi M. (2003). A reversed word length effect in coordinating the preparation and articulation of words in speaking. *Psychonomic Bulletin & Review*, *10*(3), 603–609.

Griffin, Zenzi M, & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, *11*(4), 274–279.

Grimm, A., Müller, A., Hamann, C., & Ruigendijk, E. (2011). *Production-comprehension asymmetries in child language* (Vol. 43). Walter de Gruyter.

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

Hala, S., Pexman, P. M., & Glenwright, M. (2007). Priming the meaning of homographs in typically developing children and children with autism. *Journal of Autism and Developmental Disorders*, *37*(2), 329–340. https://doi.org/10.1007/s10803-006-0162-6

Harlaar, N., Dale, P. S., & Plomin, R. (2007). Reading exposure: A (largely) environmental risk factor with environmentally-mediated effects on reading performance in the primary school years. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *48*(12), 1192–1199. https://doi.org/10.1111/j.1469-7610.2007.01798.x

Hartsuiker, R. J., & Barkhuysen, P. N. (2006). Language production and working memory: The case of subject-verb agreement. *Language and Cognitive Processes*, *21*(1-3), 181-204.

Hartsuiker, R. J., & Kolk, H. H. J. (1998). Syntactic facilitation in agrammatic sentence production. *Brain and Language*, *62*(2), 221–254.

Henderson, L., Snowling, M., & Clarke, P. (2013). Accessing, Integrating, and Inhibiting Word Meaning in Poor Comprehenders. *Scientific Studies of Reading*, *17*(3), 177–198.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393–402.

Hoff, E., Quinn, J. M., & Giguere, D. (2018). What explains the correlation between growth in vocabulary and grammar ? New evidence from latent change score analyses of simultaneous bilingual development. *Developmental Science*, *21*(2), e12536. https://doi.org/10.1111/desc.12536

Hsiao, Y., Gao, Y., & MacDonald, M. C. (2014). Agent-patient similarity affects sentence structure in language production: Evidence from subject omissions in Mandarin. *Frontiers in Psychology*, *5*, 1015.

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

Hsu, N. S., Jaeggi, S. M., & Novick, J. M. (2017). A common neural hub resolves syntactic and non-syntactic conflict through cooperation with task-specific networks. *Brain and Language*, *166*, 63–77. https://doi.org/10.1016/j.bandl.2016.12.006

Hsu, N. S., & Novick, J. M. (2016). Dynamic Engagement of Cognitive Control Modulates Recovery

From Misinterpretation During Real-Time Language Processing. *Psychological Science*, *27*(4), 572–582. https://doi.org/10.1177/0956797615625223

Huettig, F., & Hartsuiker, R. J. (2008). When you name the pizza you look at the coin and the bread: Eye movements reveal semantic activation during word production. *Memory & Cognition*, *36*(2), 341–360.

Humphreys, G. F., & Gennari, S. P. (2014). Competitive mechanisms in sentence processing: Common and distinct production and reading comprehension networks linked to the prefrontal cortex. *NeuroImage*, *84*, 354–366. https://doi.org/10.1016/j.neuroimage.2013.08.059

Humphreys, G. F., Mirković, J., & Gennari, S. P. (2016). Similarity-based competition in relative clause production and comprehension. *Journal of Memory and Language*, *89*(January), 200–221. https://doi.org/10.1016/j.jml.2015.12.007

Hussey, E. K., Harbison, J. I., Teubner-Rhodes, S. E., Mishler, A., Velnoskey, K., & Novick, J. M. (2017). Memory and language improvements following cognitive control training. *Journal of Experimental Psychology: Learning Memory and Cognition*, *43*(1), 23–58. https://doi.org/10.1037/xlm0000283

Iacobucci, D., Schneider, M. J., Popovich, D. L., & Bakamitsos, G. A. (2016). Mean centering helps alleviate “micro” but not “macro” multicollinearity. *Behavior Research Methods*, *48*(4), 1308–1317. https://doi.org/10.3758/s13428-015-0624-x

Iwasaki, N. (2011). Incremental Sentence Production: Observations from Elicited Speech Errors in Japanese. *Processing and Producing Head-Final Structures*, 131–151.

January, D., Trueswell, J. C., Thompson-Schill, S. L., January, D., Trueswell, J. C., & Thompson-Schill, S. L. (2009). Co-localization of stroop and syntactic ambiguity resolution in Broca’s area: implications for the neural basis of sentence processing. *Journal of Cognitive Neuroscience*, *21*(12), 2434–2444. https://doi.org/10.1162/jocn.2008.21179

Just, Marcel A, & Carpenter, P. A. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychological Review*, *99*(1), 122.

Just, Marcel Adam, Carpenter, P. A., Keller, T. A., Eddy, W. F., & Thulborn, K. R. (1996). Brain activation modulated by sentence comprehension. *Science*, *274*(5284), 114–116.

Kawamoto, A. H. (1993). Nonlinear dynamics in the resolution of lexical ambiguity: A parallel distributed processing account. *Journal of Memory and Language*, *32*, 474–516. https://doi.org/DOI: 10.1006/jmla.1993.1026

Kempen, G., & Huijbers, P. (1983). The lexicalization process in sentence production and naming: Indirect election of words. *Cognition*, *14*(2), 185–209.

Khanna, M. M., & Boland, J. E. (2010). Children's use of language context in lexical ambiguity resolution. *The Quarterly Journal of Experimental Psychology*, *63*(1), 160–193.

Kidd, Evan. (2013). The role of verbal working memory in children's sentence comprehension: A critical review. *Topics in Language Disorders*, *33*(3), 208–223.

Kidd, Evan, & Bavin, E. L. (2002). English-speaking children's comprehension of relative clauses: Evidence for general-cognitive and language-specific constraints on development. *Journal of Psycholinguistic Research*, *31*(6), 599–617.

Kidd, Evan, Brandt, S., Lieven, E., & Tomasello, M. (2007). Object relatives made easy: A cross-linguistic comparison of the constraints influencing young children's processing of relative clauses. *Language and Cognitive Processes*, *22*(6), 860–897.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, *30*(5), 580–602.

Konopka, A. E. (2012). Planning ahead: How recent experience with structures and words changes the scope of linguistic planning. *Journal of Memory and Language*, *66*(1), 143–162. https://doi.org/http://dx.doi.org/10.1016/j.jml.2011.08.003

Konopka, A. E. (2018). Encoding actions and verbs: Tracking the time-course of relational encoding during message and sentence formulation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Konopka, A. E., & Meyer, A. S. (2014). Priming sentence planning. *Cognitive Psychology*, *73*, 1–40.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). **lmerTest** Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

Lee, J. (2011). Size matters: Early vocabulary as a predictor of language and literacy competence. *Applied Psycholinguistics*, *32*(1), 69–92.

Lee, E. K., Brown-Schmidt, S., & Watson, D. G. (2013). Ways of looking ahead: Hierarchical planning in language production. *Cognition*, *129*(3), 544-562.

Levelt, W J M, Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–75.

Levelt, Willem J M. (1982). Linearization in describing spatial networks. In *Processes, beliefs, and questions* (pp. 199–220). Springer.

Levelt, Willem J M. (1993). *Speaking: From intention to articulation* (Vol. 1). MIT press.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177.

Lewis, R. L., & Vasishth, S. (2005). An Activation-Based Model of Sentence Processing as Skilled Memory Retrieval. *Cognitive Science*, *29*(3), 375–419.

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, *10*(10), 447–454. https://doi.org/http://dx.doi.org/10.1016/j.tics.2006.08.007

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review*, *21*(4), 861–883.

Lorsbach, T. C., Katz, G. A., & Cupak, A. J. (1998). Developmental differences in the ability to inhibit the initial misinterpretation of garden path passages. *Journal of Experimental Child Psychology*, *71*(3), 275–296.

Lorsbach, T. C., & Reimer, J. F. (1997). Developmental changes in the inhibition of previously relevant information. *Journal of Experimental Child Psychology*, *64*(3), 317–342.

Luna, B., Padmanabhan, A., & O'Hearn, K. (2010). What has fMRI told us about the development of cognitive control through adolescence? *Brain and Cognition*, *72*(1), 101–113.

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, *4*(April), 226. https://doi.org/10.3389/fpsyg.2013.00226

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996).

MacDonald, M. C., Pearlmutter, N., & Seidenberg, M. (1994). Lexical Nature of Syntactic Ambiguity Resolution. *Psychological Review*, *101*(4), 676–703.

Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word*, *15*(1), 19–44.

MacWhinney, B. (1977). Starting points. *Language*, 152–168.

MacWhinney, B., & Pleh, C. (1988). The processing of restrictive relative clauses in Hungarian. *Cognition*, *29*(2), 95–141.

Mak, W. M., Vonk, W., & Schriefers, H. (2002). The influence of animacy on relative clause processing. *Journal of Memory and Language*, *47*(1), 50–68.

Marchina, S., Wang, J., Wan, C. Y., Norton, A. C., & Schlaug, G. (2013). Predicting speech fluency and naming abilities in aphasic patients. *Frontiers in Human Neuroscience*, *7*(December), 1–13. https://doi.org/10.3389/fnhum.2013.00831

Martin, R. C., Lesch, M. F., & Bartha, M. C. (1999). Independence of input and output phonology in word processing and short-term memory. *Journal of Memory and Language*, *41*(1), 3-29.

Martin, R. C. (2005). Components of Short-Term Memory and Their Relation to Language Processing. *Current Directions in Psychological Science*, *14*(4), 204–208. https://doi.org/10.1111/j.0963-7214.2005.00365.x

McDaniel, D., McKee, C., & Bernstein, J. B. (1998). How children's relatives solve a problem for minimalism. *Language*, *74*(2), 308–334.

McDonald, J. L., Bock, J. K., & Kelly, M. H. (1993). Word and world order: Semantics, phonological, and metrical determinants of serial position. *Cognitive Psychology*, *25*, 188–230.

McElree, B. (2001). Working memory and focal attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(3), 817.

McRae, K., Ferretti, T. R., & Amyote, L. (1997). Thematic Roles as Verb-specific Concepts. *Language and Cognitive Processes*, *12*(2/3), 137–176.

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, *38*, 283–312.

Menenti, L., Gierhan, S. M. E., Segaert, K., & Hagoort, P. (2011). Shared Language: Overlap and Segregation of the Neuronal Infrastructure for Speaking and Listening Revealed by Functional MRI . *Psychological Science* , *22*(9), 1173–1182. https://doi.org/10.1177/0956797611418347

Messenger, K., Branigan, H. P., & McLean, J. F. (2011). Evidence for (shared) abstract structure underlying children's short and full passives. *Cognition*, *121*(2), 268–274.

Messenger, K., Branigan, H. P., & McLean, J. F. (2012). Is children's acquisition of the passive a staged process? Evidence from six-and nine-year-olds' production of passives. *Journal of Child Language*, *39*(5), 991–1016.

Messenger, K., & Fisher, C. (2018). Mistakes weren't made: Three-year-olds' comprehension of novel-verb passives provides evidence for early abstract syntax. *Cognition*, *178*, 118–132.

Meyer, A. S. (1996). Lexical access in phrase and sentence production: Results from picture-word interference experiments. *Journal of Memory and Language*, *35*(4), 477–496. https://doi.org/http://dx.doi.org/10.1006/jmla.1996.0026

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The Unity and Diversity of Executive Functions and Their Contributions to Complex "Frontal Lobe" Tasks: A Latent Variable Analysis. *Cognitive Psychology*, *41*(1), 49–100.

https://doi.org/10.1006/cogp.1999.0734

Montag, J L, & MacDonald, M. C. (2015). Text Exposure Predicts Spoken Production of Complex Sentences in 8-and 12-Year-Old Children and Adults. *Journal of Experimental Psychology. General*, *144*(2), 447–468. https://doi.org/10.1037/xge0000054

Montag, Jessica L, & MacDonald, M. C. (2015). Text exposure predicts spoken production of complex sentences in 8- and 12-year-old children and adults. *Journal of Experimental Psychology: General*, *144*(2), 447–468. https://doi.org/10.1037/xge0000054

Montgomery, J. W., Magimairaj, B. M., & O'Malley, M. H. (2008). Role of working memory in typically developing children's complex sentence comprehension. *Journal of Psycholinguistic Research*, *37*(5), 331–354.

Myachykov, A., Scheepers, C., Garrod, S., Thompson, D., & Fedorova, O. (2013). Syntactic flexibility and competition in sentence production: The case of English and Russian. *The Quarterly Journal of Experimental Psychology*, *66*(8), 1601–1619.

Myachykov, A., Thompson, D., Scheepers, C., & Garrod, S. (2011). Visual attention and structural choice in sentence production across languages. *Language and Linguistics Compass*, *5*(2), 95–107.

Nation, K., Adams, J. W., Bowyer-Crane, C. A., & Snowling, M. J. (1999). Working memory deficits in poor comprehenders reflect underlying language impairments. *Journal of Experimental Child Psychology*, *73*(2), 139–158.

Nigg, J. T. (2000). On inhibition/disinhibition in developmental psychopathology: views from cognitive and personality psychology and a working inhibition taxonomy. *Psychological bulletin*, *126*(2), 220.

Norbury, C. F. (2005). Barking up the wrong tree? Lexical ambiguity resolution in children with language impairments and autistic spectrum disorders. *Journal of experimental child psychology*, *90*(2), 142-171.

Norcliffe, E., Konopka, A. E., Brown, P., & Levinson, S. C. (2015). Word order affects the time course of sentence formulation in Tzeltal. *Language, Cognition and Neuroscience*, *30*(9), 1187–1208.

Novick, J. M., Drummey, A. B., Krupa, L., Kan, I. P., Nutile, L., & Teubner-Rhodes, S. (2013). To adapt or not to adapt: The question of domain-general cognitive control. *Cognition*, *129*(3), 637–651. https://doi.org/10.1016/j.cognition.2013.09.001

Novick, J. M., Hussey, E., Teubner-Rhodes, S., Harbison, J. I., & Bunting, M. F. (2014). Clearing the garden-path: improving sentence processing through cognitive control training. *Language, Cognition and Neuroscience*, *29*(2), 186–217. https://doi.org/10.1080/01690965.2012.758297

Novick, J. M., Kan, I. P., Trueswell, J. C., & Thompson-Schill, S. L. (2009). A case for conflict across multiple domains: Memory and language impairments following damage to ventrolateral prefrontal cortex. *Cognitive neuropsychology*, *26*(6), 527-567.

Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005a). Cognitive control and parsing: reexamining the role of Broca's area in sentence comprehension. *Cognitive, Affective & Behavioral Neuroscience*, *5*(3), 263–281.

Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005b). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cognitive, Affective, & Behavioral Neuroscience*, *5*(3), 263–281.

Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2010). Broca' s Area and Language Processing: Evidence for the Cognitive Control Connection. *Language and Linguistics Compass*, *4*, 906–924. https://doi.org/10.1111/j.1749-818X.2010.00244.x

Nozari, N., Dell, G. S., & Schwartz, M. F. (2011). Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive Psychology*, *63*(1), 1–33. https://doi.org/10.1016/j.cogpsych.2011.05.001

Nozari, N., Trueswell, J. C., & Thompson-Schill, S. L. (2016). The interplay of local attraction, context and domain-general cognitive control in activation and suppression of semantic distractors during sentence comprehension. *Psychonomic Bulletin and Review*, *23*(6), 1942–1953. https://doi.org/10.3758/s13423-016-1068-8

Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: A model of cumulative semantic interference during lexical access in speech production. *Cognition*, *114*(2), 227–252. https://doi.org/10.1016/j.cognition.2009.09.007

Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology*, *98*(3), 554.

Perera, C. K., & Srivastava, A. K. (2015). Animacy-Based Accessibility and Competition in Relative clause production in Hindi and Malayalam. *Journal of Psycholinguistic Research*. https://doi.org/10.1007/s10936-015-9384-0

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. *Precursors of functional literacy*, *11*, 67-86.

Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language*, *39*(4), 633–651.

Pickering, M. J., & Branigan, H. P. (1999). Syntactic priming in language production. *Trends in Cognitive Sciences*, *3*(4), 136–141.

Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: A critical review. *Psychological Bulletin*, *134*(3), 427.

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, *11*(3), 105–110. https://doi.org/http://dx.doi.org/10.1016/j.tics.2006.12.002

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *The Behavioral and Brain Sciences*, *36*(4), 329–347. https://doi.org/10.1017/S0140525X12001495

Pickering, S., & Gathercole, S. E. (2001). *Working memory test battery for children (WMTB-C)*. Psychological Corporation.

Prat-Sala, M., & Branigan, H. P. (2000). Discourse constraints on syntactic processing in language production: A cross-linguistic study in English and Spanish. *Journal of Memory and Language*, *42*(2), 168–182.

R Team Core. (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Viena, Austria.

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior*, *22*(3), 358–374.

Reali, F., & Christiansen, M. H. (2007). Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language*, *57*(1), 1–23.

Rodd, J. M., Davis, M. H., & Johnsrude, I. S. (2005). The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cerebral Cortex*, *15*(8), 1261-1269.

Roland, D., Dick, F., & Elman, J. L. (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language*, *57*(3), 348–379. https://doi.org/http://dx.doi.org/10.1016/j.jml.2007.03.002

Roth, F. P. (1984). Accelerating language learning in young children. *Journal of Child Language*, *11*(1), 89–107. https://doi.org/10.1017/S0305000900005602

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development*, *83*(5), 1762–1774.

Rubia, K., Smith, A. B., Woolley, J., Nosarti, C., Heyman, I., Taylor, E., & Brammer, M. (2006). Progressive increase of frontostriatal brain activation from childhood to adulthood during event-related tasks of cognitive control. *Human Brain Mapping*, *27*(12), 973–993.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel Distributed Processing*. Cambridge, MA: MIT Press.

Schnur, T. T., Costa, A., & Caramazza, A. (2006). Planning at the phonological level during sentence production. *Journal of Psycholinguistic Research*, *35*(2), 189–213.

Schomers, M. R., Kirilina, E., Weigand, A., Bajbouj, M., & Pulvermüller, F. (2015). Causal influence of articulatory motor cortex on comprehending single spoken words: TMS evidence. *Cerebral Cortex*, *25*(10), 3894–3902. https://doi.org/10.1093/cercor/bhu274

Scontras, G., Badecker, W., Shank, L., Lim, E., & Fedorenko, E. (2015). Syntactic complexity effects in sentence production. *Cognitive Science*, *39*(3), 559–583. https://doi.org/10.1111/cogs.12168

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General*, *125*(1), 4.

Shao, Z., Meyer, A. S., & Roelofs, A. (2013). Selective and nonselective inhibition of competitors in picture naming. *Memory & Cognition*, *41*(8), 1200–1211. https://doi.org/10.3758/s13421-013-0332-7

Shaw, P., Kabani, N. J., Lerch, J. P., Eckstrand, K., Lenroot, R., Gogtay, N., … others. (2008). Neurodevelopmental trajectories of the human cerebral cortex. *Journal of Neuroscience*, *28*(14), 3586–3594.

Simpson, G. B. (1994). Context and the processing of ambiguous words. In Morton Ann Gernsbacher (Ed.), *Handbook of psycholinguistics*: *Vol. xxii,* (pp. 359–374). San Diego, CA, US: Academic Press.

Slevc, L. R. (2011). Saying what's on your mind: Working memory effects on sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(6), 1503–1514. https://doi.org/10.1037/a0024350

Slevc, L. R., & Martin, R. C. (2016). Syntactic agreement attraction reflects working memory processes. *Journal of Cognitive Psychology*, *28*(7), 773-790.

Slobin, D. I., & Bever, T. G. (1982). Children use canonical sentence schemas: A crosslinguistic study of word order and inflections. *Cognition*, *12*(3), 229–265.

Smith, M., & Wheeldon, L. (1999). High level processing scope in spoken sentence production.

Cognition, *73*(3), 205–246. https://doi.org/http://dx.doi.org/10.1016/S0010-0277%2899%2900053-0

Smith, M., & Wheeldon, L. (2004). Horizontal Information Flow in Spoken Sentence Production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(3), 675–686. https://doi.org/http://dx.doi.org/10.1037/0278-7393.30.3.675

Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*, *49*(3), 238–299.

Spalek, K., & Thompson-Schill, S. L. (2008). Task-dependent semantic interference in language production: an fMRI study. *Brain & Language*, *107*(3), 220–228. https://doi.org/10.1016/j.bandl.2008.05.005

Staub, A. (2010). Eye movements and processing difficulty in object relative clauses. *Cognition*, *116*(1), 71–86. https://doi.org/10.1016/j.cognition.2010.04.002

Swanson, H. L. (1996). Individual and age-related differences in children’s working memory. *Memory & Cognition*, *24*(1), 70–82.

Tamm, L., Menon, V., & Reiss, A. L. (2002). Maturation of brain function associated with response inhibition. *Journal of the American Academy of Child & Adolescent Psychiatry*, *41*(10), 1231–1238.

Tan, Y., Martin, R. C., & Van Dyke, J. A. (2017). Semantic and syntactic interference in sentence comprehension: A comparison of working memory models. *Frontiers in Psychology*, *8*(FEB), 1–19. https://doi.org/10.3389/fpsyg.2017.00198

Tanaka, M. N., Branigan, H. P., & Pickering, M. J. (2010). The production of head-initial and head-final languages. In *Processing and producing head-final structures* (pp. 113–129). Springer.

Thompson-Schill, S. L., D’Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. *Proceedings of the National Academy of Sciences*, *94*(26), 14792-14797.

Thompson-Schill, S. L., Bedny, M., & Goldberg, R. F. (2005). The frontal lobes and the regulation of mental activity. *Current Opinion in Neurobiology*, *15*(2), 219–224.

Thothathiri, M., Schwartz, M. F., & Thompson- Schill, S. L. (2010). Selection for position: The role of left ventrolateral prefrontal cortex in sequencing language. *Brain and Language*, *113*, 28–38.

Tian, X., & Poeppel, D. (2013). The Effect of Imagination on Stimulation: The Functional Specificity of Efference Copies in Speech Processing. *Journal of Cognitive Neuroscience*, *25*(13), 1020–1036.

Tomlin, R. S. (1997). Mapping conceptual representations into linguistic representations: The role of attention in grammar. *Language and Conceptualization*, 162–189.

Traxler, M. J., Morris, R. K., & Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language*, *47*(1), 69–90.

Traxler, M. J., Williams, R. S., Blozis, S. A., & Morris, R. K. (2005). Working memory, animacy, and verb class in the processing of relative clauses. *Journal of Memory and Language*, *53*(2), 204–224.

Trude, A. M., & Nozari, N. (2017). Inhibitory Control Supports Referential Context Use in Language Production and Comprehension. *CogSci 2017 Proceedings*, 1218–1223.

Trueswell, J C, Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic Influences on Parsing: Use of Thematic Role Information in Syntactic Ambiguity Resolution. *Journal of Memory and Language*, *33*, 285–318.

Trueswell, John C, Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, *33*(3), 285–318.

Tunmer, W. E., & Chapman, J. W. (2012). The Simple View of Reading Redux: Vocabulary Knowledge and the Independent Components Hypothesis. *Journal of Learning Disabilities*, *45*(5), 453–466. https://doi.org/10.1177/0022219411432685

Twilley, L. C., Dixon, P., Taylor, D., & Clark, K. (1994). University of Alberta norms of relative meaning frequency for 566 homographs. *Memory & Cognition*, *22*(1), 111–126. https://doi.org/10.3758/BF03202766

Van de Velde, M., Meyer, A. S., & Konopka, A. E. (2014). Message formulation and structural assembly: Describing “easy” and “hard” events with preferred and dispreferred syntactic structures. *Journal of Memory and Language*, *71*(1), 124-144.

Van Dyke, J A., & Johns, C. L. (2012). Memory Interference as a Determinant of Language Comprehension. *Linguistics and Language Compass*, *6*, 193–211. https://doi.org/10.1002/lnc3.330

Van Dyke, J A, & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language*, *55*(2), 157–166.

Van Dyke, J. A., & McElree, B. (2011). Cue-dependent interference in comprehension. *Journal of memory and language*, *65*(3), 247-263.

Van Dyke, Julie A, Johns, C. L., & Kukona, A. (2014). Low working memory capacity is only spuriously related to poor reading comprehension. *Cognition*, *131*(3), 373–403.

Verbruggen, F., & Logan, G. D. (2008). Response inhibition in the stop-signal paradigm. *Trends in Cognitive Sciences*, *12*(11), 418–424. https://doi.org/10.1016/j.tics.2008.07.005

Verhoeven, L., van Leeuwe, J., & Vermeer, A. (2011). Vocabulary growth and reading development across the elementary school years. *Scientific Studies of Reading*, *15*(1), 8–25.

Vuong, L. C., & Martin, R. C. (2011). LIFG-based attentional control and the resolution of lexical ambiguities in sentence context. *Brain and Language*, *116*(1), 22–32. https://doi.org/10.1016/j.bandl.2010.09.012

Wagner, V., Jescheniak, J. D., & Schriefers, H. (2010). On the flexibility of grammatical advance planning during sentence production: Effects of cognitive load on multiple lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(2), 423.

Waters, G. S., & Caplan, D. (1996). The capacity theory of sentence comprehension: critique of

Just and Carpenter (1992). *Psychological Review*, *103*(4), 761--772.

Waters, G. S., & Caplan, D. (1997). Working memory and on-line sentence comprehension in patients with Alzheimer's disease. *Journal of psycholinguistic research*, *26*(4), 377-400.

Waters, G. S., & Caplan, D. (2004). Verbal working memory and on-line syntactic processing: Evidence from self-paced listening. *The Quarterly Journal of Experimental Psychology Section A*, *57*(1), 129-163.

Wechsler, D., & Hsiao-pin, C. (2011). *WASI II: Wechsler Abbreviated Scale of Intelligence. 2nd.* Psychological Corporation.

Weighall, A. R., & Altmann, G. T. M. (2011). The role of working memory and contextual constraints in children's processing of relative clauses. *Journal of Child Language*, *38*(3), 579–605.

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, *24*(11), 2143–2152.

Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., & MacDonald, M. C. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology*, *58*(2), 250–271.

Whitney, C., Kirk, M., O'Sullivan, J., Lambon Ralph, M. A., & Jefferies, E. (2011). The neural organization of semantic control: TMS evidence for a distributed network in left inferior frontal and posterior middle temporal gyrus. *Cerebral Cortex*, *21*(5), 1066–1075.

Woodard, K., Pozzan, L., & Trueswell, J. C. (2016). Taking your own path: Individual differences in executive function and language processing skills in child learners. *Journal of Experimental Child Psychology*, *141*, 187–209.

Yuill, N., Oakhill, J., & Parkin, A. (1989). Working memory, comprehension ability and the resolution of text anomaly. *British Journal of Psychology*, *80*(3), 351–361.